National and Kapodistrian University of Athens

Medical School & Department of Mathematics

Master Program of Biostatistics

Bayesian Inference for Cure Rate Models

Author:

Thomas Tsiampalis

Supervisor:

Dr. Loukia Meligkotsidou

Athens,2018

The current Master thesis was conducted in partial fulfilment of the requirements for the degree of Master in

BIOSTATISTICS

submitted to the Medical School and the Department of Mathematics, of National & Kapodistrian University of Athens.

Approved the 29/10/2018 by the examining committee:

NAME	**GRADE**	**SIGNATURE**
L. MELIGKOTSIDOU	Assistant Professor	.....
C. YIANNOUTSOS	Professor	.....
E. RIZA	Faculty member	.....

Athens,2018

To my beloved grandmother Paraskevi.

Abstract

Survival analysis consists of a set of statistical methods in the field of biostatistics, whose main aim is to study the time until the occurrence of a specified event, such as death. For the majority of these methods it is assumed that all the individuals taking part in the study are subject to the event of interest. However, there are situations where this assumption is unrealistic, since there are observations not susceptible to the event of interest or cured. For this reason, there have been developed some survival models which allow for patients that may never experience the event, usually called long-term survivors. These models, called **Cure Rate Models**, assume that, as time increases, the survival function tends to a value $p \in (0, 1)$, representing the cure rate, instead of tending to zero as in standard survival analysis.

Recently, Rocha (2016) proposed a new approach to modelling the situations in which there are long-term survivors in survival studies. His methodology was based on the use of defective distributions to model cure rates. In contrast to the standard distributions, the defective ones are characterized by having probability density functions which integrate to values less than one for certain choices of the domain of some of their parameters. The aim of the present thesis is to provide new Bayesian estimates for the parameters of the defective models used for cure rate modelling under the assumption of right censoring. We will develop Markov chain Monte Carlo (MCMC) algorithms for inferring the parameters of a broad class of defective models, both for the baseline distributions (Gompertz & Inverse Gaussian), as well as, for their extension under the Marshall-Olkin family of distributions. The Bayesian estimates of the distributions' parameters, as well as their associated credible intervals, will be obtained from the samples drawn from their joint posterior distribution.

In addition, Bayesian estimates' behaviour will be evaluated and compared with the maximum likelihood estimates obtained by Rocha (2016) through simulation experiments. Finally, we will apply the competing models and approaches to real datasets and we will compare them through various statistical measures. This work will be the first attempt to explore the advantages of the Bayesian approach to inference for defective cure rate models under the assumption of right censoring mechanism, as well as the first presentation of new Bayesian estimates for several defective distributions, but without incorporating covariate information.

Keywords: Defective distributions, Cure fraction, Bayesian Inference, Maximum likelihood, Right censoring, Survival Analysis, Gompertz distribution, Inverse Gaussian distribution, Marshall-Olkin family.

Περίληψη

Η ανάλυση επιβίωσης αποτελείται από ένα σύνολο στατιστικών μεθόδων που στοχεύει στη μελέτη του χρόνου μέχρι την εμφάνιση ενός συγκεκριμένου γεγονότος όπως ο θάνατος. Για την πλειονότητα των μεθόδων αυτών, θεωρείται πως όλα τα άτομα που συμμετέχουν υπόκεινται στο γεγονός που μας ενδιαφέρει. Ωστόσο, υπάρχουν περιπτώσεις όπου η υπόθεση αυτή δεν είναι ρεαλιστική, καθώς υπάρχουν ασθενείς που δεν θα βιώσουν το γεγονός αυτό στη διάρκεια της μελέτης. Για αυτό το λόγο, έχουν αναπτυχθεί ορισμένα μοντέλα επιβίωσης που επιτρέπουν την ύπαρξη ασθενών οι οποίοι δε βιώνουν το συμβάν και ονομάζονται μακροχρόνια επιζώντες. Τα μοντέλα αυτά ονομάζονται μοντέλα ρυθμού θεραπείας και υποθέτουν ότι, καθώς ο χρόνος αυξάνεται, η συνάρτηση επιβίωσης τείνει σε μια τιμή $p \in (0, 1)$, που αντιπροσωπεύει το ποσοστό των μακροχρόνια επιζώντων, αντί να τείνει στο μηδέν όπως στην κλασική ανάλυση επιβίωσης.

Πρόσφατα, ο Rocha (2016) πρότεινε μία νέα προσέγγιση των προβλημάτων επιβίωσης με μακροχρόνια επιζώντες. Η μεθοδολογία του για τη μοντελοποίηση του ποσοστού των μακροχρόνια επιζώντων βασίστηκε στη χρήση των «ελαττωματικών» (defective) κατανομών, οι οποίες χαρακτηρίζονται από το γεγονός ότι το ολοκλήρωμα της συνάρτησης πιθανότητάς τους δεν ισούται με τη μονάδα για ορισμένες επιλογές του πεδίου ορισμού κάποιων παραμέτρων τους. Σκοπός της παρούσας διπλωματικής εργασίας, είναι να παράσχει νέους Μπεϋζιανούς εκτιμητές των παραμέτρων των «ελαττωματικών» μοντέλων κάτω από την υπόθεση της δεξιάς λογοκρισίας. Επίσης, θα αναπτυχθούν αλγόριθμοι Markov chain Monte Carlo (MCMC) για τη συμπερασματολογία σχετικά με τις παραμέτρους μιας ευρείας κατηγορίας μοντέλων ρυθμού θεραπείας βασισμένων στις «ελαττωματικές» αυτές κατανομές, ενώ οι Μπεϋζιανοί εκτιμητές και τα αντίστοιχα διαστήματα αξιοπιστίας θα ληφθούν από τα δείγματα της από κοινού εκ των υστέρων κατανομής. Επιπλέον,

η συμπεριφορά των Μπεϋζιανών εκτιμητών θα αξιολογηθεί και θα συγκριθεί με αυτή των εκτιμητών μεγίστης πιθανοφάνειας του Rocha (2016) μέσω πειραμάτων προσομοίωσης. Ακόμη, τα προτεινόμενα αυτά μοντέλα-κατανομές θα εφαρμοσθούν σε πραγματικά σετ δεδομένων, όπου και θα συγκριθούν μεταξύ τους μέσω κατάλληλων στατιστικών μεγεθών. Τέλος, αξίζει να σημειωθεί πως η παρούσα διπλωματική εργασία αποτελεί την πρώτη προσπάθεια διερεύνησης των πλεονεκτημάτων της Μπεϋζιανής προσέγγισης στη συμπερασματολογία για τις παραμέτρους αρκετών μοντέλων ρυθμού θεραπείας, κάτω από την υπόθεση της δεξιάς λογοκρισίας, καθώς και της απόκτησης νέων Μπεϋζιανών εκτιμητών, χωρίς όμως τη συμπερίληψη της πληροφορίας από συν μεταβλητές.

Λέξεις κλειδιά: Ελλατωματικές κατανομές, Μακροχρόνια επιζώντες, Μπεϋζιανή συμπερασματολογία, Μέγιστη πιθανοφάνεια, Δεξιά λογοκρισία, Ανάλυση Επιβίωσης, Κατανομή Gompertz, Κατανομή Inverse Gaussian, Οικογένεια κατανομών Marshall-Olkin.

Acknowledgements

First of all, I would like to thank my supervisor Dr. Loukia Meligkotsidou of the Department of Mathematics at National and Kapodistrian University of Athens. The door to Dr. Meligkotsidou's office was always open whenever I ran into a trouble spot or had a question about my research or writing. She consistently allowed this thesis to be my own work, but steered me in the right direction whenever I needed it. Except for the academic help that she provided to me, I would also like to thank her for the psychological support whenever my anxiety exceeded the normal limits. I hope that our cooperation will be just as beautiful on my way to the PhD.

I would also like to acknowledge Dr. Elena Riza of the Medical School at National and Kapodistrian University of Athens and Dr. Constantin T Yannoutsos of the Department of Biostatistics at R.M Fairbanks School of Public Health in Indianapolis, as the other two members of my committee, and I am gratefully indebted to them for their very valuable comments on the present thesis.

Finally, I must express my very profound gratitude to my parents, my sister and all of my friends, for providing me with unfailing support and continuous encouragement throughout the process of researching and writing this thesis. This accomplishment would not have been possible without them. Thank you.

List of Tables

6.1 **Maximum likelihood estimates for the parameters of the defective distributions in the melanoma data set.** 89

6.2 **Maximum likelihood estimates for the parameters of the defective distributions in the colon cancer data set.** 92

6.3 **Bayesian estimates for the parameters of the defective distributions in the melanoma data set. 95% C.I represents the 95% credible interval for the parameters.** 95

6.4 **Bayesian estimates for the parameters of the defective distributions in the colon cancer data set.** 97

List of Figures

4.1 Survival function of the defective Gompertz distribution, for different values of the parameter α and with constant β. 58

4.2 Survival function of the defective Inverse Gaussian distribution, for different values of the parameter α and constant β. 59

4.3 Mean squared errors, biases and coverage probabilities of the maximum likelihood estimates $(\hat{\alpha}, \hat{\beta}, \hat{\pi})$ versus the sample size n, from the Gompertz distribution with $(\alpha, \beta, \pi) = (-1, 1, 0.3679)$. The dashed line represents the average mean squared error, bias and coverage probability, over all simulated sample sizes. 65

4.4 Mean squared errors, biases and coverage probabilities of the maximum likelihood estimates $(\hat{\alpha}, \hat{\beta}, \hat{\pi})$ versus the sample size n, from the Inverse Gaussian distribution with $(\alpha, \beta, \pi) = (-0.2, 1, 0.3297)$. The dashed line represents the average mean squared error, bias and coverage probability, over all simulated sample sizes. . . . 65

4.5 Mean squared errors, biases and coverage probabilities of the Bayesian estimates $(\hat{\alpha}, \hat{\beta}, \hat{\pi})$ versus the sample size n, from the Gompertz distribution with $(\alpha, \beta, \pi) = (-1, 1, 0.3679)$. The dashed line represents the average mean squared error, bias and coverage probability, over all simulated sample sizes. 68

4.6 Mean squared errors, biases and coverage probabilities of the Bayesian estimates $(\hat{\alpha}, \hat{\beta}, \hat{\pi})$ versus the sample size n, from the Inverse Gaussian distribution with $(\alpha, \beta, \pi) = (-0.2, 1, 0.32968)$. The dashed line represents the average mean squared error, bias and coverage probability, over all simulated sample sizes. . . . 69

5.1 Survival function of the Marshall-Olkin Gompertz distribution for different parameters' values. 73

5.2 Survival function of the Marshall-Olkin Inverse Gaussian distribution for different parameters' values. 74

5.3 Mean squared errors, biases and coverage probabilities of the maximum likelihood estimates $(\hat{\alpha}, \hat{\beta}, \hat{r}, \hat{\pi})$ versus the sample size n, from the Marshall-Olkin Gompertz distribution with $(\alpha, \beta, r, \pi) = (-1, 4, 2, 0.0360)$. The dashed line represents the average mean squared error, bias and coverage probability, over all simulated sample sizes. 76

5.4 Mean squared errors, biases and coverage probabilities of the maximum likelihood estimates $(\hat{\alpha}, \hat{\beta}, \hat{r}, \hat{\pi})$ versus the sample size n, from the Marshall-Olkin Inverse Gaussian distribution with $(\alpha, \beta, r, \pi) = (-2, 10, 2, 0.4958)$. The dashed line represents the average mean squared error, bias and coverage probability, over all simulated sample sizes. 77

5.5 Mean squared errors, biases and coverage probabilities of the Bayesian estimates $(\hat{\alpha}, \hat{\beta}, \hat{r}, \hat{\pi})$ versus the sample size n, from the Marshall-Olkin Gompertz distribution with $(\alpha, \beta, r, \pi) = (-1, 1, 0.5, 0.2254)$. The dashed line represents the average mean squared error, bias and coverage probability, over all simulated sample sizes. 80

5.6 Mean squared errors, biases and coverage probabilities of the Bayesian estimates $(\hat{\alpha}, \hat{\beta}, \hat{r}, \hat{\pi})$ versus the sample size n, from the Marshall-Olkin Inverse Gaussian distribution with $(\alpha, \beta, r, \pi) = (-1, 1, 0.5, 0.7616)$. The dashed line represents the average mean squared error, bias and coverage probability, over all simulated sample sizes. 81

6.1 Overall survival probability for the melanoma data set, as well as, separately according to the participant gender and treatment group. 85

6.2 Overall survival probability for the colon cancer data set, as well as, separately according to several participant characteristics. 87

6.3 Fit of the defective distributions to the melanoma data set, based on the maximum likelihood estimates of their parameters. 91

6.4 Fit of the defective distributions to the colon cancer data set, based on the maximum likelihood estimates of their parameters. 94

6.5 Fit of the defective distributions to the melanoma data set, based on the Bayesian estimates of their parameters. 96

6.6 Fit of the defective distributions to the colon cancer data set, based on the Bayesian estimates of their parameters. 98

Contents

Abstract **2**

Περίληψη **3**

Acknowledgements **5**

List of Tables **6**

List of Figures **7**

1 General Principles of Survival Analysis **14**

1.1 Introduction 14

1.2 Basic Functions in Survival Analysis 15

1.3 Censoring and truncation 17

1.3.1 Censoring 18

1.3.2 Truncation 19

1.4 Likelihood construction 20

1.4.1 Right censoring 20

1.4.2 General censoring 22

1.4.3 Left truncation 22

1.5 Modelling approaches in Survival Analysis 23

1.5.1 Parametric models 23

1.5.2 Semi-Parametric models 25

1.5.3 Non-parametric models 26

1.5.4 Frailty models 26

2 General Principles of Bayesian Inference 27

2.1 Introduction . . . 27

2.1.1 Bayes Theorem . . . 27

2.1.2 Bayesian versus Maximum likelihood estimation . . . 28

2.1.3 Inference . . . 29

2.2 The choice of prior distribution . . . 30

2.2.1 Conjugate Priors . . . 30

2.2.2 Absence of prior information . . . 31

2.3 Credibility Intervals . . . 31

2.4 Markov Chain Monte Carlo Methods . . . 32

2.4.1 The Gibbs sampling algorithm . . . 32

2.4.2 The Metropolis-Hastings algorithm . . . 34

2.4.3 Assessing the convergence of the Markov chain . . . 35

3 Cure Rate Models 38

3.1 Introduction . . . 38

3.2 Mixture Cure model . . . 39

3.3 Proportional Hazards Cure Rate Model . . . 42

3.4 Proportional Odds Cure Rate Model . . . 43

3.5 Transformation Cure Rate Models . . . 44

3.6 Further progress on Cure Rate Modelling . . . 47

4 Defective distributions for cure rate modelling 55

4.1 Introduction & Background . . . 55

4.2 The baseline defective distributions . . . 56

4.3 Setup of the simulation studies . . . 59

4.3.1 Artificial data generation . . . 59

4.3.2 Maximum likelihood estimation . . . 60

4.3.3 Bayesian estimation . . . 61

4.4 Maximum likelihood inference . . . 63

4.4.1 Simulation studies . . . 64

4.5 Bayesian Inference . . . 66
4.5.1 Simulation studies . . . 67
4.6 Comparative results . . . 70

5 Defective distributions under the Marshall-Olkin family 71
5.1 Introduction & Background . . . 71
5.2 The extended defective distributions . . . 72
5.3 Maximum likelihood Inference . . . 75
5.3.1 Simulation studies . . . 75
5.4 Bayesian Inference . . . 77
5.4.1 Simulation studies . . . 80

6 Application to real data sets 83
6.1 Introduction . . . 83
6.2 Description of the data sets . . . 84
6.2.1 Melanoma data set . . . 84
6.2.2 Colon cancer data set . . . 86
6.3 Maximum likelihood estimates: Applications . . . 88
6.3.1 Melanoma data set . . . 89
6.3.2 Colon cancer data set . . . 91
6.4 Bayesian estimates: Applications . . . 94
6.4.1 Melanoma data set . . . 95
6.4.2 Colon cancer data set . . . 97

7 Conclusions 99

Bibliography 102

Chapter 1

General Principles of Survival Analysis

1.1 Introduction

Survival analysis consists of statistical techniques whose aim is to describe and quantify time to event data. These types of datasets arise when some subjects are being followed for a long time period (e.g. years) or even a shorter time interval (e.g. days/ months) under controlled conditions, in order to study whether an event of interest happens or not. In medical research, such an event could be death from a disease that we are interested in, such as breast or lung cancer. For that reason, survival data are also being referred to as time-to-event data or failure-time data. We should mention here that the term failure is used to refer to the occurrence of the event of interest which could be also a success such as a recovery from a therapy or a surgery.

In the majority of studies whose aim is to collect survival data, there is missing or partial information about either the initiating or the terminating event or both, which should be taken into account in the whole analysis procedure. For example, in retrospective studies, ascertainment of the initiating event may not be possible, whereas in prospective studies the terminating event may not be observed for some subjects of the study. Finally, in cross-sectional studies with follow up, both of the above mentioned cases can happen. Such complications fall into two

general categories, which are called censoring and truncation, respectively. Censoring is generally reserved for a situation where only partial information on some subjects under study is available, while truncation refers to cases where some subjects in the population have no chance to be recruited to the study. Hence, all of this information should be included in a study with time-to-event data. These two notions, as well as other very basic notions met in survival analysis, will be presented more in detail later in this chapter. Before the basic aspects of censoring and truncation are introduced, we should also mention some basic aspects of survival analysis, such as death density, survival and hazard functions.

1.2 Basic Functions in Survival Analysis

First of all, we shall assume that T is a continuous random variable unless it is specified otherwise. The probability density function (pdf) and cumulative distribution function (cdf) are most commonly used to characterize the distribution of any random variable, and we shall denote these by $f(.)$ and $F(.)$, respectively with $F(t) = P(T \leq t)$.

However, since T is a non-negative random variable and it usually denotes the elapsed time until an event, it is commonly characterized by some other functions, such as the survival function, the hazard function and the cumulative hazard function.

Survival function $S(t)$

The survival function is a function that gives the probability that a patient who participates in our study will survive beyond any given specific time. For example, if T denotes the time until death, then $S(t)$ denotes the probability of the patient to survive beyond time t. From this definition, we can understand that there is a relationship between the survival function and the cumulative distribution function which is:

$$S(t) = 1 - F(t) = Pr(T > t). \tag{1.1}$$

Besides, the different values that T can take have a probability distribution with underlying probability density function f(t). Then the cumulative distribution function is denoted as

$$F(t) = Pr(T \leq t) = \int_0^t f(u)du \tag{1.2}$$

and according to the above equation, the survival function equals to

$$S(t) = 1 - \int_0^t f(u)du = \int_t^\infty f(u)du. \tag{1.3}$$

Hazard function $h(t)$

Another useful function to characterize the distribution of the random variable T, is the hazard function, which is defined as:

$$h(t) = \lim_{dt \to 0} \frac{Pr[t \leq T < t + dt | T \geq t]}{dt}. \tag{1.4}$$

The numerator of this expression is the conditional probability that the event will occur in the interval $[t, t + dt)$ given that it has not occurred before, and the denominator is the width of the interval. Graphically, the hazard function can have several forms. The cases most studied is where the hazard function is increasing, decreasing, constant, uni-modal and bathtub shaped. Checking the hazard behaviour is important when someone has to choose between parametric models. Finally, based on the relationship that follows, it is seen that there is an one-to-one correspondence between $S(t)$ and $h(t)$, which usually makes the algebraic manipulations easier,

$$S(t) = \exp[-\int_0^t h(u)du]. \tag{1.5}$$

Cause-specific hazard

The cause-specific hazard represents the instantaneous risk of dying of cause j and is given by the following relation:

$$\lambda_j(t, x) = \lim_{dt \to 0} \frac{Pr[t \leq T < t + dt, J = j | T \geq t, x]}{dt}.$$

In practice, we calculate the conditional probability that a subject with covariates x dies in the interval $[t, t + dt)$ and the cause of death is the j-th cause, given that the subject was alive just

before time t. Then the overall hazard rate is given by the following equation:

$$\lambda(t,x) = \sum_{j=1}^{m} \lambda_j(t,x),$$

since the overall hazard must be due to one of the m causes.

Cumulative Hazard function $H(t)$

We may think of $H(t)$ as the sum of the risks that we face going from time 0 to t. Specifically, the cumulative hazard function is defined as:

$$H(t) = \int_0^t h(u)du. \tag{1.6}$$

So given the hazard rate, we can always integrate to obtain the cumulative hazard and then exponentiate to obtain the survival function using Equation (1.5). Finally, it would be our omission if we did not mention the relationship between the survival function and the cumulative hazard function, which is:

$$S(t) = \exp[-H(t)]. \tag{1.7}$$

1.3 Censoring and truncation

In this section we will describe all the possible types of censoring and truncation which can occur in the analysis of survival data. Censoring and truncation are two well established features of time-to event data and there are various types for both of them. The knowledge about the version of censoring or truncation that exists in the dataset is very important because of its major impact on the way that the likelihood of the observed data is being computed. Each version leads to a different way of calculating the likelihood function, which is the basis for making statistical inference especially in the world of frequentist statisticians.

Except for that, censoring and truncation are connected directly with the presence of bias in the results of the analysis. Specifically, censoring should be non-informative, which means that participants who drop out of the study should do so due to reasons unrelated to the study. Informative censoring occurs when participants are lost to follow-up due to reasons related to

the study, e.g. in a study comparing disease-free survival after two treatments for cancer, the control arm may be ineffective, leading to more recurrences and patients becoming too sick to follow-up. We continue this section by introducing some of the basic aspects about censoring and truncation.

However, it should be clear that the concepts of censoring and truncation are totally different. More specifically, a censored observation is an incomplete observation containing only partial information about the event time, which means the patient is followed up for some time, but the event does not occur during this period. On the other hand, truncation which is a common problem in register data, means that subjects who fail before the date of registration are truncated and they are not included in the study population (Klein & Moeschberger (2005)).

1.3.1 Censoring

The possible censoring mechanisms which can be met in a data set are mentioned below.

Right Censoring

In general, right censoring means that the true event time happens after the observation is seized on a subject. Right censoring is perhaps the most common type of censoring and it has been extensively studied in the literature. There are several types of right censoring. Below we describe the most common types of them (Lagakos (1979)).

Type I right censoring: This type of censoring happens when the event occurs after some pre-specified time. Mathematically, let $T_1, T_2, \ldots, T_n$ be independent, identically distributed (i.i.d) random variables each with cumulative distribution function F. Assume also that t_c is a pre-assigned fixed censoring time. Instead of observing $T_1, T_2, \ldots, T_n$, the variables of interest, we can only observe $Y_1, Y_2, \ldots, Y_n$, where:

$$Y_i = T_i, \text{ if } T_i \leq t_c \text{ and } Y_i = t_c, \text{ if } T_i > t_c.$$

Type II right censoring: It happens when the starting time of the study is pre-determined

but the ending time depends on the time when the first r individuals experience the event of interest, where r is some pre-specified integer. Mathematically, let $r < n$ be fixed , and let $T_{(1)} < \ldots < T_{(n)}$, be the order statistics of $T_1, T_2, \ldots, T_n$. As we said observation ceases after the $r - th$ failure, so we can only observe $T_{(1)}, \ldots, T_{(r)}$.

Random right censoring: This type of censoring occurs when there are some other factors, except for the event of interest, which could remove some of the individuals from a trial during the study period, such as various competing events.

Left Censoring

A lifetime X is considered to be left-censored if it is less than a censoring time C_{left}, which means that the event which we are interested in has already occurred for that individual before that person is observed in the study at time C_{left}. The data from a left-censored sampling mechanism can be represented by pairs of random variables (T, ϵ), where $T = max(\text{X}, C_{left})$ and ϵ indicates whether the exact lifetime is observed ($\epsilon = 1$) or not ($\epsilon = 0$). Finally, we should note that this type of censoring is less often than the right censoring sampling mechanism.

Interval Censoring

A more general type of censoring occurs when the lifetime is only known to occur within an interval. Accordingly, we only know that the true event time is greater than the last observation time at which the change has not occurred and less than or equal to the first observation time at which the change has been observed to occur, thus giving an interval which contains the real (but unobserved) time of occurrence of the change. Finally, we note that the interval censoring contains right censoring when $R_i = \infty$ and left censoring when $L_i = 0$, where R represents the right endpoint and L the left one.

1.3.2 Truncation

The confusion of the term truncation with that of censoring is very common in the field of statistical analysis, however as mentioned above, these two terms are totally different. Truncation occurs due to the process generating the data which is such that it is only possible to

observe outcomes above (or below) a truncation limit. Moreover, truncation occurs when only those individuals whose event time lies within a certain observational window (Y_{left} , Y_{right}) are observed. If the event time of an individual does not lie in this interval, then the investigator has no information about that subject. This is in contrast to censoring where there is at least partial information for each participant. Because we are only aware of individuals with event times in the observational window, the inference for truncated data is restricted to conditional estimation (Steelman (2015)).

Left Truncation

When $Y_{right} = \infty$ we talk about left truncation. In this case, we only observe lifetime X if and only if $Y_{left} < X$. In this type of truncation any subjects who experience the event of interest prior to the truncation time are not observed.

Right Truncation

It occurs when Y_{left} equals zero. That is, we observe the survival time X only when $X \leq Y_{right}$. More specifically, right truncation occurs when only subjects who have experienced the event of interest are included in the sample.

1.4 Likelihood construction

In this section we obtain the likelihood function of survival times under different censoring and truncation mechanisms.

1.4.1 Right censoring

Let $T_1^*, T_2^*, \ldots, T_n^*$ be i.i.d survival times with cumulative distribution function F and let $C_1, C_2, \ldots, C_n$ be i.i.d censoring times with cumulative distribution function G. Also, let f and g be the probability density functions with respect to F and G. We are only able to observe the bivariate data $(T_1, \delta_1), (T_2, \delta_2), \ldots, (T_n, \delta_n)$, where $T_i = min(T_i^*, C_i)$ denotes the observation time for every

$i = 1, 2, \ldots, n$ and

$$\delta_i = \begin{cases} 1, & \text{if } T_i^* \leq C_i \text{ observed failure-event,} \\ 0, & \text{if } T_i^* > C_i \text{ observed censoring.} \end{cases} \tag{1.8}$$

From the above notation of the observed pairs, we note that each pair consists of a continuous component (T_i) and a binary component (δ_i). For that reason the pair (T_i, δ_i) can take the two following forms:

$$(T_i, \delta_i) = \begin{cases} (t_i, 1), & \text{if } T_i \text{ is uncensored at time } t_i, \\ (t_i, 0), & \text{if } T_i \text{ is censored at time } t_i. \end{cases} \tag{1.9}$$

Let K be the distribution function of the observation time $T = min(T^*, C)$. Then the following relation holds: $K(t) = Pr(min(T^*, C) \leq t) = 1 - Pr(min(T^*, C) > t) = 1 - Pr(T^* > t, C > t)$. Assuming independence between the event time T^* and censoring time C is very crucial, since it implies the following simplification: $K(t) = 1 - Pr(T^* > t)Pr(C > t) = 1 - (1 - F(t))(1 - G(t))$. Based on the above, the probability density function of the survival data (T, δ) is:

$$f(t, \delta) = (f(t)(1 - G(t)))^{\delta}(g(t)(1 - F(t)))^{1-\delta}; \text{ Wienke (2010).}$$

In addition to the independence between the censoring and event times we also assume non informative censoring. That means that the censoring distribution must not depend on the same parameter as the event distribution. For that reason, the terms $g(t)$ and $G(t)$ in the previous relation become constants with respect to the parameter of interest. As a consequence, the contribution of right censored survival data $(t_i, \delta_i), i = 1, 2, \ldots, n$, to the likelihood function is the following:

$$L_i((t_i, \delta_i)|\theta) \propto [f(t_i|\theta)]^{\delta_i}[S(t_i|\theta)]^{1-\delta_i}.$$

If any of the two assumptions changes, then the final relation does not hold and some transformations should take place.

1.4.2 General censoring

Consider at this point that we could have any type of the censoring mechanisms which were mentioned above (Wienke (2010)). Then the likelihood function would be as follows:

$$L((t_i,\delta)|\theta) = \prod_{d\in D} f(t_d) \times \prod_{r\in R} S(t_r) \times \prod_{l\in L}[1-S(t_l)] \times \prod_{i\in I}[S(U_i)-S(V_i)],$$

where D is the set of death times, R is the set of right censored times, L is the set of left censored observations and I is the set of interval censored observations with the only knowledge that the real survival time T_i is in the interval $[U_i, V_i]$. Furthermore we should note that $S(U_i) - S(V_i) = Pr[U_i \leq T_i \leq V_i]$, which is the probability that the real survival time $T_i \in [U_i, V_i]$.

1.4.3 Left truncation

As mentioned above, sometimes we may have datasets in which there are left truncated data. In such cases, suppose that the real survival time T_i is left truncated at some time point Y_i (Wienke (2010)). Then we have to consider the conditional distribution of T_i given that $T_i \geq Y_i$:

$$g(t|T_i \geq Y_i) = \frac{f(t)}{Pr(T_i \geq Y_i)} = \frac{f(t)}{S(Y_i)}.$$

So the probability to observe an event at t_d is proportional to:

$$g(t_d|T_d \geq Y_d) = \frac{f(t_d)}{S(Y_d)}.$$

Furthermore, the probability that the real survival time T_r is right censored at t_r is:

$$Pr[T_r \geq t_r|T_r \geq Y_r] = \frac{S(t_r)}{S(Y_r)}.$$

The probability that the real survival time T_l is left censored at t_l is:

$$Pr[T_l \leq t_l|T_l \geq Y_l] = \frac{S(Y_l) - S(t_l)}{S(Y_l)},$$

and the probability that the real survival time $T_i \in [U_i, V_i]$ where $U_i \geq Y_i$ is:

$$Pr(U_i \leq T_i \leq V_i|T_i \geq Y_i) = Pr(T_i \geq U_i|T_i \geq Y_i) - Pr(T_i \geq V_i|T_i \geq Y_i) = \frac{S(U_i) - S(V_i)}{S(V_i)}$$

So in this case, the likelihood function is given by the following formula:

$$\begin{aligned} L((t,\delta)|\theta) &= \prod_{d\in D} \frac{f(t_d)}{S(Y_d)} \times \prod_{r\in R} \frac{S(t_r)}{S(Y_r)} \times \prod_{l\in L} \frac{S(Y_l) - S(t_l)}{S(Y_l)} \times \prod_{i\in I} \frac{S(U_i) - S(V_i)}{S(Y_i)} \\ &= \frac{\prod_{d\in D} f(t_d) \times \prod_{r\in R} S(t_r) \times \prod_{l\in L}[S(Y_l) - S(t_l)] \times \prod_{i\in I}[S(U_i) - S(V_i)}{\prod_{i=1}^{n} S(Y_i)} \end{aligned}$$

Likelihood construction for right truncated data

In the case of right truncation the probability to observe a death at Y_i conditional on that the

survival time T_i is less than or equal to Y_i is proportional to: $\frac{f(Y_i)}{1-S(Y_i)}$. So the likelihood function is: $L(t^*, \delta|\theta) = \prod_{i=1}^{n} \frac{f(Y_i)}{1-S(Y_i)}$.

1.5 Modelling approaches in Survival Analysis

There are basically three ways to model time-to-event data. The parametric, the semi-parametric and the non-parametric techniques. If the distributional assumption on the survival times is valid, the parametric models result in more efficient estimates for the parameters in the sense of having smaller standard errors as compared to those obtained by the non-parametric models and the interpretation of the results is easier (Ibrahim et al. (2013)).

1.5.1 Parametric models

The parametric models most commonly found in the literature are the following:

1. **The exponential distribution:** The simplest choice of parametric model is the Exponential distribution. The characteristic of this distribution is the fact that it assumes a constant hazard over time, which reflects its property called lack of memory. However, the specific property makes the exponential model a poor choice for modelling human survival and for that reason other parametric models have been developed.

2. **The standard Weibull distribution:** Another very common choice of parametric model is the one in which the standard Weibull distribution (2-parameter Weibull distribution) is being used. The Weibull model was introduced by Weibull (1939) and is the most popular generalization of the exponential model with two positive parameters. It constitutes a better choice compared to the Exponential one, due to the fact that it has the ability to assume the characteristics of many different types of distributions. However, this distribution is inappropriate when the hazard rate is indicated to be uni-modal or bathtub-shaped and for that reason, a generalization was proposed by Mudholkar et al. (1996) in order to be able to include such kind of shapes.

3. **The Gamma distribution:** The Gamma distribution constitutes an extension of the Exponential distribution, yet it is of limited use in survival analysis, since it does not

have closed form for the survival and the hazard function because both of them include the incomplete gamma integral: $I_k(x) = \frac{\int_0^x s^{k-1}\exp(-s)ds}{\Gamma(k)}$. In addition, it should be mentioned that the great advantage of this distribution is the fact that for different values of the parameters, it can model different hazard forms, such as constant, monotonically increasing or monotonically decreasing.

4. **The Gompertz distribution:** Benjamin Gompertz (1825) proposed the Gompertz distribution which is widely used especially in actuarial, biological and demographic applications. The Gompertz distribution, just as the previously mentioned parametric models, captures several hazard forms, such as increasing and decreasing forms for different parameter values. In addition, by adding one more parameter to the 2-parameter Gompertz distribution, it is generalized to the Gompertz-Makenhaam distribution. Finally, a very important characteristic of this distribution is the fact that when the domain of its parameters changes, then it becomes defective. More specifically, a distribution is called defective if the integral of its density function does not result in unity, but in a value $p \in (0, 1)$. Then, it can model situations in which there are individuals who never experience the event of interest, known as long-term survivors. The specific version of the Gompertz distribution will be investigated in more detail to the next chapters, under the maximum likelihood and the Bayesian approach to inference.

5. **The log-Logistic distribution:** An alternative modelling approach is the use of the log-Logistic distribution. The log-Logistic distribution has a fairly flexible functional form and therefore it is one of the parametric survival models in which the hazard rate may be decreasing, increasing, as well as hump-shaped.

6. **The log-Normal distribution:** The last most commonly used parametric model is the log-Normal distribution. A key feature of the distribution and its appropriateness as a model for survival data, is that the hazard function is non-monotone. The hazard starts at zero, rises rapidly to a peak and then falls off gradually. Therefore, this model should be considered only when such behaviour makes biological sense. Except for that, the log-Normal distribution yields non-proportional hazards, which is often encountered in real

life problems (Royston; 2001).

1.5.2 Semi-Parametric models

Throughout the present subsection the following notation will be used. For each study subject i, T_i denotes the survival time and C_i the censoring time. Also, by $Z_i(t)$ are denoted the covariates which may depend on time. Due to the censoring effect, the observable data are $[Y_i, \delta_i, Z_i(.), i = 1, ..., n]$ where $Y_i = min(T_i, C_i)$ and $\delta_i = I(T_i \leq C_i)$.

1. **The Cox proportional hazards model:** As it is known, the most popular semi-parametric model is the Cox proportional hazard model introduced by Cox (1972). The basic assumption of this survival model is the proportionality of the hazards and since some times this assumption is very strict, various models have been proposed to overcome the problem of non-proportionality.

2. **The proportional odds model:** A model that constitutes an alternative choice when the hazards' proportionality does not hold is the proportional odds model introduced by Bernett (1983). The difference between this model and the Cox model is the fact that the hazard ratio between two sets of covariate values converges to unity rather than staying constant as time passes. This property is very desirable especially when the differences between stages of a disease at diagnosis tend to diminish with time or even when the initial effects of a treatment tend to diminish with time. It should be noted though that in the case of this model the estimation procedure is not as easy as in the Cox model and for more detail we refer to the work done by Dabrowska & Doksum (1988), Cheng et.al (1995), Murphy et.al (1997) and Shen (1998).

3. **The additive hazards model:** Another approach is the additive hazards model, which has been studied by Cox & Oakes (1984) as well as by Lin & Ying (1994; 1995). The specific model has been introduced especially for dealing with the cases when the effects of the covariates are time dependent. In this case, thr Cox model is difficult to fit, since smoothing parameters need to be chosen.

4. **The accelerated failure time model:** A very useful model for the analysis of survival

data is the accelerated failure time model (Zeng & Lin (2007)), an alternative to the Cox proportional hazards model since it provides a natural formulation of the effects of covariates on potentially censored response variable and is in many ways more appealing because of its direct interpretation, especially in cases when the response variable does not pertain to survival time.

1.5.3 Non-parametric models

As it is widely known, the standard non-parametric estimator of the survival function is the Kaplan-Meier (K-M) estimator, also known as the product-limit estimator, introduced by Kaplan & Meier (1958). The KME is a step function estimator of the survival distribution $S(t)$ which takes into account the fact that the observed survival times may be censored. To define the KME, suppose that we have a sample which consists of the observed survival times $t_1, t_2, \ldots, t_n$. The effect of censoring is to limit the length of survival time. So for censored observations we know only that $t_i^* \geq t_i$, however for uncensored observations we have that $t_i^* = t_i$.

1.5.4 Frailty models

The notion of frailty provides a convenient way to introduce random effects, association and unobserved heterogeneity into models for survival data. This term was introduced by Vaupel et al. (1979) in univariate survival models and was substantially promoted by its application to multivariate survival data (Clayton (1978)) on chronic disease incidence in families. In its simplest form, a frailty is an unobserved random proportionality factor that modifies the hazard function of an individual. Normally, in most clinical applications, survival analysis implicitly assumes a homogeneous population to be studied. This means that all individuals sampled into that study are subject, in principle, to the same risk. However, in many applications the study population must be considered as a heterogeneous sample, since it is impossible to measure all relevant covariates related to the disease of interest. Therefore, the frailty approach is a statistical modelling concept which aims to account for heterogeneity, caused by unmeasured covariates.

Chapter 2

General Principles of Bayesian Inference

2.1 Introduction

As it is widely known, in statistical inference there are two broad schools of inference, classical inference and Bayesian inference. The second one, on which the present thesis emphasizes, is a method of statistical inference in which Bayes theorem is used to update the probability for a hypothesis as more evidence or information becomes available and constitutes an important technique in mathematical statistics finding application in a wide range of fields, such as engineering, sports and medicine. More specifically, Bayesian inference derives the posterior probability as a combination of two elements: the prior probability and the likelihood function which is derived from a statistical model concerning the observed data.

2.1.1 Bayes Theorem

Let A and B be the two possible outcomes of a given situation and in addition assume that $A = A_1 \cup A_2 \cup \ldots \cup A_n$, with $A_i \cap A_j = \emptyset$ for every $i \neq j$. Then, the Bayes theorem provides an expression for the conditional probability of A_i given B, which is equal to:

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)},$$

while more usually we write:

$$P(A_i|B) \propto P(B|A_i)P(A_i).$$

The last equation which is based on the proportionality of the probabilities is also called the Bayes rule introduced by Laplace, stating the proportionality of the two parts by considering that the $P(B)$ in the denominator of the first equation being constant (Hoffmann & Jorgensen; 1994).

2.1.2 Bayesian versus Maximum likelihood estimation

The two schools of statistical inference have several major differences. The principal difference between the two statistical approaches is the fact that the Bayesian approach does not consider the parameters as constants but as random variables characterized by prior distributions. Therefore, Bayesian inference does not take into account only one value of the parameter, but a whole distribution. It is known that the frequentist statisticians used to treat the Bayesians as a minority until the late 80s. This happened since the Bayesian approach needed computer power to reach its real potential, while maximum likelihood did not.

However, when Markov Chain Monte Carlo techniques, which will be mentioned below, were introduced in the field of statistics, as well as the informatics made progress, Bayesian statistics started getting close to what they could truly offer to the statistical society, becoming a valuable tool to every researcher. Let's see the principal ideas of each inferential approach:

Frequentist-Maximum likelihood approach

- The parameters of the population are unknown fixed constants.
- Statistical procedures have a long-term meaning, like an infinite repetition of the same experiment.
- Probabilities are interpreted as a frequency after a long number of experiments.

Bayesian approach:

- The parameters are considered as random variables, as we are not certain of their real values.
- The way to make inference is just the use of the rules of probabilities.
- Each person has his own way of thinking, so the prior beliefs naturally vary across people.
- There can be a continuous update of our beliefs as data come to our hand.

So it is seen that, the Bayesian approach, especially based on the last two ideas, gets more related to real life situations and a more sensible and natural way of quantifying problems.

2.1.3 Inference

Let us consider a random sample $y = (y_1, \ldots, y_n)$, with $f(y_i|\theta)$ as the distribution function which describes the random variables $y_1, \ldots, y_n$. Therefore, the likelihood function is given by $f(y|\theta) = \prod_{i=1}^{n} f(y_i|\theta)$ which sets the probability of observing y_i under different values of the parameter. As already mentioned, Bayes' Theorem incorporates the information already gathered, our prior beliefs for the parameter, represented by one or more prior distributions, then takes into account the observed data and makes inference. For the case where θ is continuous, the following equation is obtained representing the posterior distribution:

$$\pi(\theta|y) = \frac{f(y|\theta)f(\theta)}{\int f(y|\theta)f(\theta)d\theta},$$

where $\pi(\theta)$ is the prior distribution, $\int f(y|\theta)f(\theta)d\theta = f(y)$ is the marginal likelihood function and $f(y|\theta)$ is the likelihood of the data given the parameter θ. In the case when the parameter θ is discrete, the previous relationship changes and takes the following form:

$$\pi(\theta|y) = \frac{f(y|\theta)\pi(\theta)}{\sum_{\theta\in\Theta} f(y|\theta)\pi(\theta)},$$

where $\sum_{\theta\in\Theta} f(y|\theta)\pi(\theta) = f(y)$ is again the marginal likelihood of the observed data.

By controlling the prior distribution we can express either the fact that we are very certain about our beliefs by setting the variance of the distribution to a low value, or our prior ignorance by placing a large variance. There are various techniques for specifying our prior ignorance and the researcher should take into account the cases in which the placing of an uninformative

prior is improper (i.e. does not integrate to 1) and therefore will cause computation troubles, especially if the concluding posterior is simultaneously an improper one.

2.2 The choice of prior distribution

One of the most basic issues that we are called to face is the choice of the appropriate parameter's prior distribution. For this issue the following points should be noted:

- It should be understood that the data analysis under the framework of Bayesian inference is more subjective, since the prior distribution represents our prior beliefs about the parameter. Therefore, it follows that the analysis is unique for each statistical analyst, since someone else's prior would lead to a different posterior analysis.

- We should also have in mind that as long as the prior distribution does not seem unreasonable for a specific problem, then its effect becomes less influential as the number of available data increases.

- When we do not have a clear idea about the form of the prior distribution, it is suggested to use a more convenient form which is consistent with our rough prior beliefs, but also makes the computation more easily handled.

- Finally, in cases when we feel that we do not have a specific prior information about the parameter, we should choose a form that depicts this ignorance.

2.2.1 Conjugate Priors

In the context of Bayesian theory, when the posterior distribution of a parameter $\pi(\theta|x)$ is in the same probability distribution family as its prior $\pi(\theta)$, then the prior and the posterior distributions are called conjugate and the prior is called conjugate prior for the specific likelihood function $L(x|\theta)$. It should be made clear that, a conjugate prior is an algebraic convenience, giving a closed-form expression for the posterior, otherwise numerical integration techniques may be necessary. However, the conjugate family should not be used without limits only for the convenience that offers in the statistical analysis.

2.2.2 Absence of prior information

As already mentioned, when we do not have any specific information about the parameter, we usually choose a prior distribution which reflects this ignorance. One way to do this, is to choose a prior distribution from the conjugate family but with a large enough variance. This would mean the absence of information concerning the concentration of the parameter values around a certain value.

Another choice of less-informative prior distribution is the Uniform distribution. The fundamental problem by using the uniform distribution as our prior is the fact that the uniform distribution is not invariant under re parametrization, as well as if the parameter space is infinite the uniform prior is improper which means it does not integrate to one. This is however not always a serious problem since improper prior distributions often lead to proper posterior distributions.

2.3 Credibility Intervals

The credibility intervals constitute the analogous of confidence intervals that we meet in the framework of classical statistics. However, there is a fundamental difference between these two notions. In Classical inference, a 95% confidence interval can not be interpreted as the interval in which the parameter lies with probability 95% since the parameter θ is considered to be constant, whereas in the framework of Bayesian inference, this is the interpretation of a 95% credibility interval since the parameter is considered to be random. A 95% credibility interval satisfies the following equation:

$$C_\alpha(x) = [\theta : \pi(\theta|x) \geq \gamma] \tag{2.1}$$

where γ is considered to be such that: $\int_{C_\alpha(x)} \pi(\theta|x)d\theta = 1 - \alpha$, with such regions being called *highest posterior density regions.*

2.4 Markov Chain Monte Carlo Methods

A fairly important reason that Bayesian analysis has grown rapidly in recent years, is the development of Markov Chain Monte Carlo (MCMC) methodology. These methods allow us to overcome problems of great computational difficulty that could not be handled before. They also allow us to be more realistic in our modelling approaches, as it is possible to infer more complex models. In particular, the MCMC methods are simulation techniques that give us samples from the joint posterior distribution of the parameters based on the creation of a Markov chain that after a large number of steps converges to that distribution. In particular, a Markov chain is a sequence of discrete or continuous random variables θ with the property that, given the present outcome, the past and future outcomes are independent:

$$p(\theta^{k+1} = y|\theta^k = x, \theta^{k-1} = x_{k-1}, \ldots, \theta^0 = x_0) = p(\theta^{k+1} = y|\theta^k = x).$$

In the following subsections, we will briefly describe two of the most well known MCMC methods, the Gibbs sampler and the Metropolis Hastings algorithm.

2.4.1 The Gibbs sampling algorithm

The Gibbs sampler is a Markov chain Monte Carlo (MCMC) algorithm for obtaining a sequence of observations which are approximated from a specified multivariate probability distribution, when direct sampling is difficult. This sequence can be used to approximate the joint distribution; to approximate the marginal distribution of one of the variables, or some subset of the variables; or to compute an integral (such as the expected value of one of the variables). Typically, some of the variables correspond to observations whose values are known, and hence do not need to be sampled. It is a randomized algorithm and is an alternative to deterministic algorithms for statistical inference such as the expectation-maximization algorithm (EM).

As with other MCMC algorithms, Gibbs sampling generates a Markov chain of samples, each of which is correlated with nearby samples. As a result, care must be taken if independent samples are desired. In general, samples from the beginning of the chain (the burn-in period) may not accurately represent the desired distribution and are usually discarded. In practice, we assume

that after k iterations, the chain has reached its target distribution and we can throw away the early portion and use the remaining draws for posterior inference. The value of k is the burn-in length of the burn-in period.

As far as the selection of the burn-in period is concerned, it is made through some tests which are known as convergence diagnostics. One of the most widely known convergence diagnostics is the Gelman and Rubin test, proposed by Gelman & Rubin (1992) which compares the variances between the chains. Also, Geweke (1992) proposed the comparison of the means calculated from distinct segments of one Markov chain, as well as Raftery & Lewis (1992) introduced a test based on the estimation of the minimum chain length needed in estimate a percentile to some precision. Except for these, one possible remedy is thinning the resulting chain of samples if needed. It has been shown, however, that using a longer chain instead leads to better estimates of the true posterior. Thus, thinning should only be applied when time or computer memory are restricted.

Mathematical Notation

Suppose that a sample X is taken from a distribution depending on a parameter vector $\theta \in \Theta$ of length d with prior distribution $g(\theta_1, \ldots, \theta_d)$. The Gibbs algorithm is being implemented through the following steps:

1. We start with $\theta = (\theta_1^0, \ldots, \theta_d^0)$, generated by the prior distribution of the parameter.
2. We simulate $\theta_1^{(1)}$ by the conditional posterior distribution $f(\theta_1|x, \theta_2^0, \ldots, \theta_d^0)$.
3. We simulate $\theta_2^{(1)}$ by the conditional posterior distribution $f(\theta_1|x, \theta_1^{(1)}, \theta_3^0, \ldots, \theta_d^0)$.
4. ...
5. We simulate $\theta_d^{(1)}$ by the conditional posterior distribution $f(\theta_1|x, \theta_1^{(1)}, \theta_2^{(1)}, \theta_3^{(1)}, \ldots, \theta_{d-1}^{(1)})$.
6. Repeat the above procedure times the number of samples we would like to generate.

The convergence of the above Markov chain in the posterior distribution $f(\theta_1, \theta_2, \ldots, \theta_d|x)$ is guaranteed. The above procedure is completed after a large number of iterations, but by first

deducting the first samples we get from the burn-in period, which are not "realistic" samples of the joint posterior distribution.

2.4.2 The Metropolis-Hastings algorithm

The Metropolis Hastings algorithm is an MCMC method for obtaining a sequence of random samples from a probability distribution for which direct sampling is difficult. This sequence can be used to approximate the distribution, or to compute an integral, such as an expected value. Metropolis Hastings is generally used for sampling from multi-dimensional distributions, especially when the number of dimensions is high.

The fundamental difference between this algorithm and the Gibbs Sampling Algorithm lies on the fact that Gibbs can be used only when the conditional posterior distributions of the unknown parameters are recognizable. However, in many cases the conditional posterior distributions are not written in closed form, which results in not being able to use the Gibbs sampler. Instead, the Metropolis-Hastings algorithm is one of the most common choices in such cases.

Mathematical Notation

Before looking at the specific algorithm in more detail, it is important to look at the steps followed by generalized simulation algorithms.

1. We divide the unknown parameters into d sets $\theta_1, \ldots, \theta_d$ where each set has dimension ≥ 1.

2. Start with $(\theta_1^{(0)}, \theta_2^{(0)}, \ldots, \theta_d^{(0)})$.

As in the case of the Gibbs sampler, we remove the initial samples that we will get from the procedure and the rest can be considered to be from the distribution required.
Let us assume that we have reached the j^{th} iteration with the values $\theta_1^{(j)}, \theta_2^{(j)}, \ldots, \theta_d^{(j)}$ and we want to simulate the value $\theta_1^{(j+1)}$, the next value of θ_1. The updating mechanism of the Metropolis-Hastings algorithm is as follows:

1. We propose a candidate value θ_1^{can} which is from a random distribution with a function of density $q(\theta_1^{can}|\theta_1^{(j)}, \ldots, \theta_d^{(j)})$.

2. We choose as the next value of θ_1 in the Markovian chain the $\theta_1^{(j+1)}$ where

$$\theta_1^{(j+1)} = \begin{cases} \theta_1^{can}, & \text{with probability p} \\ \theta_1^{(j)}, & \text{with probability 1-p} \end{cases} \tag{2.2}$$

where

$$p = \min[1, \frac{f(\theta_1^{can}|x, \theta_2^{(j)}, \ldots, \theta_d^{(j)})}{f(\theta_1^{(j)}|x, \theta_2^{(j)}, \ldots, \theta_d^{(j)})} \frac{q(\theta_1^{(j)}|\theta(can)_1, \theta_2^{(j)}, \ldots, \theta_d^{(j)})}{q(\theta_1^{can}|\theta_1^{(j)}, \theta_2^{(j)}, \ldots, \theta_d^{(j)})}] \tag{2.3}$$

and $f(\theta_1^{can}|x, \theta_2^{(j)}, \ldots, \theta_d^{(j)})$ is the conditional probability density function of θ_1 calculated for $\theta_1 = \theta_1^{can}$ and respectively for $f(\theta_1^{j}|x, \theta_2^{(j)}, \ldots, \theta_d^{(j)})$.

As it can be seen, the Gibbs sampler is a special sub-case of Metropolis-Hastings algorithm where $q(\theta_1^{can}|\theta_1^{(j)}, \theta_2^{(j)}, \ldots, \theta_d^{(j)}) = f(\theta_1^{can}|x, \theta_2^{(j)}, \ldots, \theta_d^{(j)})$.

2.4.3 Assessing the convergence of the Markov chain

As it was mentioned above, MCMC techniques generate a Markov chain that ultimately provides a sample from the posterior distribution and the summary measures calculated from this chain consistently estimate the corresponding true posterior summary measures. In addition, it should be noted that when the probabilities mentioned before do not depend on k, the Markov chain is called homogeneous.

The fundamental question resulting from the sampling procedure is what happens with the Markov chain when k goes to infinity. One can show that if the generated chain has a limiting distribution π, then the distribution is also stationary which means that further elements of the chain also have π as distribution. Further, we speak for a reversible Markov chain if the rate at which the Markov chain moves from x to y is the same as the rate at which it moves from y to x, which is the detailed balance condition. Furthermore, we say that a Markov chain satisfies the ergodicity criteria if it satisfies the following three criteria:

1. **Irreducibility**: the chain can reach each possible outcome whatever the starting position;

2. **Aperiodicity**: there is no cyclic behaviour in the chain and

3. **Positive Recurrence**: the chain visits every possible outcome an infinite number of times and the expected time to return to a particular outcome, irrespective of where we start in the chain, is finite.

In practical terms, ergodicity means that the chain will explore the posterior distribution exhaustively. In the following theorems the most important results are given concerning the convergence of the Markov chains generated by Gibbs sampler or Metropolis-Hastings.

Graphical approaches to assess convergence

Trace Plot: A simple exploration of the trace plot gives a first and insightful impression of the characteristics of the Markov chain. Trace plots are produced for each parameter separately and evaluate the chain univariately, but it is also useful to monitor the Markov chains jointly, i.e. the total parameter vector θ.

Autocorrelation Plot: When future positions in the chain are highly predictable from the current position, then the posterior is slowly explored and one says that the chain has a low mixing rate. The mixing rate is measured by autocorrelations of different lags. The autocorrelation of lag m, denoted as ρ_m, is defined as the correlation between θ^k and θ^{k+m} (for $k = 1, 2, \ldots$) and can be simply estimated by the Pearson correlation or a time series approach.When the autocorrelation decreases only slowly with increasing lag, the mixing rate is low. Note that the autocorrelation plot can also indicate the minimum number of iterations for the chain to ignore from its starting position. Once the Markov chain converged, the ACF plot does not change anymore, irrespective of the magnitude of the autocorrelations. When all autocorrelations are close to zero then MCMC sampling is done in an almost independent manner and stationarity will be attained quickly.

Running mean plot: Upon stationarity at k_0, the mean (and all other characteristics) of $\pi_k(\theta)$ shows stability for $k > k_0$. The running mean or ergodic mean plot can display this stability. It is a time-series plot of the running mean θ_k, i.e. the mean of all sampled values up to and including iteration k.The initial variability of the running -mean plot is always relatively

high (even when sampling from the correct posterior), but stabilizes with increasing k in case of stationarity.

Q-Q plot: Another graphical tool that was proposed is a Q-Q plot with the first half of the chain on the x-axis and the second half of the chain on the y-axis. A Q-Q plot deviating from the bisecting line is an indication of non-stationarity of the chain.Finally,

Cross-correlation plot: The correlation between θ_1^k with θ_2^k $(k = 1, \ldots, n)$ is called the cross-correlation of θ_1 with θ_2. The scatterplot of θ_1^k versus θ_2^k produces a cross-correlation plot. This plot is useful in case of convergence problems to indicate if model parameters are strongly related and thus is a diagnostic for an over specified model.

Chapter 3

Cure Rate Models

3.1 Introduction

In the field of survival analysis, a strict and implicit assumption is that all subjects participating in the study will eventually experience the event of interest if the follow-up time is sufficiently long enough. However, this assumption may not be true in many cases. For instance, in some cancer research studies there may be a certain portion of patients who respond favourably to treatment and appear to be risk free after a sufficient follow-up time. Such patients are called cured of the disease. In other words, only a proportion of patients from the population are susceptible to the event of interest and other subjects are not susceptible to the event. The proportion of the cured subjects in the population is called the **cure fraction** or **cure rate**. A long and stable plateau at the tail of the Kaplan-Meier survival curves is a clear indication concerning the existence of a cure fraction in the dataset.

As it is widely known, the model for the survival data with a cure fraction is called the cure rate model or simply cure model and the model to analyse the survival data without a cure fraction is referred to as non-cure model. Before introducing the cure models, the notation of the survival data with a cure fraction should be first given. Survival data with a cure fraction are similar to the survival data without a cure fraction, as defined in the first chapter, yet, the true failure time Y of the subjects can be ∞. However, the case where $Y = \infty$ is not observable

due to the existence of right censoring and therefore the survival data with a cure fraction looks identical to the survival data without a cure fraction.

In the context of cure rate modelling, the survival function $S(t)$, the hazard function $\lambda(t)$ or $h(t)$, the cumulative hazard function $\Lambda(t)$ or the probability density function $f(t)$ are said to be improper functions, in the sense that $\lim_{t\to\infty} S(t) > 0$, $\lim_{t\to\infty} \Lambda(t) < \infty$ and $\int_0^\infty f(t)dt < 1$. As far as the other properties of the specific quantities are concerned, they remain the same as the one in the survival models without a cure fraction. Thus in these situations we still have that $S(t)$ of cure model is monotone decreasing with $S(0) = 1$, $h(t)$ is non-negative and $\Lambda(t) = \int_0^t h(t)dt$ is monotone non-decreasing, as well as $f(t)$ is always non-negative.

3.2 Mixture Cure model

The Mixture Cure model, introduced by Boag (1949), is a popular method for analysing time-to-event data in which some of the participating subjects are believed to be cured in a reasonable way. More specifically, this model assumes that a proportion (π) will be cured and these subjects are not at risk of experiencing the occurrence of the event of interest. The remaining proportion $(1-\pi)$ is for the rest of the participants who are expected to experience the event some time in the future. Under those assumptions, the mixture cure model takes the following form:

$$S(t|z,x) = \pi(z) + (1-\pi(z))S_0(t|x)$$

where $S(t|z,x)$ is the improper population survival function, $\pi(z)$ is the cure rate fraction, $S_0(t|x)$ is the proper survival function for the non cured participants, z is the covariate associated with $\pi(z)$ and x is the covariate associated with $S_0(t|x)$, where z and x may share some common elements.

As far as the part for non cured individuals is concerned, several different parametric distributions have been used, including the Exponential distribution (Ghitany et.al. 1994), the Weibull distribution (Farewell, 1986), as well as, the Log Normal distribution (Boag, 1949). Except for the choice of parametric distributions, many non parametric approaches have also been considered in the literature so far (Taylor, 1995; Sy & Taylor, 2000).

In addition, a generalized link function of logistic distribution was suggested in order to access the effects of covariates for the cured probability ($\pi(z)$), while parametric functions were suggested for fitting the $S_0(t|x)$. More specifically, Farewell (1982) used logistic regression for modelling the cured probability and a Weibull distribution for fitting the survival function for the non cured participants. That means that the cured probability was modelled as follows:

$$\log(\frac{\pi(z)}{1-\pi(z)}) = \beta' z,$$

where $\beta' s$ are the logistic regression's coefficients.

Except for that, Maller and Zhou (1996) provided a very comprehensive treatment of the cure model based on different parametric failure time regression models, as well as, they investigated one-sample non-parametric failure time models. Also, Zhang & Peng (2009) considered separate modelling of covariate effects on the cure probability and the distribution of the survival time for the non cured participants. It should be noted though, that unlike the traditional mixture cure rate models, the last one of Zhang & Peng (2009), allowed the covariate effects on the failure time distribution of the non cured patients to be negligible at time zero and to grow gradually with time. This property of the specific modelling is very useful especially in several cancer treatments when their effect gradually increases from zero.

Furthermore, there were several semi and non parametric methodologies proposed for the estimation of the survival function when the distribution of the survival times was not specified. More specifically, Tsodikov et.al. (2003) provided a very useful summary of the non parametric model proposed by Maller & Zhou (1996) who approached a homogeneous sample. Also, Kuk & Chen (1992) proposed a semi-parametric approach of the cure rate model, in which the survival times were estimated using a proportional hazards regression model while the cure probability was determined by the logistic regression model and they also developed a Monte Carlo approximation in order to estimate the model's parameters. The specific model (proportional hazards regression model) was further studied and developed by Peng & Dear (2000), Sy & Taylor (2000), as well as, Lam et.al. (2005) among various authors, aiming to develop a new

methodological approach for calculating the joint parametric-non-parametric likelihood function. Finally it should be noted that the approaches mentioned above are basically based on the EM (Expectation-Maximization) algorithm which focuses on the computation of the parametric as well as the non-parametric components.

Certainly there are many studies that have been conducted within this particular model. For instance, Kim & Jhun (2008) suggested using the Mixture cure rate model for interval censored data. More specifically, they developed the likelihood function based on an approximate approach which was proposed by Goetghebeur & Ryan (2000) and they introduced a frailty model to characterize the association between the cure probability and the survival time. Furthermore, Kim et.al. (2009) suggested a new cure rate model by incorporating latent cure rate markers, which were modelled via a multinomial logistic regression and the individuals who were sharing the same cure rate were classified into the same risk group.

In addition, Seppa et. al. (2010) proposed a new mixture cure rate model with random effects to cause-specific survival data concerning female breast cancer patients. More specifically, they applied two different sets of random effects in order to capture the regional variation in the cure probability and in the survival of the non cured patients. Another very useful contribution in the specific scientific field, was made by Peng & Taylor (2010) who also studied the mixture cure rate model with random effects, yet they obtained the maximum likelihood estimates of the model for clustered survival data with a cure fraction, based on several estimation methods, such as rejection sampling and importance sampling.

Finally, another scientific contribution to the field that deserves to be noted is that made by Ma (2010). In that specific work, Ma presented a semi-parametric cure rate model for mixed case interval-censored data (Note: Mixed case interval-censored data arise when the event time of interest is only known to lie in an interval obtained from a sequence of k random examinations, where k is a random integer), in which a generalized linear model was used to describe the probability of cure and a Cox model was applied for the estimation of the non cured part.

3.3 Proportional Hazards Cure Rate Model

Another important cure model is the **Proportional Hazards Cure model**, which is given as follows:

$$S(t|z,x) = \exp\left(-\theta(z)F_0(t|x)\right),$$

where $S(t|z,x)$ is the improper population survival function, $F_0(t|x)$ is a proper baseline cumulative distribution function and $\theta(z)$ is a positive function of z which is usually formulated as $\exp(\beta' z)$. Then, the cure rate of this model is given by $S(\infty|z,x) = \exp(-\theta(z))$.

The proportional hazards cure model was first proposed by Yakovlev et al. (1993) and then studied by various researchers. Chen et al. (1999) studied this model in the Bayesian framework. Also, Tsodikov (1998b, 2001) studied the model by assuming that $F_0(t)$ is non-parametrically specified, Tsodikov (2003) and Tsodikov and Garibotti (2007) studied the model again by assuming a Cox proportional hazards model for $F_0(t)$ with unspecified baseline hazard function. Tsodikov (1998a) studied this model when the covariate z is time-dependent and finally, Chen et al. (2002) studied the application of this model to multivariate survival data.

The advantage of the proportional hazards cure model over the mixture cure model is that it has a proportional hazards structure, which is preferable for Bayesian inference, and a biological interpretation, which is explained as follows: For a given subject, let N be the number of tumor cells that is capable of metastasizing after treatment and $(X_1, \ldots, X_N)$ be the survival times of each cell. Then the failure time of the subject is the $min(X_1, \ldots, X_N)$. Usually, N is assumed to follow a Poisson distribution with mean θ and $(X_1, \ldots, X_N)$ are assumed to be independent and identically distributed with a common cumulative distribution function $F(t)$ independent of N. Then the survival function is given by the following relation:

$$\begin{aligned} S(t) &= Pr(N=0) + Pr(N>1) \times Pr(X_1 > t, \ldots, X_N > t | N > 1) = \\ &= \exp(-\theta) + \sum_{j=1}^{\infty} \frac{(1-F(t))^j \theta^j \exp(-\theta)}{j!} = \exp(-\theta F(t)) \end{aligned}$$

Because of its biologic derivation, the specific model is also known as the **Promotion Time Cure model** in the literature. As far as the estimation approaches are concerned, Ibrahim et.al.

(2001) discussed various parametric and semi-parametric approaches. In addition, Tsodikov et.al. (2003) provided a very nice review of these non-mixture cure modelling techniques in cure rate estimation and various statistical problems associated with them. They have also highlighted one more advantage over the classical mixture cure model, which is the fact that it is very attractive for computations, since it has a simple structure for the survival function which can provide a natural technical structure for maximum likelihood estimation techniques.

Except for the scientific works mentioned above, the non mixture cure rate model was also studied by Herring & Ibrahim (2002) who introduced a parametric estimation approach by incorporating random effects. Furthermore, they proposed a methodology to account for non-ignorable missing covariates in the framework of the specific model. In addition, Brown & Ibrahim (2003) extended the non mixture cure rate model to include longitudinal covariates. Uddin et. al. (2006a, b) proposed two different approaches (non-parametric and parametric) for the estimation of the cure rate, based on the specific model, yet they based their conclusion on uncensored data. Furthermore, Liu & Chen (2009) introduced a semi-parametric non mixture cure rate model for the analysis of interval censored data by introducing a semi-parametric maximum likelihood estimation procedure for the model using the EM algorithm. Finally, Lopesa & Bolfarine (2012) investigated the specific model with random effects and they estimated the parameters by classical and Bayesian methods.

3.4 Proportional Odds Cure Rate Model

The third cure model, which is less popular than the previous two though, is the **Proportional Odds Cure model**, which is defined analogous to the proportional odds model for survival data without a cure fraction. The survival function of the model is given as shown below:

$$S(t|z,x) = \frac{1}{1+\exp(\beta' z)F_0(t|x)},$$

where $S(t|z,x)$ is the improper population survival function and $F_0(t|x)$ is a proper cumulative distribution function. The cure rate of the proportional odds cure model is $S(\infty|z,x) = \frac{1}{1+\exp(\beta' z)}$. Tsodikov (2003) studied the model when $F_0(t)$ is non-parametrically specified and Zeng et al. (2006) studied the model as a special case of their transformation cure model. Also,

Gu et al. (2011) provided a biological derivation of the model and studied it parametrically.

3.5 Transformation Cure Rate Models

First of all, a little notation about the transformation models should be given. Semi-parametric **transformation models** have attracted much interest in the last two decades. In the transformation model, a family of transformation functions is imposed on the failure time, the hazard function or the survival function. The transformation function is usually parametrically specified. When the transformation function changes within the family, the transformation model generates a class of survival models, including some well-known survival models as its special cases. There are many examples of transformation survival models. For example, Ciampi et al. (1989) used the Box-Cox transformation for survival data generation. However, the transformation models considered in this section are imposed on the survival function instead of the failure time.

The first transformation model applied on the survival function was proposed by Cheng et al. (1995) and it has the following form:

$$g(S(t|z)) = h_0(t) + \beta' z,$$

where $g(.)$ is the transformation function, $S(t|z)$ is the population survival function and $h_0(t)$ is the baseline hazard function. When $g(.) = \log(-\log(.))$, the model becomes the Cox proportional hazards model, whereas when $g(.) = -logit(.)$, it becomes the proportional odds model. Therefore, this model nests the proportional hazards model and the proportional odds model as its special cases. Since the origination of the model, the transformation models have been widely studied by many researchers. A nice summary of the transformation models could be found in Zeng & Lin (2007), where the authors presented several classes of semi-parametric transformation models, as well as, various corresponding estimating methods.

After listing some basic information on these models, we are now in a position to proceed to the reference to the transformation cure models which have also been proposed in the literature. Lu & Ying (2004) proposed a transformation cure model in the mixture cure pattern with

the following form:

$$\begin{cases} S(t|z) = \pi(z) + (1 - \pi(z))S_0(t|z), \\ S_0(t|z) = \exp(-\Lambda(H_0(t) + \beta' z)), \end{cases}$$

where $S(t|z)$ is the improper population survival function, $\pi(z)$ is the cure rate, $S_0(t|z)$ is the improper survival function of the non cured subjects which follows a similar transformation model as the model mentioned above, $H_0(t)$ is an unspecified monotone increasing function and $\Lambda(x)$ is a specified transformation function. When $\Lambda(x) = \exp(x)$, $S_0(t|z)$ follows the Cox proportional hazards model with unspecified hazard function, whereas when $\Lambda(x) = \log(\frac{\exp(x)}{1+\exp(x)})$, $S_0(t|z)$ follows the proportional odds model.

Similarly to the previous transformation cure model, Mao & Wang (2010) proposed another transformation cure model by applying a parametric transformation model for $S_0(t|z)$ which takes the following form:

$$\begin{cases} S(t|z) = \pi(z) + (1 - \pi(z))S_0(t|z), \\ S_0(t|z) = \dfrac{\exp(-\frac{\beta' z}{\rho})}{(\exp(-\frac{\beta' z}{\rho}) + \rho H_0(t))^{\frac{1}{\rho}}}, \end{cases}$$

where $H_0(t)$ is a proper, but unspecified, baseline cumulative hazard function. The proper survival function for non cured patients $S_0(t|z)$ follows a transformation model, which is called the generalized proportional odds model proposed by Dabrowska & Doksum (1988). When the transformation parameter $\rho = 1$ in the previous model, $S_0(t|z)$ follows the proportional odds model. However, when $\rho \to 0$, $S_0(t|z)$ follows the proportional hazards model.

It can be seen that both models mentioned above are based on the mixture cure model and the transformation model is applied only on the proper baseline survival function. Transformation cure models which are not based on the mixture cure model have also been proposed in the literature. More specifically, Yin and Ibrahim (2005a) proposed a transformation cure model,

by applying the Box-Cox Transformation model on the survival function as follows:

$$S(t|z) = (1 - \frac{\alpha \exp(\beta' z)}{1 + \alpha \exp(\beta' z)} F_0(t))^{\frac{1}{\alpha}},$$

where $S(t|z)$ is the improper population survival function and $F_0(t)$ is a proper cumulative distribution function. Also, α is the transformation parameter. The cure rate of this model is $S(\infty|z) = (1 + \alpha \exp(\beta' z))^{-\frac{1}{\alpha}}$. When $\alpha = 1$, the model becomes a mixture cure model with the following form:

$$S(t|z) = \frac{1}{1 + \exp(\beta' z)} + \frac{\exp(\beta' z)}{1 + \exp(\beta' z)} S_0(t|z),$$

where $S_0(t) = 1 - F_0(t)$. On the other hand, when $\alpha \to 0$, the model becomes a proportional hazards cure model with the following form:

$$S(t|z) = \exp(-\exp(\beta' z) F_0(t)).$$

In addition, when $0 < \alpha < 1$, the model becomes an intermediate model between the mixture cure model and the proportional hazards cure model. It should be noted that the specific model is still a valid cure model when $\alpha > 1$.

Another transformation cure rate model has also been proposed by Zeng et al. (2006) which does not follow the mixture cure pattern and has the following form:

$$S(t|z) = [1 + \alpha \theta(z) F_0(t)]^{-\frac{1}{\alpha}},$$

where $S(t|z)$ is the improper population survival function and $F_0(t)$ is a proper cumulative distribution function. α is the transformation parameter and the cure rate is $S(\infty|z) = [1 + \alpha \theta(z)]^{-\frac{1}{\alpha}}$. As before the following cases should be noted as well:

When $\alpha = 1$, it results in the proportional odds cure model:

$$S(t|z) = 1 + \theta(z) F_0(t).$$

When $\alpha \to 0$, it results in the proportional hazards cure model:

$$S(t|z) = \exp(\theta(z) F_0(t)).$$

Finally, even when $0 < \alpha < 1$, the specific model produces a new intermediate cure model between the proportional odds cure model and the proportional hazards cure model and when $\alpha > 1$, the model is still a valid cure rate model.

3.6 Further progress on Cure Rate Modelling

Research in the field of cure rate modelling does not stop in the classic model choices mentioned above. Many researchers have dealt with this issue and have extended much of the relevant statistical knowledge. This specific section will cover some of the scientific work done on cure rate modelling under the frequentist, as well as, under the Bayesian perspective, in recent years.

Pal & Balakrishnan (2016) assumed that the number of competing causes follow an exponentially weighted Poisson distribution. More specifically, let M be a random variable denoting the initial number of competing causes related to the occurrence of an event of interest. Then, M is assumed to follow an exponentially weighted Poisson distribution with parameter $\eta > 0$ and weight function $\exp(\phi m)$, where ϕ is a real number and $m = 0, 1, 2, \ldots$. The probability mass function of M is given by the following relation:

$$Pr[M = m|\eta, \phi] = \exp(-\eta \exp(\phi)) \frac{(\eta \exp(\phi))^m}{m!}.$$

The researchers, under the assumption that the population of interest has a cure fraction, developed an EM algorithm in order to determine the maximum likelihood estimates of the model parameters. In addition, this model is characterized as more flexible than the promotion time cure model mentioned above, and also it provides an interesting and realistic interpretation of the biological mechanism of the occurrence of an event of interest.

Gallardo et al. (2017) introduced a new cure rate model based on a new distribution, called the Yule-Simon distribution, for modelling the concurrent events. They studied some properties of this distribution and the model arising when the distribution of the competing causes is the Weibull model. They called this distribution the Weibull-Yule-Simon distribution, which assumes the following forms:

$$S_{WYS}(t|\rho, \lambda) = \rho(\rho + 1) \int_0^1 \frac{u^\rho du}{1 - (1 - u)\exp(-\exp(\alpha)t^\nu)} - \rho,\ t > 0 \text{ and}$$

$$f_{WYS}(t|\rho, \lambda) = \rho(\rho + 1)\nu t^{\nu-1} \exp(\alpha - \exp(\alpha)t^\nu) \int_0^1 \frac{u^\rho (1 - u) du}{[1 - (1 - u)\exp(-\exp(\alpha)t^\nu)]^2}$$

Under those assumptions, the authors conducted the maximum likelihood estimation procedure

for the model parameters, as well as, they conducted a small scale simulation study indicating satisfactory parameter recovery by the estimation approach. Finally, the results on which they concluded were applied to the melanoma data set illustrating the fact that this model can outperform traditional alternative models in terms of model fitting.

Borges (2017) developed a regression model for survival data in the presence of long-term survivors based on the generalized Gompertz distribution introduced by El -Gohary et al. (2013) in a defective version. The specific model also includes as a special case the Gompertz cure rate model proposed by Gieser et al.(1998). Finally, Borges (2017) developed an EM algorithm for determining the maximum likelihood estimates (MLEs) of the parameters of the model and discussed the construction of confidence intervals for the parameters using the asymptotic distribution of the MLEs and the parametric bootstrap method as well. More specifically, the regression model proposed had the following properties:

$$f(t|\lambda,\alpha,\theta) = \lambda\theta \exp(\alpha t - \frac{\lambda}{\alpha}(\exp(\alpha t - 1)))(1 - \exp(-\frac{\lambda}{\alpha}(\exp(\alpha t - 1))))^{\theta-1},$$

$$S(t|\lambda,\alpha,\theta) = 1 - (1 - \exp(-\frac{\lambda}{\alpha}(\exp(\alpha t - 1))))^{\theta}.$$

Finally, it should be noted that the author chose to use this modelling approach since the defective distributions have the potential to model data with a cure fraction and they also have the great advantage that the proportion of cured is always estimated using a distribution with one parameter less than the standard mixture model, which brings plenty of benefits in terms of estimation.

Koutras & Milienos (2017) introduced a flexible family of cure rate models, motivated by the biological derivation of the classical promotion time cure rate model and assuming that a metastasis-competent tumor cell produces a detectable -tumor mass only when a specific number of distinct biological factors affect the cell. The proposed model has as special cases, among others, the promotion time (proportional hazards), the geometric (proportional odds), and the negative binomial cure rate model. In addition, their model generalizes specific families of transformation cure rate models and some well-studied destructive cure rate models. Ex-

act likelihood inference was carried out by the aid of the EM algorithm, as well as, a profile likelihood approach was exploited for estimating the parameters of the model while the model discrimination problem was treated by the aid of the likelihood ratio test. More specifically the authors proposed the following model:

$$S_p(t) = \begin{cases} \exp\left(\frac{\alpha}{(1-\alpha)\delta}[1-(1+\delta\alpha^{-1}H(t))^{1-\alpha}]\right), \alpha \neq 1, \\ (1+\delta H(t))^{-\frac{1}{\delta}}, \alpha = 1, \end{cases}$$

where $\alpha, \delta \geq 0$ with $H(t)$ being a cumulative hazard function and the probability someone to be cured is non zero only when $\alpha > 1$.

Balakrishnan et al. (2017) assumed a Conway-Maxwell-Poisson distribution under a competing cause scenario and they studied a flexible cure rate model in which the lifetimes of non-cured individuals were described by a Cox proportional hazard model with a Weibull hazard as the baseline function. As far as inference is concerned, they developed an approach for right censored data by the maximum likelihood method with the use of the EM algorithm and a profile likelihood approach for the estimation of the dispersion parameter of the Conway-Maxwell-Poisson distribution.

More specifically, the Conway -Maxwell -Poisson distribution accommodates and generalizes some well known discrete distributions and it is a flexible family of distributions since it can be over-dispersed or under-dispersed depending on the value of the dispersion parameter. Therefore, if the number of competing causes M follows a Conway -Maxwell -Poisson distribution, its probability mass function is given by the following relation:

$$Pr[M = m|\eta, \phi] = \frac{1}{Z(\eta, \phi)} \times \frac{\eta^m}{(m!)^\phi},$$

where

$$Z(\eta, \phi) = \sum_{j=0}^{\infty} \frac{\eta^j}{(j!)^\phi},$$

with $\phi \geq 0$ and $\eta > 0$. If $\phi = 1$, M is an equi-dispersed Poisson random variable with $E(M) = \eta$, while if $\phi \to \infty$, M becomes an under-dispersed Bernoulli random variable with parameter $\frac{1}{1+\eta}$. Moreover, if $\phi = 0$ and $\eta < 1$, then M is an over-dispersed geometric random

variable with parameter $1-\eta$. Therefore, according to the value of ϕ we can have various different distributions (over, under or equi dispersed). Finally, the cure rate is given by the following relation:

$$p_0 = Pr[M = 0|\eta, \phi] = Z(\eta, \phi)^{-1}.$$

Shen (2015) proposed a semi-parametric non mixture cure model for the regression analysis of left-truncated and interval-censored data and developed semi parametric maximum likelihood estimation for the non mixture cure model after conducting a simulation study in order to investigate the performance of the proposed estimators.

Rodrigues & Castro (2011) developed a flexible cure rate survival model by assuming the number of competing causes of the event of interest to follow a compound weighted Poisson distribution. In fact, their model is more flexible in terms of dispersion than the promotion time cure model and it gives an interesting and realistic interpretation of the biological mechanism of the occurrence of event of interest as it includes a destructive process of the initial risk factors in a competitive scenario.

Rodrigues et al. (2015) published a paper whose purpose was to make the standard promotion cure rate model (Yakovlev & Tsodikov, 1996) more flexible by assuming that the number of altered cells after some a treatment follows a fractional Poisson distribution (Laskin, 2003). It was proved that the well-known Mittag-Leffler relaxation function (Berberan-Santos, 2005) was a simple way to obtain a new cure rate model which is a compromise between the promotion and geometric cure rate models allowing for super-dispersion. So, the relaxed cure rate model developed by the authors can be considered as a natural and less restrictive extension of the popular Poisson cure rate model at the cost of an additional parameter but a competitor to negative-binomial cure rate models (Rodrigues et al., 2009b).

Louzada (2015) extended the promotion cure rate model by incorporating excess of zeros in the modelling. Despite allowing to relate the covariates to the cure fraction, the specific ap-

proach, which is based on a biological interpretation of the causes that trigger the event of interest, did not enable to relate the covariates to the fraction of zeros.

Of course there has been a wealth of research activity not only under the frequentist approach but also under the Bayesian perspective. Below, some of the typical tasks that have been done under the Bayesian perspective will be presented. Rodrigues et al. (2010) proposed a new Bayesian flexible cure rate survival model, which generalised the stochastic model of Klebanov et al. (1993) and had much in common with the destructive model formulated by Rodrigues et al. (2009). For the development of the Bayesian inference for the proposed model the authors used Markov Chain Monte Carlo (MCMC) methods.

Yin & Nieto-Barajaw (2009) proposed a class of Bayesian cure rate models by incorporating a baseline density function as well as multiplicative and additive covariate structures. Their model naturally accommodated zero and non-zero cure rates, which provided an objective way to examine the existence of a survival fraction in the failure time data. Within the Bayesian paradigm, they took a Markov gamma process prior modelling the baseline hazard rate, and mixture prior distributions for the parameters in the additive component of the model. Finally, they implemented a Markov chain Monte Carlo computational scheme to sample from the full conditional distributions of the posterior and they conducted simulation studies to assess the estimation and inference properties of the proposed model.

Souza et al. (2017) proposed a flexible probability distribution induced by a discrete frailty, and then presented some special discrete probability distributions. More specifically, they focused on a special hyper-Poisson distribution and then developed the corresponding Bayesian simulation, influence diagnostics by means of intensive Markov chain Monte Carlo algorithm.

Furthermore, Kim et al. (2007) proposed a new class of semi-parametric cure rate models and they constructed dynamic models for piecewise hazard functions over a finite partition of the time axis. Allowing the size of partition and the levels of baseline hazard to be random, their

proposed models provide a great flexibility in controlling the degree of parametricity in the right tail of the survival distribution and the amount of correlations among the log-baseline hazard levels. In addition, several properties of the proposed models were derived, and propriety of the implied posteriors with improper non informative prior distributions for regression coefficients based on the proposed models was established for the fixed partition of the time axis. Finally, they developed an efficient reversible jump computational algorithm for carrying out posterior computation.

Ortega et al. (2017) proposed a four-parameter extended fatigue lifetime model called the odd Birnbaum -Saunders geometric distribution, which extends the odd Birnbaum-Saunders and Birnbaum-Saunders distributions. The authors derived some properties of the new distribution that included expressions for the ordinary moments and quantile functions. In addition, they adopted the method of maximum likelihood and a Bayesian approach in order to estimate the model parameters. Also, they performed various simulations for different parameter settings and sample sizes and they proposed two new models with a cure rate called the odd Birnbaum-Saunders mixture and odd Birnbaum-Saunders geometric models by assuming that the number of competing causes for the event of interest has a geometric distribution. More specifically, they adopted proper prior distributions according to variations of the parametric space, but ensuring non-informativeness according to the fixed hyper-parameters that lead to such a situation.

Yiqi et al. (2016) developed a Bayesian approach for the Weibull-Negative-Binomial regression model with cure rate under latent failure causes and in the presence of randomized activation mechanisms. They assumed the number of competing causes of the event of interest followed a Negative Binomial distribution while the latent lifetimes were assumed to follow a Weibull distribution. In addition, the Bayesian procedure was developed through Markov chain Monte Carlos methods and the authors also discussed the model selection for the comparison of the fitted models. Moreover, they developed case deletion influence diagnostics for the joint posterior distribution based on the ψ-divergence, which has several divergence measures as particular cases.

Cantor and Shuster (1992) made a constructive discussion about parametric and non-parametric methods for estimating cure rates based on censored survival data. They used the Kaplan-Meier method for non-parametric estimation of the cure proportion. On the other hand for parametric estimation of cure rates, they assumed a survival function $S(t)$ for which $\lim_{t\to\infty} S(t) = S(\infty) > 0$ i.e., in a proportion of patients the event never occurs. Using MLE one can estimate $S(\infty)$, which is considered as the cure rate fraction. Two survival functions considered by the authors are:

$$S_1(t) = \pi + (1-\pi)\exp(-\lambda t),\ 0 < \pi < 1,\ \lambda > 0$$

$$S_2(t) = \exp[\beta^{-1}\alpha(1-\exp(\beta(t)))],\ \alpha > 0 \text{ and } \beta < 0$$

The first model represents the case where a proportion π of patients are cured, while the remaining $1-\pi$ have an exponential failure rate. The second one is a modified Gompertz survival distribution, for which $S_2(t)$ approaches $\exp(\frac{\alpha}{\beta})$ asymptotically. In addition, the first model which is based on the exponential distribution, has been extensively discussed by Goldman (1984), who studied the performance of maximum likelihood estimates of the parameters of $S_1(t)$ in the context of a Monte Carlo study, whose main aim was the study of the power, bias, variance and other characteristics of maximum likelihood estimates of the binomial-exponential model, by involving simulation of a large number of clinical studies with differing designs.

Also, the second model was developed by Gompertz (1825) and was motivated by observed population life tables. Garg et al. (1970) discussed the Gompertz distribution and the maximum likelihood estimation of its parameters. However, they did not consider the situation in which the hazard is decreasing, causing $S(\infty)$ to be positive. Gehan and Siddiqui (1973) presented a least squares procedure for estimating the parameters of the Gompertz distribution. Given a set of survival data, either of the above survival functions can be used as the basis of the likelihood function, whose logarithm can be maximized to find the estimates of the parameter. Estimates of the cure rates are given by $\hat{\pi}$ or by $\exp(\frac{\hat{\alpha}}{\hat{\beta}})$. However, the maximization of the likelihood function based on $S_1(t)$ creates problems and therefore Newton-type methods often

fail or require a large number of iterations.

Wu (2010) proposed a novel extension of the classical mixture cure rate model in order to incorporate the additional information about the status of cure. This modelling approach also showed that with the specific additional information, more efficient estimators could be obtained. Finally it should be noted that the author mentions that both proportional hazards cure rate models and accelerated failure times cure rate models can use this extension.

Chapter 4

Defective distributions for cure rate modelling

4.1 Introduction & Background

In the previous chapter we presented some of the most widely known models used in the field of cure rate modelling, as well as some of the most recent progress conducted in this field. Except for these approaches, Rocha (2016) introduced in his doctoral thesis, a new modelling approach for the cases where there are long term survivors. More specifically, he proposed the the use of a class of distributions which are called **defective**. The characteristic of these distributions is the fact that the integral of their density function does not equal to unity, but to a value $p \in (0, 1)$, which represents the proportion of long term survivors in the data set, known as the cure rate or cure fraction, when the domain of their parameters changes. By this definition, it is made clear that it is not necessary to assume by hand that there are indeed long-term survivors, as is the case of the traditional cure rate models, but the conclusion whether they exist or not arises through the process of estimating the parameters.

A great advantage of these distributions, is the fact that the proportion of cured participants is always estimated using a model with one parameter less than the respective mixture model, as well as the calculation of the cure rate becomes very easy, since it is calculated through a

simple function of the distributions' parameters. In more detail, Rocha proposed the use of the Gompertz and the Inverse Gaussian distribution, in their simple form shown below, as well as the Marshall-Olkin family of distributions, shown in the next chapter. These distributions are based on positive values of their parameters, however, when the parameter α takes negative values they can be used for cure rate modelling, since they can instantly assess the proportion of long term survivors in the data set. It is noted that, when the point estimate of the parameter α is negative, it means that there are cured patients in the data set. For the statistical significance of this existence, as proposed by Rocha, we should look at the confidence intervals of the estimates. When both edges of the confidence interval are negative, then the cure rate is statistically significant. All these distributions and their respective properties are presented in more detail in this chapter, as well as the next one.

The aim of the present chapter is to reproduce the maximum likelihood estimates for the parameters of the baseline functions, namely the defective Gompertz and the defective Inverse Gaussian distribution, under the right censoring mechanism. More specifically, we are going to obtain the maximum likelihood estimates with the same procedure presented in the PhD of Rocha (2016). Except for that, we are going to investigate these two baseline defective distributions under the Bayesian inferential procedure, in order to propose a new Bayesian approach to the defective Gompertz distribution, compared to the one published by MR dos Santos et.al. (2017), as well as to obtain for the first time Bayesian estimates for the parameters of the defective Inverse Gaussian distribution, under the right censoring assumption.

4.2 The baseline defective distributions

The Gompertz distribution is used for modelling survival data in various areas of knowledge, especially where there is a suspicion of exponential hazard and it takes the following form:

$$f(t|\alpha,\beta) = \beta \exp(\alpha t) \exp[-\frac{\beta}{\alpha}(\exp(\alpha t) - 1)],$$

where $\alpha > 0, \beta > 0$ and $t > 0$ as well. By using the specific parametrization, α is called

the shape parameter and β is called the scale parameter. In addition, its survival function takes the following form:

$$S(t|\alpha,\beta) = \exp[-\frac{\beta}{\alpha}(\exp(\alpha t) - 1)].$$

According to the literature, the Gompertz distribution takes its defective version when the shape parameter (α) takes values outside of its domain, which means that it takes negative values. Then, the cure fraction is given by the following relation:

$$p_0 = \lim_{t\to\infty} S(t|\alpha,\beta) = \lim_{t\to\infty} \exp[-\frac{\beta}{\alpha}(\exp(\alpha t) - 1)] = \exp[\frac{\beta}{\alpha}] \in (0,1)$$

In **Figure 4.1** we can see various different forms of the survival function based on different negative choices for the shape parameter, while the parameter β was kept constant and equal to unity. Smaller values of the shape parameter indicate a higher cure fraction. Therefore, the cure fraction estimated in each case is as follows:

$p_1 = 60.65\%$, for $\alpha = -2$

$p_2 = 71.65\%$, for $\alpha = -3$

$p_3 = 77.88\%$, for $\alpha = -4$ and

$p_4 = 81.87\%$, for $\alpha = -5$

In probability theory, the Inverse Gaussian distribution is a two-parameter family of continuous probability distributions with support on $(0,\infty)$. In the present thesis this particular distribution is going to be used as a model for cure rates, an idea introduced for the first time by Lee & Whitmore (2006). The probability density function of the Inverse Gaussian distribution takes the following form:

$$f\left(t|\alpha,\beta\right) = \frac{1}{\sqrt{2\beta\pi t^3}} \exp\left[-\frac{1}{2\beta t}(1-\alpha t)^2\right],$$

for $\alpha > 0$, $\beta > 0$ and $t > 0$. In addition, its survival function is given by the following

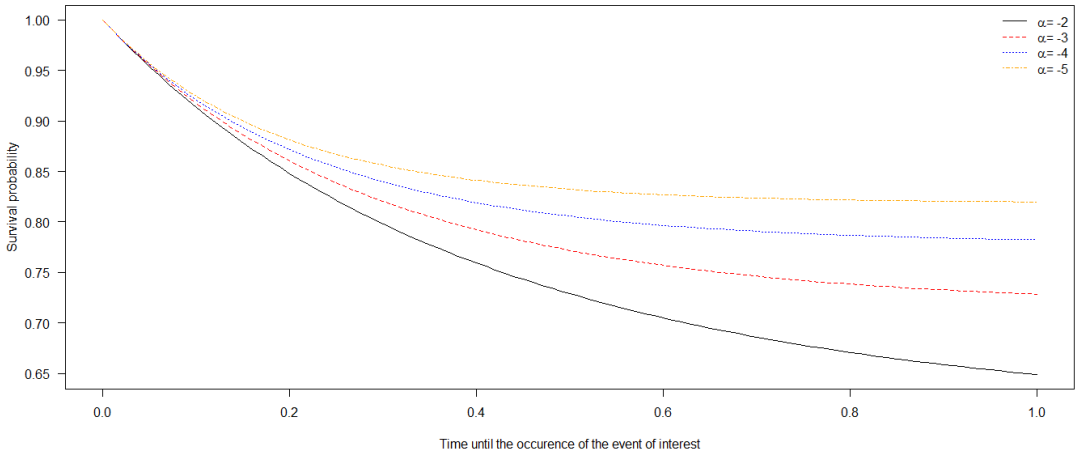

Figure 4.1: Survival function of the defective Gompertz distribution, for different values of the parameter α and with constant β.

formula:

$$S\left(t|\alpha,\beta\right)=1-\left[\Phi\left(\frac{\alpha t-1}{\sqrt{\beta t}}\right)+\exp\left(\frac{2\alpha}{\beta}\right)\Phi\left(\frac{-\alpha t-1}{\sqrt{\beta t}}\right)\right],$$

where $\Phi(.)$ denotes the cumulative distribution function of a standard normal random variable. The Inverse Gaussian distribution takes its defective form when $\alpha<0$. In this case the cure fraction is given by the following relation:

$$\begin{gathered}p_0=\lim_{t\to\infty}S(t|\alpha,\beta)=\\ =\lim_{t\to\infty}\left[1-\left[\Phi\left(\frac{\alpha t-1}{\sqrt{\beta t}}\right)+\exp\left(\frac{2\alpha}{\beta}\right)\Phi\left(\frac{-\alpha t-1}{\sqrt{\beta t}}\right)\right]\right]=\\ =1-\exp\left(\frac{2\alpha}{\beta}\right).\end{gathered}$$

As can be seen, the cure fraction can be estimated using the estimated parameters α and β. In **Figure 4.2** we can see graphs of the survival function of the defective Inverse Gaussian distribution for different values of the parameter α and $\beta=4$. Again, when α takes lower values, the Inverse Gaussian distribution captures a higher proportion of cured individuals. More specifically, we get the following cure fractions:

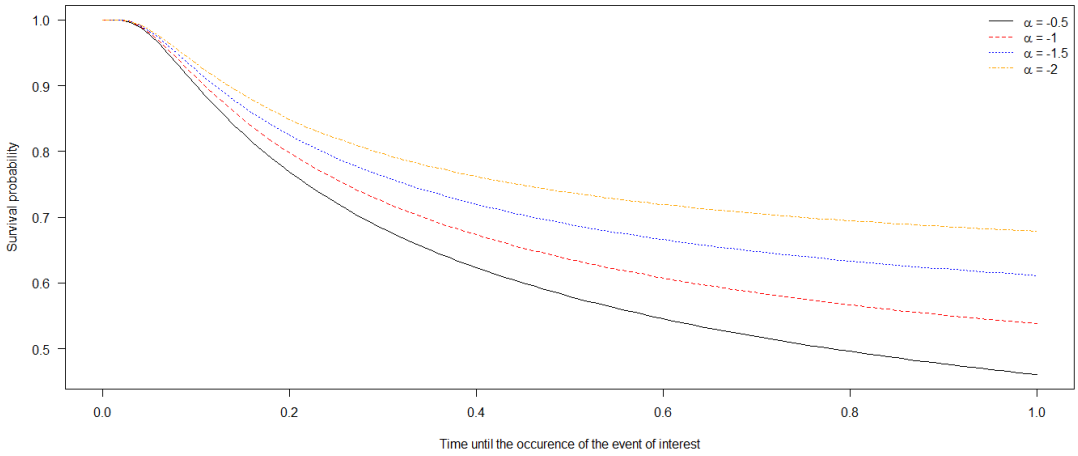

Figure 4.2: Survival function of the defective Inverse Gaussian distribution, for different values of the parameter α and constant β.

$p_1 = 22.12\%$, for $\alpha = -0.5$

$p_2 = 39.35\%$, for $\alpha = -1$

$p_3 = 52.76\%$, for $\alpha = -1.5$ and

$p_4 = 63.21\%$, for $\alpha = -2$.

4.3 Setup of the simulation studies

4.3.1 Artificial data generation

Before introducing the maximum likelihood and the respective Bayesian estimation procedure for the parameters of the baseline defective distributions, we present the simulation algorithm, based on which we simulated right censored survival times from all the defective distributions presented in this thesis. The survival times were simulated according to the following procedure published by Rocha (2016):

- Determine the desired parameter values, as well as the value of the cure fraction p;
- Generate $M_i \sim Bernoulli(1-p)$;
- If $M_i = 0$ set $t_i = \infty$. If $M_i = 1$ take t_i as the root of $F(t) = u$, where $u \sim Uniform(0, 1-p)$;

- Generate $u_i' \sim Uniform(0, \max(t_i))$, considering only the finite t_i;
- Calculate $t_i = \min(t_i, u_i)$. If $t_i < u_i$ set $\delta_i = 1$, otherwise set $\delta_i = 0$.

It is noted that the range of $F(t)$ has been changed, since instead of $(0, 1)$ the interval $(0, 1-p)$ was used. In addition, the censoring distribution chosen is the $Uniform(0, \max(t_i))$, while the limit $\max(t_i)$ was taken in order to control the censoring regardless of the initial parameter choices.

4.3.2 Maximum likelihood estimation

Assume that the data are independently and identically distributed according to a distribution with density and survival functions specified by $f(., \theta)$ and $S(., \theta)$, respectively, where $\theta = (\theta_1, \ldots, \theta_k)$ denotes the vector of parameters which we are interested in. Then, in survival analysis the full data set takes the form $D = (t, \delta)$, where $t = (t_1, \ldots, t_n)$ are the survival times and $\delta = (\delta_1, \ldots, \delta_n)$ are the failure indicators, with $\delta_i = 1$, if a failure is observed and $\delta_i = 0$ otherwise. Then the likelihood function for the full data set can be written as follows:

$$L(D|\theta) \propto \prod_{i=1}^{n} \left[f(t_|\theta)^{\delta_i} S(t_i|\theta)^{1-\delta_i} \right],$$

while the log likelihood function takes the following form:

$$\log L(D|\theta) = c + \sum_{i=1}^{n} \delta_i \log f(t_i|\theta) + \sum_{i=1}^{n} (1-\delta_i) \log S(t_i|\theta),$$

where c is a constant number.

It is noted that the above expression of the likelihood and the log likelihood function for the full data set is valid only when the censoring mechanism is type I, type II or random and non-informative. Since the maximum likelihood estimates of the parameters usually do not have a closed expression, it is necessary to use computational methods to calculate them. In R, where all the simulations were carried out, there are various routines available for numerical maximization, one of which was used in the present thesis. More specifically, we used the *optim* function in order to obtain the maximum likelihood estimates, as suggested by Rocha (2016), based on the BFGS maximization algorithm (Liu & Nocedal (1989)).

As far as the simulation procedure is concerned, we simulated right censored survival times with different sample size ($n = 200, \ldots, 2000$) and then we always chose the value of $S = 1000$ simulations per sample size. In each sample size, we calculated the bias, the mean squared error and the coverage probability for each parameter. The $\hat{\theta}$ was taken as the average of θ_i, for $i = 1, \ldots, S$, while for the calculation of the previous quantities, the following equations were used:

$$Var(\hat{\theta}) = \frac{1}{S}\sum_{i=1}^{S}(\hat{\theta}_i - \theta)^2$$
$$Bias(\hat{\theta}) = \hat{\theta} - \theta \text{ and}$$
$$MSE(\hat{\theta}) = Var(\hat{\theta}) + Bias^2(\hat{\theta})$$

The coverage probability is the frequency in which the real parameter value stays in the confidence region, for each simulation. The cure fraction was calculated based on the estimated parameters and the appropriate equation according to each defective distribution, while for its variance we used the first order Taylor approximation of the Delta method (Oehlert (1992)).

4.3.3 Bayesian estimation

Let $D = (t, \delta)$ be the full data set of survival times, as denoted in section 4.3.2 and let $\theta = (\theta_1, \ldots, \theta_k)$ be the parameters of interest. Then for the Bayesian estimation of these parameters we used the likelihood function of the observed data ($L(D|\theta)$), the prior distribution of the parameters ($\pi(\theta)$) and based on Bayes' rule described in chapter 2, we calculated the conditional posterior distribution of the parameters ($\pi(\theta_i|D, \theta_{-i})$), where θ_{-i} denotes the parameter vector without the i-th element.

First of all it should be noted that the parameter α was replaced from another parameter γ, as $\gamma = -\alpha$, in all the defective distributions (chapter 4 and chapter 5). This replacement was done, since the defective version of the Gompertz and the Inverse Gaussian distribution is based on the assumption that $\alpha < 0$. Therefore, with this replacement the parameter γ takes

only positive values and it is easier to specify a prior distribution. Furthermore, for all the parameters as prior distribution we assumed the Uniform distribution, since we did not aim to include any prior information into the Bayesian inferential procedure.

The MCMC algorithms constructed in both chapters were based on the Gibbs algorithm and the Metropolis Hastings algorithm with normal increments, both of which presented in chapter 2. More specifically, we used the Gibbs algorithm for the parameter β of the defective Gompertz distribution, while for all the other parameters and in all defective distributions we used the Metropolis Hastings algorithm with normal increments, since the conditional posterior distribution of the parameters was not of a recognizable form. In the context of the Metropolis Hastings algorithm, the variance of the normal candidates was defined by trial and error, so that the proportion of accepted draws ranges between 20% and 30% on average. After this step, we used appropriate graphs in order to diagnose the convergence of the simulated chains (trace plots), as well as to investigate the autocorrelation at different lags (autocorrelation plots). The final chain for each parameter was obtained after the appropriate removal of draws and the appropriate thinning, according to these graphs.

For obtaining the final Bayesian estimates of the parameters and the respective cure fraction, we simulated right censored survival times with different sample size ($n = 200, \ldots, 2000$) and then we always chose the value of $S = 1000$ simulations per sample size. The point estimate in each simulation was taken as the mean of the thinned chains and for its 95% credible interval we used the 25%-th percentile and the 75%-th percentile of the chains. The final point estimate for each parameter and each sample size, was calculated as the average value over all the simulations of those mean values, as in the case of the maximum likelihood estimates.

The bias and the variance of the final estimates for each sample size, were calculated according to the equations shown in section 4.3.2, while we used again the Delta method in order to calculate the variance of the cure fraction. The variance of the filal point estimates was used for the construction of their 95% credible intervals. Finally, it is noted that in all distributions in both

chapters we removed the first 2,000 iterations from the beginning of the chains and then we thinned the remaining chains, by keeping every 20-th value in chapter 4 (baseline distributions) and every 300-th value in chapter 5 (extended distributions under the Marshall-Olkin family).

4.4 Maximum likelihood inference

Let $D = (t, \delta)$ be the full data set of right censored survival times, where $t = (t_1, \ldots, t_n)$ are the survival times and $\delta = (\delta_1, \ldots, \delta_n)$ are the failure indicators, with $\delta_i = 1$ if the t_i is a failure time and $\delta_i = 0$ if t_i is a censored time. Then the log likelihood function for the two baseline defective distributions introduced in section 4.2, takes the following form:

For the Gompertz distribution:

$$\log[L(D|\alpha,\beta)] = c + \log(\beta)\sum_{i=1}^{n}\delta_i + \alpha\sum_{i=1}^{n}\delta_i t_i - \frac{\beta\sum_{i=1}^{n}\exp(\alpha t_i)}{\alpha} + \frac{n\beta}{\alpha},$$

while for the Inverse Gaussian:

$$\log[L(D|\alpha,\beta)] = c - \frac{1}{2}\sum_{i=1}^{n}\delta_i\log(2\beta\pi t_i^3) - \frac{\alpha^2}{2\beta}\sum_{i=1}^{n}\delta_i t_i + \frac{\alpha}{\beta}\sum_{i=1}^{n}\delta_i - \frac{1}{2\beta}\sum_{i=1}^{n}\frac{\delta_i}{t_i} +$$
$$+\sum_{i=1}^{n}(1-\delta_i)\log\left[1-\left(\Phi\left(\frac{-1+\alpha t_i}{\sqrt{\beta t_i}}\right)+\exp(\frac{2\alpha}{\beta})\Phi\left(\frac{-1-\alpha t_i}{\sqrt{\beta t_i}}\right)\right)\right],$$

where c is constant. By maximizing numerically the two equations shown above we get the maximum likelihood estimates of the parameters α and β. In addition, the confidence intervals for the parameters were based on asymptotic normality. Therefore, in order to check the asymptotes of the maximum likelihood estimates, as well as the performance of the Delta method used for the calculation of the cure fraction variance, we conducted the simulation study described in the next section. Furthermore, the simulation study aims to investigate whether the sample size affects the estimates' bias and mean squared error, as well as to examine whether the coverage probability remains close to the level of 95% for different sample sizes.

4.4.1 Simulation studies

In this section we propose two simulation scenarios in order to check the maximum likelihood estimates when the sample size increases. First of all we generated data according to the algorithm described in section 4.3, from the defective Gompertz and the defective Inverse Gaussian distribution.

In the first simulation scenario we simulated 1000 random samples each of size $n = 200, \ldots, 2000$. These random samples were assumed to come from the defective Gompertz distribution with $(\alpha, \beta, \pi) = (-1, 1, 0.3679)$. We calculated the maximum likelihood estimates $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\pi}$, as well as their standard errors, which were used to compute the bias, the mean squared error and the coverage probability for each parameter. As for the standard deviation of the cure fraction, the delta method was used. **Figure 4.3** shows the obtained results from this simulation scenario.

The second scenario was simulated from the defective Inverse Gaussian distribution with parameters $(\alpha, \beta, \pi) = (-0.2, 1, 0.3297)$ following the same procedure described above. The obtained results from this simulation scenario are shown in **Figure 4.4**. Both scenarios produced similar results. More specifically, it is noticed that:

1. The mean squared error decreases very smoothly as the sample size increases and its value remains small for any n. In addition, the smallest level of MSE is obtained for the cure fraction estimate.

2. The bias is very small for both parameters (α and β), as well as for the cure fraction (π).

3. The coverage probability stays around 95% even for the smallest value of n, for both parameters, while the coverage probability of the cure fraction is observed at higher levels.

The third point suggests that the Delta method constitutes a good approximation to the variance of the cure fraction, as well as that the asymptotic normality provides good confidence intervals for the distribution's parameters. Therefore, it can be seen that both distributions can give a good point and interval estimation with no need for a large amount of data.

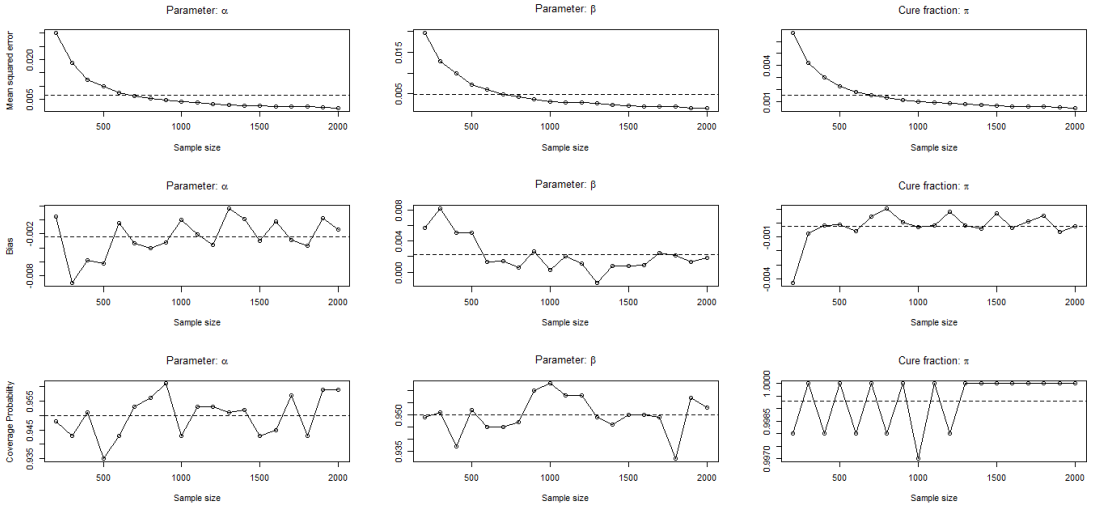

Figure 4.3: Mean squared errors, biases and coverage probabilities of the maximum likelihood estimates $(\hat{\alpha}, \hat{\beta}, \hat{\pi})$ versus the sample size n, from the Gompertz distribution with $(\alpha, \beta, \pi) = (-1, 1, 0.3679)$. The dashed line represents the average mean squared error, bias and coverage probability, over all simulated sample sizes.

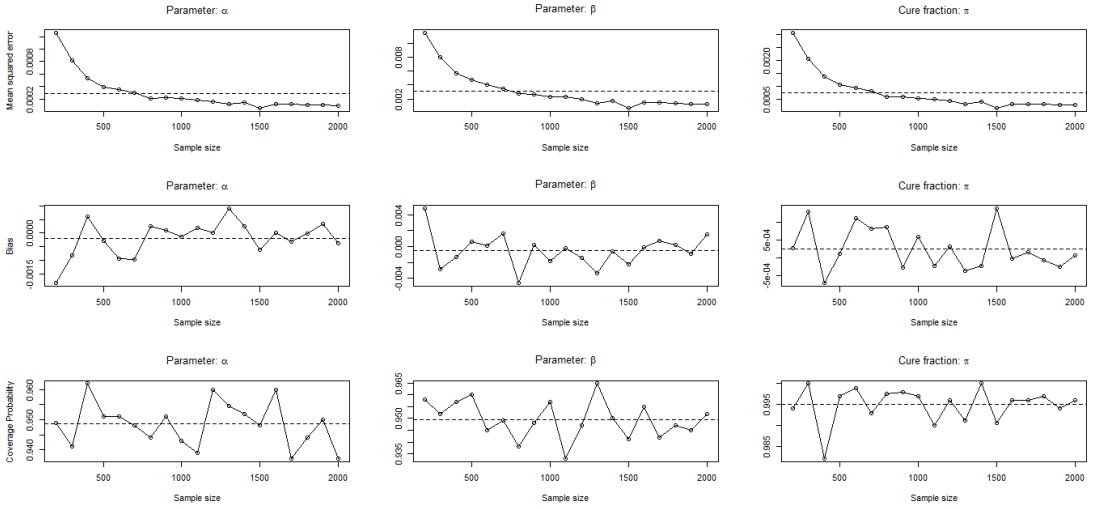

Figure 4.4: Mean squared errors, biases and coverage probabilities of the maximum likelihood estimates $(\hat{\alpha}, \hat{\beta}, \hat{\pi})$ versus the sample size n, from the Inverse Gaussian distribution with $(\alpha, \beta, \pi) = (-0.2, 1, 0.3297)$. The dashed line represents the average mean squared error, bias and coverage probability, over all simulated sample sizes.

4.5 Bayesian Inference

In this section we will present the Bayesian inferential procedure for the defective Gompertz and the defective Inverse Gaussian distribution. As mentioned in section 4.3.3, the parameter α was replaced by the parameter γ as $\gamma = -\alpha$ so that $\gamma > 0$. Then the likelihood function of both distributions, takes the following form:

For the Gompertz distribution:

$$L((T_1,\delta_1),(T_2,\delta_2),\ldots,(T_n,\delta_n)) = \beta^{\sum_{i=1}^{n}\delta_i}\exp[-\gamma\sum_{i=1}^{n}\delta_i t_i + \frac{\beta}{\gamma}\sum_{i=1}^{n}[\exp(-\gamma t_i)-1]],$$

while for the defective Inverse Gaussian distribution, it takes the following form:

$$L\left((T_1,\delta_1),(T_2,\delta_2),\ldots,(T_n,\delta_n)\right) = (2\pi)^{-\frac{\sum_{i=1}^{n}\delta_i}{2}}\exp(-\frac{3}{2}\sum_{i=1}^{n}\delta_i\log(t_i))\beta^{-\frac{\sum_{i=1}^{n}\delta_i}{2}}\times$$
$$\times\exp\left(-\frac{\gamma^2}{2\beta}\sum_{i=1}^{n}\delta_i t_i - \frac{\gamma}{\beta}\sum_{i=1}^{n}\delta_i - \frac{1}{2\beta}\sum_{i=1}^{n}\frac{\delta_i}{t_i}\right)\exp\left(\sum_{i=1}^{n}(1-\delta_i)\log(S(t_i|\gamma,\beta))\right),$$

where $S(t_i|\gamma,\beta) = 1 - \left[\Phi\left(\frac{-1-\gamma t}{\sqrt{\beta t}}\right) + \exp\left(-\frac{2\gamma}{\beta}\right)\Phi\left(\frac{-1+\gamma t}{\sqrt{\beta t}}\right)\right]$.

Since there is no joint conjugate prior distribution for the parameters (γ,β), we assumed that the two parameters were independently distributed according to the Uniform distribution in both defective distributions. Then, the joint posterior distribution of the parameters takes the following form:

$$\pi\left(\beta,\gamma|(T_1,\delta_1),(T_2,\delta_2),\ldots,(T_n,\delta_n)\right)\propto$$
$$\propto L((T_1,\delta_1),(T_2,\delta_2),\ldots,(T_n,\delta_n)|\beta,\gamma)\times\pi(\beta)\times\pi(\gamma)\propto$$
$$\propto L((T_1,\delta_1),(T_2,\delta_2),\ldots,(T_n,\delta_n)|\beta,\gamma).$$

So, after the above formulation of the joint posterior distribution we get the following relations for the conditional posterior distribution of β and γ.

For the Gompertz distribution:

$$\pi(\beta|\gamma,(T_1,\delta_1),(T_2,\delta_2),\ldots,(T_n,\delta_n)) \propto \beta^{\sum_{i=1}^{n}\delta_i} \exp\left[\frac{\sum_{i=1}^{n}[\exp(-\gamma t_i)-1]}{\gamma}\beta\right] \text{ and}$$

$$\pi(\gamma|\beta,(T_1,\delta_1),(T_2,\delta_2),\ldots,(T_n,\delta_n)) \propto \exp\left[-\gamma\sum_{i=1}^{n}\delta_i t_i + \frac{\beta}{\gamma}\sum_{i=1}^{n}[\exp(-\gamma t_i)-1]\right],$$

while for the Inverse Gaussian distribution:

$$\begin{aligned}
&\pi\left(\beta|\gamma,(T_1,\delta_1),(T_2,\delta_2),\ldots,(T_n,\delta_n)\right) \propto \\
&\propto \beta^{-\frac{\sum_{i=1}^{n}\delta_i}{2}} \exp\left(-\frac{\gamma^2}{2\beta}\sum_{i=1}^{n}\delta_i t_i - \frac{\gamma}{\beta}\sum_{i=1}^{n}\delta_i - \frac{1}{2\beta}\sum_{i=1}^{n}\frac{\delta_i}{t_i}\right) \times \\
&\times \exp\left(\sum_{i=1}^{n}(1-\delta_i)\log(S(t_i|\gamma,\beta))\right) \text{ and}
\end{aligned}$$

$$\begin{aligned}
&\pi\left(\gamma|\beta,(T_1,\delta_1),(T_2,\delta_2),\ldots,(T_n,\delta_n)\right) \propto \\
&\propto \exp\left(-\frac{\gamma^2}{2\beta}\sum_{i=1}^{n}\delta_i t_i - \frac{\gamma}{\beta}\sum_{i=1}^{n}\delta_i\right) \exp\left(\sum_{i=1}^{n}(1-\delta_i)\log(S(t_i|\gamma,\beta))\right).
\end{aligned}$$

As it can be seen, the conditional posterior distribution of β in the case of the defective Gompertz distribution is the Gamma distribution with parameters $p = \sum_{i=1}^{n}\delta_i + 1$ and $q = -\frac{1}{\gamma}\sum_{i=1}^{n}[\exp(-\gamma t_i)-1]$, while the form of the conditional posterior distribution of γ can not be recognized in both distributions.

4.5.1 Simulation studies

In this section we present two simulation scenarios, one for each defective distribution, in order to obtain the Bayesian estimates for their parameters. As mentioned in section 4.1, this is the first attempt for obtaining the Bayesian estimates of the parameters of the defective Inverse Gaussian distribution and, in addition, a different Bayesian inferential procedure is provided for the parameters of the defective Gompertz distribution compared to that of the paper of MR dos Santos et.al (2017).

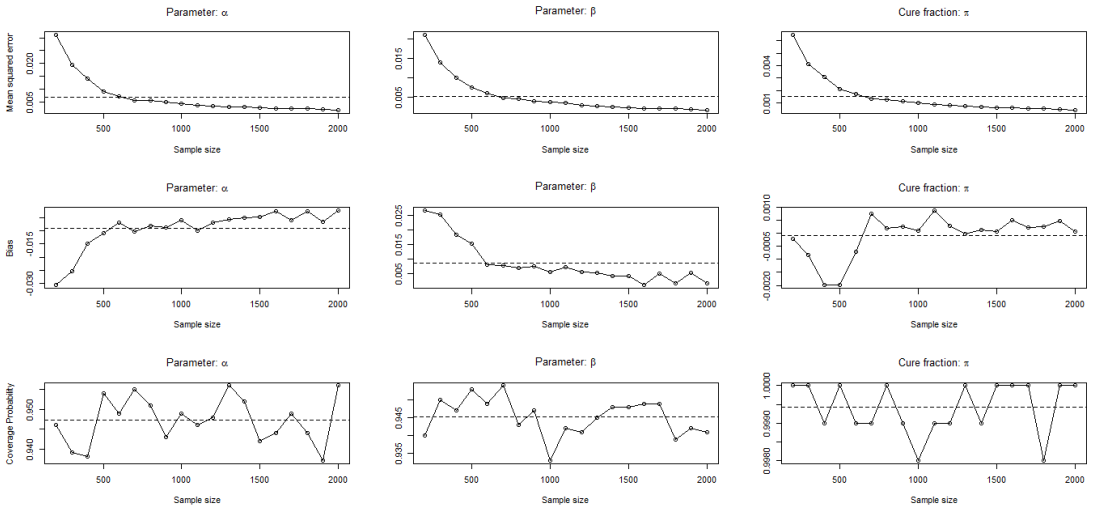

Figure 4.5: Mean squared errors, biases and coverage probabilities of the Bayesian estimates $(\hat{\alpha}, \hat{\beta}, \hat{\pi})$ versus the sample size n, from the Gompertz distribution with $(\alpha, \beta, \pi) = (-1, 1, 0.3679)$. The dashed line represents the average mean squared error, bias and coverage probability, over all simulated sample sizes.

First of all, we simulated right censored survival times of different sample sizes (n =200,. . .,2000) from the defective Gompertz distribution with parameters $\alpha = -1$, $\beta = 1$ and $\pi = 0.3679$. For the calculation of the parameters' Bayesian estimates we followed the procedure described in section 4.5. In **Figure 4.5** we present the results concerning this simulation scenario and it is observed that:

- The mean squared error of both parameters and the cure fraction decreases smoothly with the sample size, with the lowest being succeeded in the estimation of the cure fraction and the biggest in the estimation of the parameter α.
- The bias of the Bayesian estimates seems to be at very low levels, with the smallest bias being observed in the estimation of the cure rate.
- Finally, as far as the coverage probability is concerned, it is observed that it is very high for both parameters and the cure fraction, with the average probability being very close to 95.0%. In addition, it is seen that the coverage probability for π is the highest one, compared to those obtained for the other two parameters.

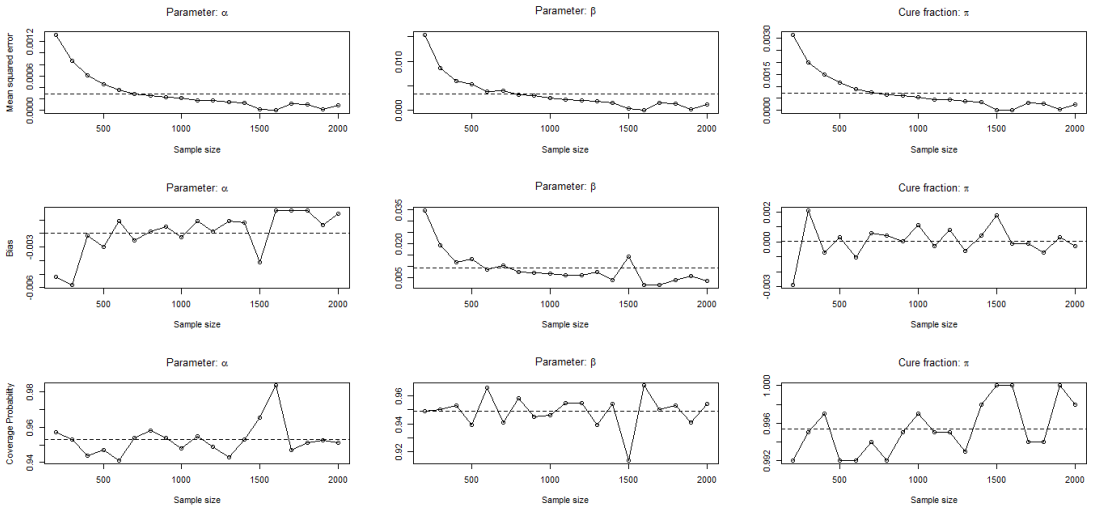

Figure 4.6: Mean squared errors, biases and coverage probabilities of the Bayesian estimates $(\hat{\alpha}, \hat{\beta}, \hat{\pi})$ versus the sample size n, from the Inverse Gaussian distribution with $(\alpha, \beta, \pi) = (-0.2, 1, 0.32968)$. The dashed line represents the average mean squared error, bias and coverage probability, over all simulated sample sizes.

Subsequently, we simulated right censored survival times of different sample sizes ($n = 200, \ldots, 2000$) from the defective Inverse Gaussian distribution with parameters $\alpha = -0.2$, $\beta = 1$ and $\pi = 0.32968$. The conclusions seem to be the same as in the previous case of the defective Gompertz distribution. More specifically, in **Figure 4.6**, we present the results regarding this specific simulation scenario, where it is observed that:

- The mean squared error decreases smoothly as the sample size increases, with the one concerning the parameter α being the smallest compared to the MSE obtained from the estimation of the parameter β and the cure fraction π.
- In addition, as far as the bias is concerned, we can see that it remains at very low levels, with the one obtained from the estimation of the cure fraction being the smallest one.
- Finally, the coverage probability is very satisfactory for both parameters, as well as for the cure fraction, since it can be seen that it is approximately equal to 95% in all cases.

4.6 Comparative results

In this section we are going to compare the two estimating methodologies, concerning the two baseline distributions. After presenting the results from the simulation experiments under the frequentist and the Bayesian perspective the following conclusions can be drawn:

1. *G*ompertz distribution

    - **Mean squared error:** The two inferential procedures achieve approximately the same mean squared errors for both parameters and the cure fraction.
    - **Bias:** The absolute bias achieved by the maximum likelihood estimates is smaller than the one achieved by the respective Bayesian estimates for both parameters and the cure fraction.
    - **Coverage Probability:** Finally, the coverage probability under the frequentist approach seems to be a little bit bigger than the one based on the Bayesian parameter estimates.

2. *I*nverse Gaussian distribution

    - The conclusions drawn concerning the mean squared error and the bias of the estimates are similar to those mentioned before.
    - **Coverage Probability:** As far as the parameter α and the cure fraction (π) is concerned, we can see that the coverage probability achieved under the Bayesian framework is greater than the one based on the maximum likelihood inferential procedure. However, it seems that for the parameter β the two inferential procedures succeed approximately the same probability.

Chapter 5

Defective distributions under the Marshall-Olkin family

5.1 Introduction & Background

Except for the two baseline distributions (Gompertz & Inverse Gaussian) presented in the previous chapter, Rocha (2016) also suggested the use of their extension under various families of distributions for cure rate modelling. More specifically, he proposed two new defective distributions under the Marshall-Olkin family (Marshall & Olkin, 1997), a family which is obtained by adding an extra parameter to a known baseline distribution. He showed that if the baseline distribution is defective, then its extension under this specific family is also defective and can be used for cure rate modelling. Let $S(t)$ be the baseline survival function. Then, the extension under the Marshall-Olkin family of distributions is given by the following relation:

$$S^*(t) = \frac{rS(t)}{1-(1-r)S(t))},$$

for $r > 0$ and $t \in \mathbb{R}$. In addition, by making some simple algebraic manipulation we get the following relation for its probability density function:

$$f^*(t) = \frac{rf(t)}{[1-(1-r)S(t)]^2}.$$

In the literature, there are many Marshall-Olkin G distributions which are studied, such as the Marshall-Olkin asymmetric Laplace distribution (Krishna & Jose, 2011), the Marshall-Olkin beta distribution (Jose et al., 2009), the Marshall-Olkin Birnbaum-Saunders distribution (Lemonte, 2013) etc. These distributions have been already used to model several types of data, among which are some daily ozone measurements in New York (Jose et al., 2009), the remission times of a random sample of bladder cancer patients (Ghitany et al., 2005, 2007), as well as the survival times of guinea pigs injected with different doses of tubercle bacilli (Krishna et al., 2013). Since the use of the family was broad enough, Rocha decided to propose two new defective distributions, extending the Gompertz and the Inverse Gaussian distributions, through the Marshall-Olkin family. In addition, he performed a simulation study in order to assess the performance of the maximum likelihood estimators and he illustrated the proposed distributions using several real data sets.

As mentioned above, the main result presented by Rocha was the fact that, if a given distribution (baseline) is defective, then its extension under the Marshall-Olkin family is also defective. This means that if the limit of $S(t)$ equals $p_0 \in (0, 1)$, then:

$$\lim_{t\to\infty} S^*(t) = \lim_{t\to\infty} \frac{rS(t)}{1-(1-r)S(t)} = \frac{rp_0}{1-(1-r)p_0} = \frac{rp_0}{rp_0+1-p_0}.$$

Based on this specification, Rocha presented the two new defective distributions, the Marshall-Olkin Gompertz and Marshall-Olkin Inverse Gaussian distributions. Our aim is to reproduce the maximum likelihood estimates for the parameters of the two extended defective distributions obtained by Rocha, as well as to obtain, for the first time, the Bayesian estimates for their parameters.

5.2 The extended defective distributions

Using the two relations shown in section (5.1) with the density and the survival function given by the Gompertz distribution as the baseline distribution, we get the following relations for the density and the survival function of the Marshall-Olkin Gompertz distribution:

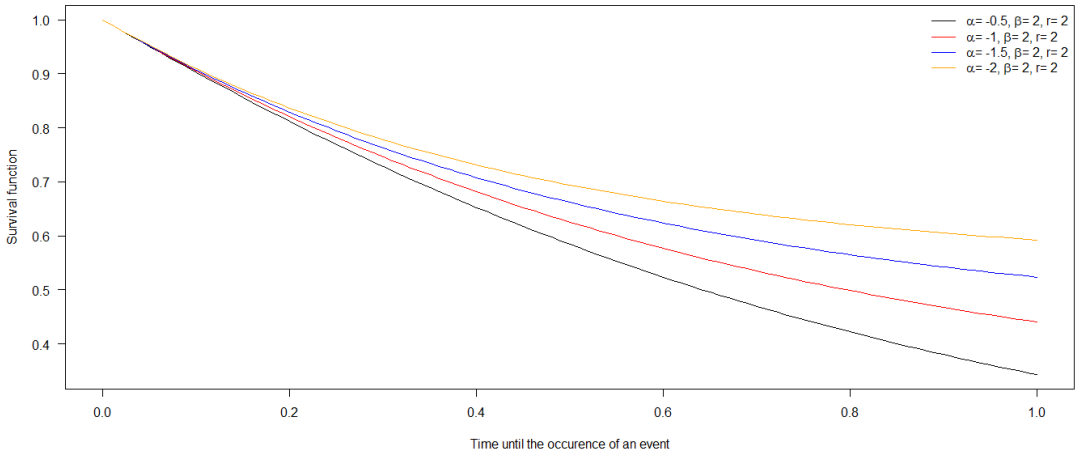

Figure 5.1: Survival function of the Marshall-Olkin Gompertz distribution for different parameters' values.

$$f^*(t) = \frac{\beta r \exp(\alpha t) \exp\left(-\frac{\beta}{\alpha}\right)(\exp(\alpha t) - 1)}{\left[1 - (1-r)\exp\left(-\frac{\beta}{\alpha}\right)(\exp(\alpha t) - 1)\right]^2} \text{ and}$$

$$S^*(t) = \frac{r \exp\left(-\frac{\beta}{\alpha}\right)(\exp(\alpha t) - 1)}{1 - (1-r)\exp\left(-\frac{\beta}{\alpha}\right)(\exp(\alpha t) - 1)}.$$

In **Figure 5.1** we show the graph of the survival function of the Marshall-Olkin Gompertz distribution. More specifically, we display the survival function of this distribution for different values of the parameter α while the other two parameters were kept fixed. As it can be seen from the figure, as the α parameter takes lower values, the cure fraction in the data set seems to increase. In addition, if the parameter r equals 1, the Marshall-Olkin Gompertz distribution diminishes to the Gompertz distribution. According to Rocha, the Marshall-Olkin Gompertz distribution takes its defective form when $\alpha < 0$. In this case, the cure fraction which can be estimated is given in the following relation:

$$p = \lim_{t\to\infty} S^*(t) = \lim_{t\to\infty}\left[1 - \frac{1}{r\exp(\frac{\beta(\exp(\alpha t)-1)}{\alpha}) - r + 1}\right] = \frac{rp_0}{1-(1-r)p_0} = \frac{rp_0}{rp_0 + 1 - p_0},$$

where $p_0 = \exp\left(\frac{\beta}{\alpha}\right)$ is the cure fraction estimated by the defective Gompertz distribution.

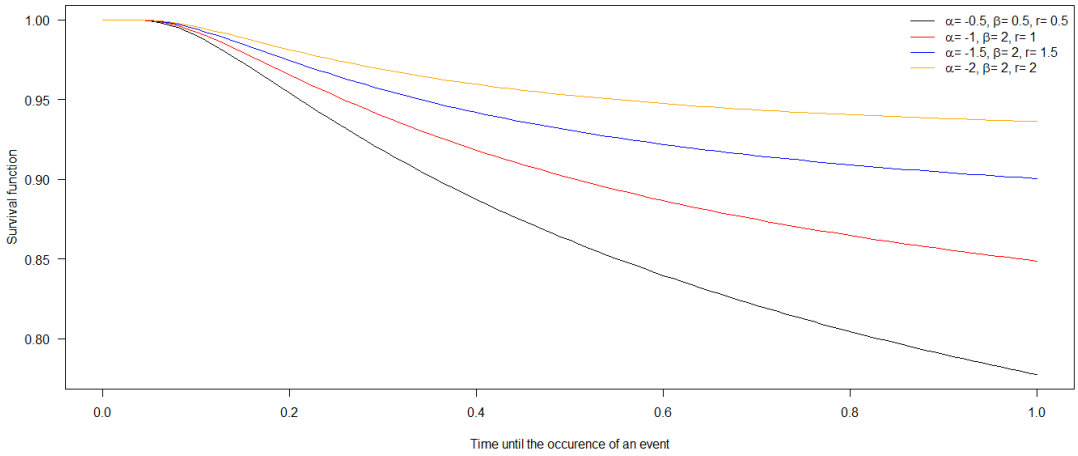

Figure 5.2: Survival function of the Marshall-Olkin Inverse Gaussian distribution for different parameters' values.

Instead, if the baseline distribution is the Inverse Gaussian distribution, then we get the Marshall-Olkin Inverse Gaussian distribution which satisfies the following properties:

$$f^*(t) = \frac{r\left\{\frac{1}{\sqrt{2\pi\beta t^3}}\exp\left[-\frac{1}{2\beta t}(\alpha t-1)^2\right]\right\}}{\left\{1+(r-1)\left[1-\left(\Phi\left(\frac{-1+\alpha t}{\sqrt{\beta t}}\right)+\exp(\frac{2\alpha}{\beta})\Phi\left(\frac{-1-\alpha t}{\sqrt{\beta t}}\right)\right)\right]\right\}^2} \text{ and}$$

$$S^*(t) = \frac{r\left[1-\left(\Phi\left(\frac{-1+\alpha t}{\sqrt{\beta t}}\right)+\exp(\frac{2\alpha}{\beta})\Phi\left(\frac{-1-\alpha t}{\sqrt{\beta t}}\right)\right)\right]}{\left\{1+(r-1)\left[1-\left(\Phi\left(\frac{-1+\alpha t}{\sqrt{\beta t}}\right)+\exp(\frac{2\alpha}{\beta})\Phi\left(\frac{-1-\alpha t}{\sqrt{\beta t}}\right)\right)\right]\right\}}.$$

In **Figure 5.2**, it is observed that when the parameter α takes higher values, then the cure fraction in the data set declines. Furthermore, when $\alpha < 0$ the Marshall-Olkin Inverse Gaussian distribution becomes defective and it estimates the following cure fraction:

$$p = \lim_{t\to\infty} S^*(t) = \frac{rp_0}{rp_0+1-p_0},$$

where p_0 is the cure fraction of the defective Inverse Gaussian distribution and is given by:

$$p_0 = 1-\exp\left\{\frac{2\alpha}{\beta}\right\}.$$

In the rest of this chapter we are going to obtain the maximum likelihood estimates for the parameters of the two extended defective distributions, as well as the respective Bayesian estimates, through a simulation study following the setup presented in the previous chapter.

5.3 Maximum likelihood Inference

Let $D = (t, \delta)$ be the full data set of right censored survival times, where $t = (t_1, \ldots, t_n)$ are the survival times and $\delta = (\delta_1, \ldots, \delta_n)$ are the failure indicators. Then, the log likelihood function of the Marshall-Olkin Gompertz distribution takes the following form:

$$\log[L(D|\alpha,\beta,r)] = \log(\beta)\sum_{i=1}^{n}\delta_i + n\log(r) + \alpha\sum_{i=1}^{n}\delta_i t_i - \frac{\beta}{\alpha}\sum_{i=1}^{n}(\exp(\alpha t_i)-1)$$
$$-\sum_{i=1}^{n}(1+\delta_i)\log\left(1-(1-r)\exp(-\frac{\beta}{\alpha}(\exp(\alpha t_i)-1))\right),$$

while, the respective quantity for the defective Marshall-Olkin Inverse Gaussian distribution takes the following form:

$$\log[L(D|\alpha,\beta,r)] = n\log(r) - \frac{1}{2}\sum_{i=1}^{n}\delta_i\log\left(2\pi\beta t_i^3\right) - \frac{1}{2\beta}\sum_{i=1}^{n}\delta_i\left(\alpha^2 t_i - 2\alpha + \frac{1}{t_i}\right)$$
$$+\sum_{i=1}^{n}(1-\delta_i)\log\left(1-\left\{\Phi\left(\frac{-1+\alpha t_i}{\sqrt{\beta t_i}}\right)+\exp\left(\frac{2\alpha}{\beta}\right)\Phi\left(\frac{-1-\alpha t_i}{\sqrt{\beta t_i}}\right)\right\}\right)$$
$$-\sum_{i=1}^{n}(1+\delta_i)\log\left(1+(r-1)\left[1-\left\{\Phi\left(\frac{-1+\alpha t_i}{\sqrt{\beta t_i}}\right)+\exp\left(\frac{2\alpha}{\beta}\right)\Phi\left(\frac{-1-\alpha t_i}{\sqrt{\beta t_i}}\right)\right\}\right]\right).$$

5.3.1 Simulation studies

Here we perform one simulation experiment. We simulated one thousand random samples each of size $n = 200, \ldots, 2000$, according to the algorithm described in section 4.3.1. The random samples were taken to come from i) the Marshall-Olkin Gompertz distribution with $(\alpha, \beta, r, \pi) = (-1, 4, 2, 0.0360)$ and ii) the Marshall-Olkin Inverse Gaussian distribution with $(\alpha, \beta, r, \pi) = (-2, 10, 2, 0.4958)$. We computed the maximum likelihood estimates,$\hat{\alpha}, \hat{\beta}, \hat{r}$ and $\hat{\pi}$, as well as their standard errors for each sample. These were used to compute the bias, the mean squared error and the coverage probability, as we did in the previous chapter for the baseline

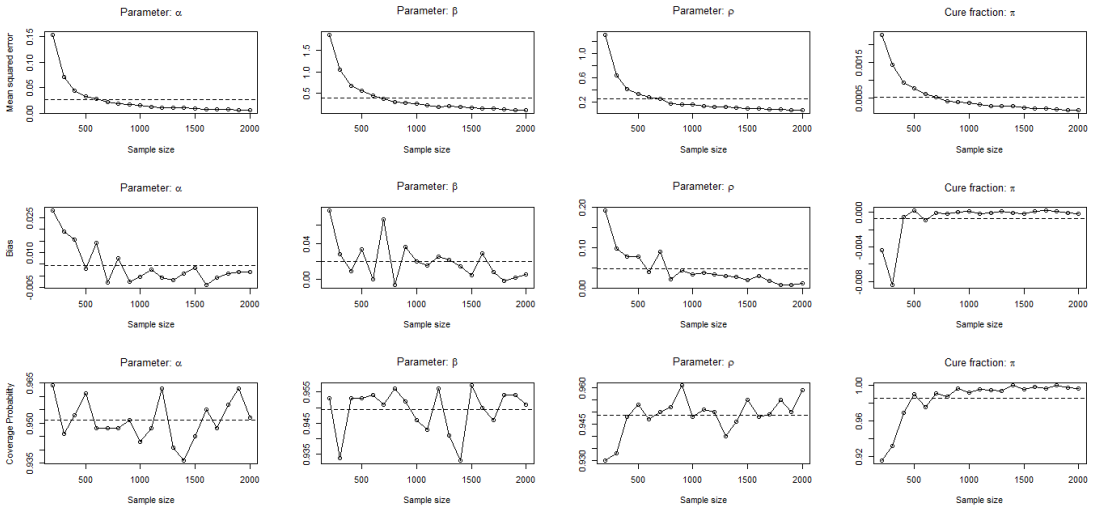

Figure 5.3: Mean squared errors, biases and coverage probabilities of the maximum likelihood estimates $(\hat{\alpha}, \hat{\beta}, \hat{r}, \hat{\pi})$ versus the sample size n, from the Marshall-Olkin Gompertz distribution with $(\alpha, \beta, r, \pi) = (-1, 4, 2, 0.0360)$. The dashed line represents the average mean squared error, bias and coverage probability, over all simulated sample sizes.

defective distributions.

Figure 5.3 and **Figure 5.4** show the mean squared errors, the biases and the coverage probabilities of $(\hat{\alpha}, \hat{\beta}, \hat{r}, \hat{\pi})$ versus the sample size n for the simulated data from the defective Marhall-Olkin Gompertz and the defective Marshall-Olkin Inverse Gaussian distributions. From both figures we can observe the following:

- The mean squared error seems to decrease to zero for all the parameters with increasing sample size.
- The smallest mean squared errors seem to be achieved in the estimation of the cure fraction, while the largest one, in the estimation of the parameter β and then in the estimation of the parameter r.
- The biases seem to be very small in the estimation of all the parameters, as well as in the estimation of the cure fraction.
- Finally, the coverage probabilities are very close to the level of 95%, with those obtained

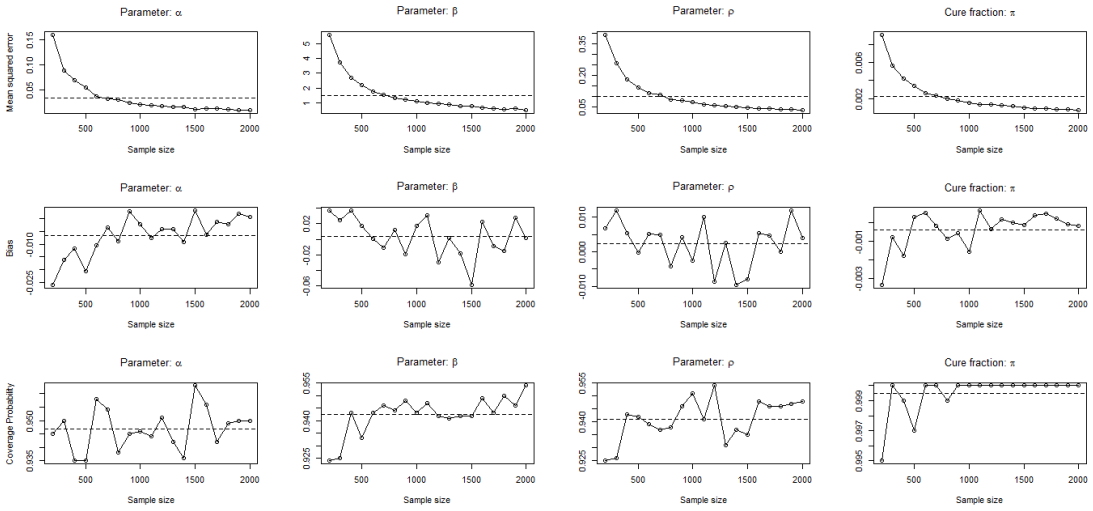

Figure 5.4: Mean squared errors, biases and coverage probabilities of the maximum likelihood estimates $(\hat{\alpha}, \hat{\beta}, \hat{r}, \hat{\pi})$ versus the sample size n, from the Marshall-Olkin Inverse Gaussian distribution with $(\alpha, \beta, r, \pi) = (-2, 10, 2, 0.4958)$. The dashed line represents the average mean squared error, bias and coverage probability, over all simulated sample sizes.

by the $\hat{\pi}$ being the largest.

5.4 Bayesian Inference

In this section, we will obtain the Bayesian estimates for the parameters of the defective Gompertz and the defective Inverse Gaussian distribution, under the Marshall-Olkin family, after replacing the parameter α with the parameter γ as in the section 4.5. In section 5.2 we presented the probability density function and the respective survival function of the defective Gompertz and the defective Inverse Gaussian distributions under the Marshall-Olkin family. Based on these relations and after replacing the parameter α, the likelihood function for the full data set of right censored survival times takes the following form:

For the Marshall-Olkin Gompertz distribution:

$$L((T_1,\delta_1),(T_2,\delta_2),\ldots,(T_n,\delta_n)|\gamma,\beta,r) = \beta^{\sum_{i=1}^{n}\delta_i} r^n \exp[-\gamma\sum_{i=1}^{n}\delta_i t_i]\exp\left[\frac{\beta}{\gamma}\sum_{i=1}^{n}(\exp(-\gamma t_i)-1)\right]\times$$

$$\times \exp\left\{-\sum_{i=1}^{n}(1+\delta_i)\log\left[1-(1-r)\exp\left(\frac{\beta}{\gamma}(\exp(-\gamma t_i)-1)\right)\right]\right\}$$

and for the Marshall-Olkin Inverse Gaussian distribution:

$$L((T_1,\delta_1),(T_2,\delta_2),\ldots,(T_n,\delta_n)|\gamma,\beta,r) = r^n(2\pi)^{-\frac{\sum_{i=1}^{n}\delta_i}{2}}\beta^{-\frac{\sum_{i=1}^{n}\delta_i}{2}}\exp\left[-\frac{3}{2}\sum_{i=1}^{n}\delta_i\log(t_i)\right]\times$$
$$\times\exp\left[-\frac{1}{2\beta}(\gamma^2\sum_{i=1}^{n}\delta_i t_i+2\gamma\sum_{i=1}^{n}\delta_i+\sum_{i=1}^{n}\frac{\delta_i}{t_i})\right]\times$$
$$\times\exp\left\{\sum_{i=1}^{n}(1-\delta_i)\log\left[1-\left(\Phi\left[\frac{-1-\gamma t_i}{\sqrt{\beta t_i}}\right]+\exp\left(\frac{-2\gamma}{\beta}\right)\Phi\left[\frac{-1+\gamma t_i}{\sqrt{\beta t_i}}\right]\right)\right]\right\}\times$$
$$\times\exp\left\{-\sum_{i=1}^{n}(1+\delta_i)\log\left[1+(r-1)\left(1-\left[\Phi\left(\frac{-1-\gamma t_i}{\sqrt{\beta t_i}}\right)+\exp\left(\frac{-2\gamma}{\beta}\right)\Phi\left(\frac{-1+\gamma t_i}{\sqrt{\beta t_i}}\right)\right]\right)\right]\right\}.$$

It can be seen that there is no conjugate prior distribution for the parameters (γ,β,r). Therefore, we set the Uniform distribution as their prior, so that no prior beliefs would be included in the inferential procedure. After this assumption, the conditional posterior distributions of the parameters take the following form:

For the Marshall-Olkin Gompertz distribution:

Parameter γ:

$$\pi(\gamma|\beta,r,(T_1,\delta_1),(T_2,\delta_2),\ldots,(T_n,\delta_n))\propto$$
$$\propto\exp[-\gamma\sum_{i=1}^{n}\delta_i t_i]\exp\left[\frac{\beta}{\gamma}\sum_{i=1}^{n}(\exp(-\gamma t_i)-1)\right]\times$$
$$\times\exp\left\{-\sum_{i=1}^{n}(1+\delta_i)\log\left[1-(1-r)\exp\left(\frac{\beta}{\gamma}(\exp(-\gamma t_i)-1)\right)\right]\right\},$$

Parameter β:

$$\pi(\beta|\gamma,r,(T_1,\delta_1),(T_2,\delta_2),\ldots,(T_n,\delta_n))\propto$$
$$\propto\beta^{\sum_{i=1}^{n}\delta_i}\exp\left[\frac{\beta}{\gamma}\sum_{i=1}^{n}(\exp(-\gamma t_i)-1)\right]\times$$
$$\times\exp\left\{-\sum_{i=1}^{n}(1+\delta_i)\log\left[1-(1-r)\exp\left(\frac{\beta}{\gamma}(\exp(-\gamma t_i)-1)\right)\right]\right\}$$

and parameter r:

$$\pi(r|\gamma,\beta,(T_1,\delta_1),(T_2,\delta_2),\ldots,(T_n,\delta_n)) \propto$$
$$\propto r^n \exp\left\{-\sum_{i=1}^{n}(1+\delta_i)\log\left[1-(1-r)\exp\left(\frac{\beta}{\gamma}(\exp(-\gamma t_i)-1)\right)\right]\right\},$$

while for the Marshall-Olkin Inverse Gaussian distribution we get the following equations for the conditional posterior distributions of the parameters:

Parameter γ:

$$\pi(\gamma|\beta,r,(T_1,\delta_1),(T_2,\delta_2),\ldots,(T_n,\delta_n)) \propto$$
$$\propto \exp\left[-\frac{1}{2\beta}(\gamma^2\sum_{i=1}^{n}\delta_i t_i + 2\gamma\sum_{i=1}^{n}\delta_i + \sum_{i=1}^{n}\frac{\delta_i}{t_i})\right]\times$$
$$\times\exp\left\{\sum_{i=1}^{n}(1-\delta_i)\log\left[1-\left(\Phi\left[\frac{-1-\gamma t_i}{\sqrt{\beta t_i}}\right]+\exp\left(\frac{-2\gamma}{\beta}\right)\Phi\left[\frac{-1+\gamma t_i}{\sqrt{\beta t_i}}\right]\right)\right]\right\}\times$$
$$\times\exp\left\{-\sum_{i=1}^{n}(1+\delta_i)\log\left[1+(r-1)\left(1-\left[\Phi\left(\frac{-1-\gamma t_i}{\sqrt{\beta t_i}}\right)+\exp\left(\frac{-2\gamma}{\beta}\right)\Phi\left(\frac{-1+\gamma t_i}{\sqrt{\beta t_i}}\right)\right]\right)\right]\right\},$$

Parameter β:

$$\pi(\beta|\gamma,r,(T_1,\delta_1),(T_2,\delta_2),\ldots,(T_n,\delta_n)) \propto$$
$$\propto \beta^{-\frac{\sum_{i=1}^{n}\delta_i}{2}}\exp\left[-\frac{1}{2\beta}(\gamma^2\sum_{i=1}^{n}\delta_i t_i + 2\gamma\sum_{i=1}^{n}\delta_i + \sum_{i=1}^{n}\frac{\delta_i}{t_i})\right]\times$$
$$\times\exp\left\{\sum_{i=1}^{n}(1-\delta_i)\log\left[1-\left(\Phi\left[\frac{-1-\gamma t_i}{\sqrt{\beta t_i}}\right]+\exp\left(\frac{-2\gamma}{\beta}\right)\Phi\left[\frac{-1+\gamma t_i}{\sqrt{\beta t_i}}\right]\right)\right]\right\}\times$$
$$\times\exp\left\{-\sum_{i=1}^{n}(1+\delta_i)\log\left[1+(r-1)\left(1-\left[\Phi\left(\frac{-1-\gamma t_i}{\sqrt{\beta t_i}}\right)+\exp\left(\frac{-2\gamma}{\beta}\right)\Phi\left(\frac{-1+\gamma t_i}{\sqrt{\beta t_i}}\right)\right]\right)\right]\right\}$$

and parameter r:

$$\pi(r|\gamma,\beta,(T_1,\delta_1),(T_2,\delta_2),\ldots,(T_n,\delta_n)) \propto r^n\times$$
$$\times\exp\left\{-\sum_{i=1}^{n}(1+\delta_i)\log\left[1+(r-1)\left(1-\left[\Phi\left(\frac{-1-\gamma t_i}{\sqrt{\beta t_i}}\right)+\exp\left(\frac{-2\gamma}{\beta}\right)\Phi\left(\frac{-1+\gamma t_i}{\sqrt{\beta t_i}}\right)\right]\right)\right]\right\}.$$

We can see that none of the above conditional posterior distribution are of recognizable form. For that reason, in both distributions and for all the parameters, we used the Metropolis Hastings algorithm with normal increments, as mentioned in section 4.3.3.

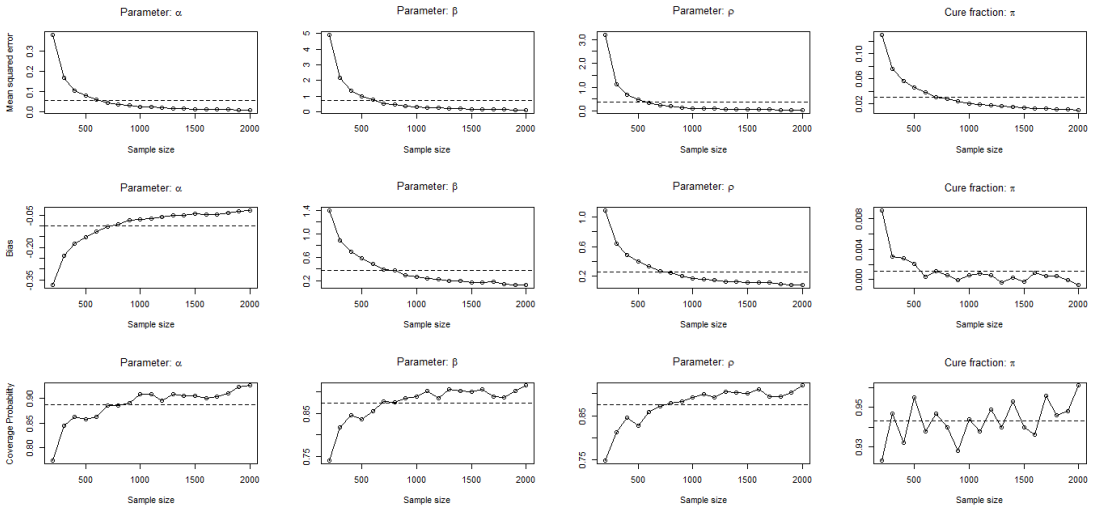

Figure 5.5: Mean squared errors, biases and coverage probabilities of the Bayesian estimates $(\hat{\alpha}, \hat{\beta}, \hat{r}, \hat{\pi})$ versus the sample size n, from the Marshall-Olkin Gompertz distribution with $(\alpha, \beta, r, \pi) = (-1, 1, 0.5, 0.2254)$. The dashed line represents the average mean squared error, bias and coverage probability, over all simulated sample sizes.

5.4.1 Simulation studies

In this section we present two simulation scenarios, one for each extended defective distribution under the Marshall-Olkin family of distributions. This is the first presentation of the Bayesian inferential procedure for the parameters of the defective Marshall-Olkin Gompertz and the defective Marshall-Olkin Inverse Gaussian distribution in the literature.

First of all we simulated 1,000 random samples each one with sample size $n = 200, \ldots, 2000$, according to the algorithm described in section 4.3.1, from the defective Marshall-Olkin Gompertz distribution with $(\alpha, \beta, r, \pi) = (-1, 1, 0.5, 0.2254)$. In **Figure 5.5** we present the results concerning this simulation scenario, and more specifically we present the mean squared error, the bias and the coverage probability for all the distribution parameters, versus the sample size n. According to **Figure 5.5** the following conclusions can be drawn:

- **Mean squared error:** It can be seen that the mean squared error decreases smoothly as the sample size increases. In addition, it should be noted that the largest mean squared

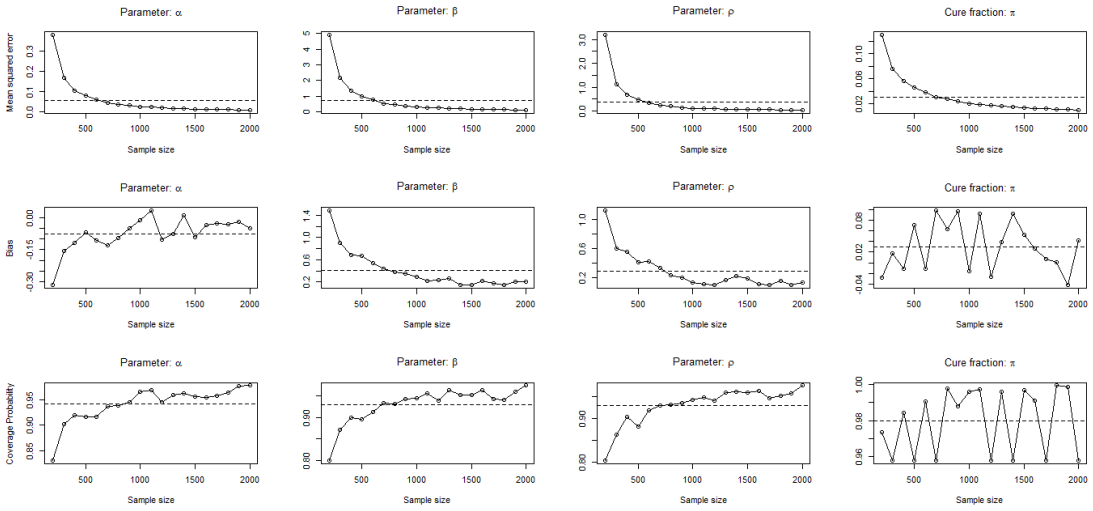

Figure 5.6: Mean squared errors, biases and coverage probabilities of the Bayesian estimates $(\hat{\alpha}, \hat{\beta}, \hat{r}, \hat{\pi})$ versus the sample size n, from the Marshall-Olkin Inverse Gaussian distribution with $(\alpha, \beta, r, \pi) = (-1, 1, 0.5, 0.7616)$. The dashed line represents the average mean squared error, bias and coverage probability, over all simulated sample sizes.

error is obtained from the estimation of the parameter β, while the smallest one, from the estimation of the cure fraction.

- **Bias of the estimates:** We can see that the bias of the estimation procedure is relatively small, while it is seen that as the sample size increases the bias tends to zero for all the distribution parameters, as well as for the cure fraction. The smallest bias is achieved from the estimation of the cure fraction, while the largest one, from the estimation of the parameter β.

- **Coverage probability:** Finally, as we can see, the coverage probability in the estimation procedure for the three parameters is lower than 90% with the smallest one obtained from the estimation of the parameter β, however the coverage probability achieved in the estimation of the cure fraction is very close to 95%. Yet, we can see that as the sample size increases, the coverage probability increases too.

For the last simulation experiment, we simulated 1,000 random samples from the Marshall-Olkin Inverse Gaussian distribution each of sample size $n = 200, \ldots, 2000$ with the following parameter

values $(\alpha, \beta, r, \pi) = (-1, 1, 0.5, 0.7616)$. In **Figure 5.6** we can see the results obtained for the Bayesian estimates of the parameters as in the previous simulation experiments. The conclusions drawn in the case of the Marshall-Olkin Inverse Gaussian distribution are very similar to those drawn in the previous case. More specifically we can see that:

- As the sample size increases, the mean squared error decreases for all the parameters and the cure fraction. In addition, we can see that the lowest mean squared error is achieved from the estimation of the cure fraction, while the largest one, from the estimation of the parameter β.
- As far as the estimation bias is concerned, we can see that it remains at low levels for all the parameters and for the cure fraction. The smallest bias is obtained from the estimation of the cure fraction and the largest one, from the estimation of the parameter β.
- Finally, the coverage probability for all the parameters and the cure fraction is very close to 95%, while as the sample size increases, the coverage probability of the three parameters (α, β and r) increases too.

Chapter 6

Application to real data sets

6.1 Introduction

Cure rate models become more and more popular in modelling survival data from clinical trials concerning various types of cancer, such as breast cancer, prostate cancer, colon cancer, or melanoma. In these diseases a significant proportion of patients is cured and for that reason they can not experience the event of interest, however long the follow-up time is. As already seen, the baseline defective distributions, as well as their extension under the Marshall-Olkin family of distributions, are able to model situations where there are long term survivors. Therefore, in order to understand better their usefulness in modelling such situations, we applied them to two real data sets.

The one dataset includes survival data from a clinical trial concerning melanoma, while the other contains survival data on colon cancer. The choice of these two data sets seems reasonable, since there is a substantial proportion of patients who survive from both types of cancer. We applied both the baseline distributions and their extensions to the data sets in order to model the survival probability of the total sample. Our aim was to find the maximum likelihood and the Bayesian estimates of their parameters, in order to investigate whether they can capture the existence of long term survivors or not, as well as to compare the two estimating methods concerning the parameters' estimates, their respective 95% confidence and credible intervals and

other statistical concepts.

6.2 Description of the data sets

In this section we present the most important information concerning the two analysed data sets, the melanoma data set and the colon cancer data set.

6.2.1 Melanoma data set

The first application of the defective distributions was carried out to the melanoma data set, which is available in R in the library *smcure*. The data set contains information about 285 patients which have to do with the following characteristics:

- The treatment group which they belong to, where the control group is coded as zero and the experimental group (IFN regimen) as one.
- The observed relapse-free time.
- The censoring indicator, where 1 represents the event of interest and 0 the censoring. The event of interest is death from melanoma.
- The participants' age, and
- The participants' gender, 0 for male and 1 for female.

This information comes from the ECOG phase III clinical trial, whose aim was to evaluate the high dose interferon alpha-2b (IFN) regimen, against the placebo as the postoperative adjuvant therapy. More specifically, the researcher was trying to find if the high dose interferon alpha-2b (IFN) regimen group would have higher survival times than the placebo group. It is noted that this data set is widely used in the field of cure rate modelling (Kirkwood, et al., 1996). Below we present an analytical descriptive investigation that we conducted on the data set.

In **Figure 6.1**, the survival curve is shown, not only concerning the total sample of the patients, but also separately according to the participants' gender, as well as, the treatment group to which they belong. As it can be seen, the survival curve presents a plateau almost at 0.30

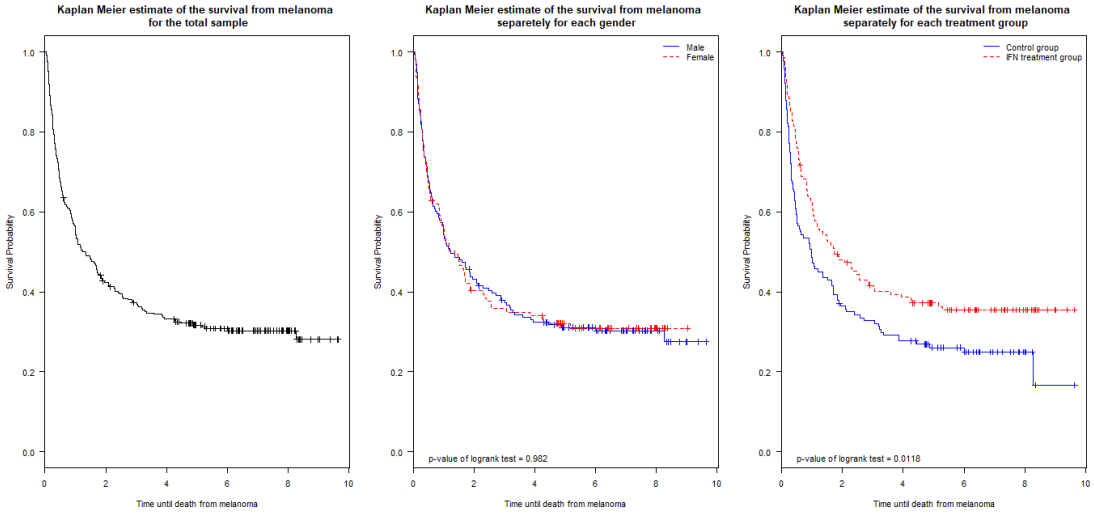

Figure 6.1: Overall survival probability for the melanoma data set, as well as, separately according to the participant gender and treatment group.

($S(t) \approx 0.30$) which indicates the existence of long term survivors in the data set. Furthermore, regarding the total sample it can be seen that the median survival time equals 1.35 with 95% confidence interval of [1.00, 1.85]. In addition, we can see that there is no statistically significant difference in the survival probability between males and females, whereas there is a statistically significant difference between the two treatment groups ($p = 0.0118 < 0.05$). Also:

- The median survival time of males equals 1.24 with 95% C.I [0.95, 2.13] and the median survival time of females equals 1.36 with 95% C.I [0.90, 2.30]. Also, the 25%-ile survival time of males is 0.34 while for females it is 0.36.

- Concerning the two treatment groups it can be seen that the median survival time of the control group equals 0.98 with 95% C.I [0.52, 1.70], whereas for the IFN group, equals 1.73 with 95% C.I [1.15, 3.02]. In addition, it was found that the 25%-ile survival time of the control group is 0.28, while for the IFN treatment group is 0.55.

- There was no statistically significant effect of the participants' age on the risk of death from melanoma ($HR = 1.01$; 95% C.I $= [0.99, 1.02]$; $p = 0.369 > 0.05$).

6.2.2 Colon cancer data set

The second data set used is the colon cancer data set, which is available in R, in the *survival* package. This data set consists of data from one of the first successful trials of adjuvant chemotherapy for colon cancer. Levamisole is a low-toxicity compound previously used to treat worm infestations in animals, while 5-FU is a moderately toxic chemotherapy agent. There are two records per person, one for recurrence and one for death. The study is originally described in Laurie (1989), whereas the main report is found in Moertel (1990). This data set is closer to that of the final report in Moertel (1991) and a version of the data with less follow-up time was used in the paper by Lin (1994). This data set contains the following information:

- Personal identification number
- Treatment group-Observation, Levamisole, Levamisole+5-FU
- Participants' gender.
- Participants' age in years
- Obstruction of colon by tumour
- Perforation of colon
- Adherence to nearby organs
- Number of lymph nodes with detectable cancer
- Days until event or censoring; converted to years
- Censoring status
- Differentiation of tumour
- Extent of local spread
- Time from surgery to registration
- Event type (recurrence, death).

As it is seen, there are two types of event. In the context of the present thesis, we included the part of the data set which had to do with the possible death from colon cancer. Thus, the data

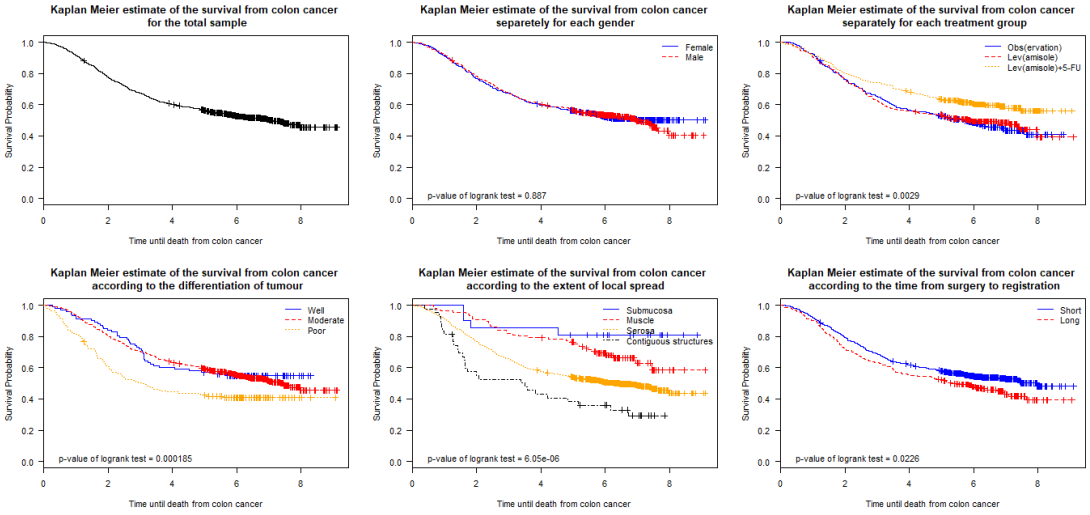

Figure 6.2: Overall survival probability for the colon cancer data set, as well as, separately according to several participant characteristics.

set used throughout this chapter has information about the 929 patients with median follow-up time 5.1 years, concerning only death as the event of interest.

In **Figure 6.2**, the survival curve for the total sample is shown, as well as, separately according to several participants' characteristics. More specifically, we generated these plots for the total sample, as well as according to the participants' gender, treatment group, differentiation of tumour, extent of local spread and the time from surgery to registration. In addition, in the survival curve of the total sample of the participants, we can see that there is a plateau at the end of the curve approximately at 0.4, which could be indicative of the existence of long term survivors in the data set. Furthermore, there is a statistically significant difference in the survival probability according to all participants' characteristics, except for the participants' gender. More specifically the following results were obtained:

- **For the total sample:** The median survival time equals 6.99 years, while the estimated 25%-ile survival time is 2.21 years.

- **According to participants' gender:** It was found that the median survival time of the females could not be calculated, since no woman reached the survival probability of 50%, whereas the respective quantity for males is equal to 6.99 years. Moreover, the estimated

25%-ile survival time for females is 2.2 years, while for males it is 2.3 years.

- **According to the treatment group:** There is a statistically significant difference among the 3 treatment groups. More specifically, it was found that the median survival time for the three treatment groups is equal to 5.7 and 5.9 years for the *Observation* and the *Levamisole* group respectively, whereas for the third group, the median survival time could not be calculated, for the same reason as mentioned above for women. Indeed, the patients belonging in the third treatment group seem to have a much better survival probability compared to the other two treatment groups, which seem to be almost the same.

- **According to the differentiation of tumour:** It can be seen that the patients characterized by poor differentiation of tumour seem to have lower survival probability compared to the other two statuses. More specifically, it was found that the median survival time of the patients equals 7.4 and 3.04 years for the patients with moderate and poor differentiation of their tumour.

- **According to the extent of local spread:** The median survival time for the patients with extent of local spread defined as "Serosa" and as "Contiguous structures" was 6.35 and 3.49 years, respectively.

- **According to the time from surgery to registration:** It was found that the median survival time equals 7.97 and 5.34 years for the patients with short and long time from surgery to registration, respectively. In addition, the difference between the two groups of patients is statistically significant, with those belonging in the short time group having a statistically significant higher survival probability, compared to those in the long time group.

6.3 Maximum likelihood estimates: Applications

In this section we are going to apply the defective distributions introduced in the previous chapters to the data sets described in section 6.2, based on the maximum likelihood estimates of their parameters.

6.3.1 Melanoma data set

As seen in section 6.2, the survival curve for the total sample of the participants, presents a plateau at its right part which indicates the existence of long term survivors in the data set. For that reason we applied all the defective distributions in order to investigate whether there is a statistically significant proportion of cured patients or not. In **Table 6.1** we can see the results after applying the defective distributions to the data set.

Distribution	**Parameters**	**Estimate**	**Lower 95% C.I**	**Upper 95% C.I**	**AIC**
Gompertz	$\hat{\alpha}$	-0.62	-0.75	-0.51	770.4205
	$\hat{\beta}$	0.77	0.63	0.92	
	$\hat{\pi}$	29.11%	19.34%	38.88%	
Inverse Gaussian	$\hat{\alpha}$	-0.31	-0.49	-0.16	747.7463
	$\hat{\beta}$	2.51	2.11	3.03	
	$\hat{\pi}$	22.14%	11.47%	32.81%	
Marshall-Olkin Gompertz	$\hat{\alpha}$	-0.55	-0.74	-0.46	771.1574
	$\hat{\beta}$	0.50	0.38	1.13	
	$\hat{r}$	0.60	0.09	0.66	
	$\hat{\pi}$	28.81%	23.00%	34.63%	
Marshall-Olkin Inverse Gaussian	$\hat{\alpha}$	-0.29	-0.50	-0.07	749.6181
	$\hat{\beta}$	2.63	1.98	3.73	
	$\hat{r}$	1.09	0.67	1.80	
	$\hat{\pi}$	21.19%	10.26%	32.09%	

Table 6.1: **Maximum likelihood estimates for the parameters of the defective distributions in the melanoma data set.**

Based on **Table 6.1**, we can draw the following conclusions:

1. *Existence of long term survivors in the data set.*

   - All distributions estimate the parameter α with a negative value, which means that there are long term survivors in the data set.
   - Both edges of the 95% confidence interval for the parameter α are negative, under

all the defective distributions. This means that the existence of long term survivors is statistically significant, at 5% level of significance.

2. *Comparison between the different distributions.*

   - The proportion of cured patients estimated by the Gompertz distribution and its extension under the Marshall-Olkin family is higher compared to the one estimated by the Inverse Gaussian distribution and its respective extension.
   - These differences do not seem to be statistically significant, due to the overlap of 95% confidence intervals of the cure fraction.

3. *Comparison between the baseline distributions and their extension under the Marshall-Olkin family*

   - There is statistically significant difference between the Gompertz distribution and the Marshall-Olkin Gompertz distribution, since the parameter r differs from unity in a statistically significant manner.
   - However, the Inverse Gaussian distribution and its extension do not differ significantly, since the parameter r of the defective Marshall-Olkin Inverse Gaussian distribution does not differ significantly from unity.

4. *Fit to the data according to the AIC values*

   - According to the AIC values, it seems that the Inverse Gaussian distribution provides the best fit to the data set, with its extension under the Marshall-Olkin family having a similar AIC value. Finally, the Gompertz distribution and its extension under the Marshall-Olkin family, seem to provide a worse fit to the data set.
   - However, it is noted that the Gompertz and the Marshall-Olkin Gompertz distributions estimate the cure fraction at a higher level compared to the other two estimates, which is expected due to the survival curve in **Figure 6.1**.

In addition to the above results of the table, it is worthwhile to see the behaviour of the four distributions in one figure. In **Figure 6.3** we present the fit of the four defective distributions to the melanoma data set, based on the maximum likelihood estimates of their parameters. It is

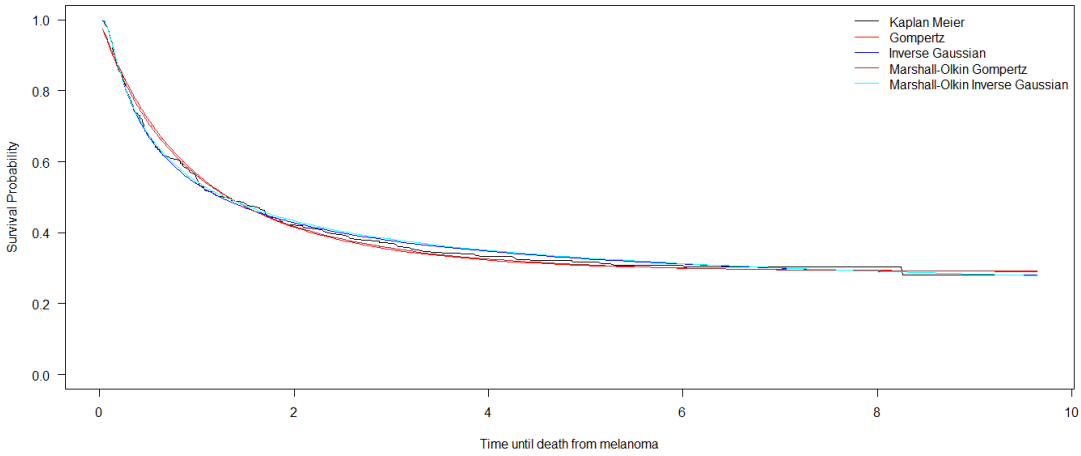

Figure 6.3: Fit of the defective distributions to the melanoma data set, based on the maximum likelihood estimates of their parameters.

observed that all distributions provide a very good fit to the data set, since they are close to the Kaplan Meier estimate and they follow its behaviour at a very satisfactory degree. Furthermore, we can see that there is no difference between the baseline distributions and their extensions under the Marshall-Olkin family, but the different distributions seem to differ significantly, which we have also seen in the previous table. However, in **Table 6.1** we have noted that the Gompertz distribution and its extension are significantly different, something that is not shown in **Figure 6.3**. Finally, compared to the conclusions drawn from the table, we can see that the Gompertz distribution and the Marshall-Olkin Gompertz distribution are closer to the Kaplan Meier curve, although, as mentioned before, these two distrbutions have higher AIC values compared to the Inverse Gaussian and the Marshall-Olkin Inverse Gaussian distributions.

6.3.2 Colon cancer data set

Our second application was carried out to the colon cancer data set, described in section 6.2. **Table 6.2** presents the results obtained after applying the four defective distributions to the data set.

Based on **Table 6.2** the following conclusions can be drawn:

1. *Existence of long term survivors in the data set.*

Distribution	Parameters	Estimate	Lower 95% C.I	Upper 95% C.I	AIC
Gompertz	$\hat{\alpha}$	-0.10	-0.15	-0.05	2917.049
	$\hat{\beta}$	0.14	0.12	0.16	
	$\hat{\pi}$	24.53%	6.67%	42.38%	
Inverse Gaussian	$\hat{\alpha}$	-0.06	-0.09	-0.02	2318.419
	$\hat{\beta}$	0.45	0.40	0.51	
	$\hat{\pi}$	21.97%	10.48%	33.47%	
Marshall-Olkin Gompertz	$\hat{\alpha}$	-0.51	-0.59	-0.42	2872.494
	$\hat{\beta}$	1.89	1.33	2.52	
	$\hat{r}$	36.34	18.10	71.85	
	$\hat{\pi}$	47.31%	43.31%	51.07%	
Marshall-Olkin Inverse Gaussian	$\hat{\alpha}$	0.16	0.07	0.27	2904.892
	$\hat{\beta}$	0.84	0.62	1.21	
	$\hat{r}$	3.92	2.43	6.37	
	$\hat{\pi}$	.	.	.	

Table 6.2: **Maximum likelihood estimates for the parameters of the defective distributions in the colon cancer data set.**

- All defective distributions, except for the Marshall-Olkin Inverse Gaussian distribution, estimate the parameter α with a negative value, which means that there is a cure rate in the data set. The Marshall-Olkin Inverse Gaussian distribution suggests that they do not exist cure patients in this data set.
- The cure rate suggested by the Gompertz, the Inverse Gaussian and the Marshall-Olkin Gompertz distribution is statistically significant, since the 95% confidence interval of the parameter α is fully negative.

2. *Comparison between the different distributions.*

- It is obvious that there is a statistically significant difference between the Gompertz distribution and the Marshall-Olkin Inverse Gaussian distribution, as well as between the Marshall-Olkin Gompertz and the Marshall-Olkin Inverse Gaussian distribution.

The first two distributions estimate a statistically significant cure fraction in the data set, whereas the last one does not suggest the existence of cured patients at all.

- The difference between the Inverse Gaussian distribution and the Marshall-Olkin Gompertz distribution is statistically significant and it seems that they suggest a different proportion of cured patients in the data set.
- However, the baseline distributions do not seem to differ, with the cure fraction suggested by the Gompertz distribution being statistically equal with the respective quantity proposed by the Inverse Gaussian distribution.

3. *Comparison between the baseline distributions and their extension under the Marshall-Olkin family.*

   - The baseline distributions differ from their extension under the Marshall-Olkin family significantly, since the parameter r is significantly different from unity, both in the case of the Marshall-Olkin Inverse Gaussian distribution and in the case of the Marhsall-Olkin Gompertz distribution.
   - The cure fraction proposed by the Gompertz distribution is significantly lower than the one proposed by its extension under the Marshall-Olkin family.
   - As already mentioned, the Marshall-Olkin Inverse Gaussian distribution does not capture the existence of long term survivors in the data set, whereas its baseline form suggests a statistically significant cure fraction.

4. *Fit to the data set according to the AIC values.*

   - The Inverse Gaussian distribution seems to provide the best fit to the colon data set, since it achieves the lowest AIC value. The worst fit seems to be suggested by the Gompertz distribution, while the AIC value of the Marshall-Olkin Inverse Gaussian distribution is close.
   - However, we should note that the Marshall-Olkin Gompertz distribution which provides the second best fit to the data set, suggests a cure fraction which is in the expected range based on the **Figure 6.2**.

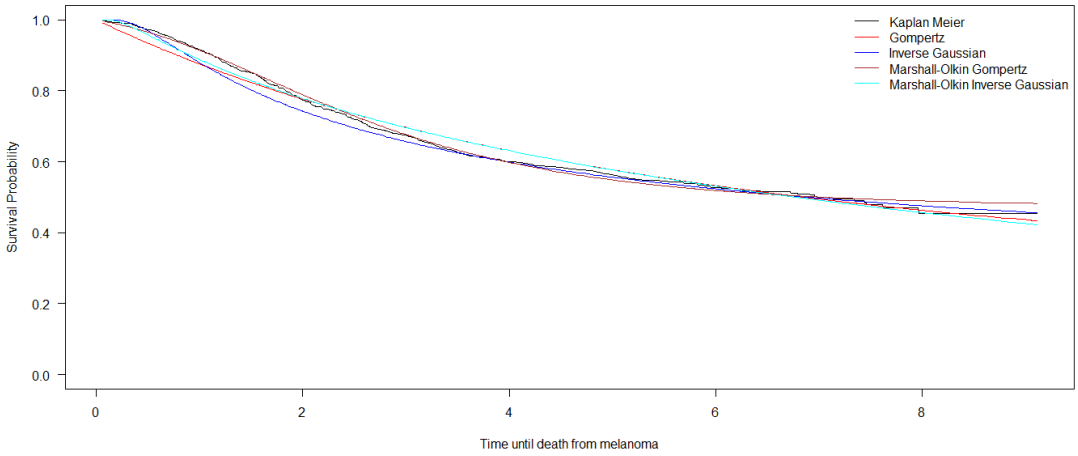

Figure 6.4: Fit of the defective distributions to the colon cancer data set, based on the maximum likelihood estimates of their parameters.

In **Figure 6.4** we present the fit of the four defective distributions to the colon cancer data set. We can see that the Marshall-Olkin Gompertz distribution is very close to the Kaplan Meier curve and it captures its behaviour at a very satisfactory degree. In addition, the Inverse Gaussian distribution seems to capture very well the raw estimate and especially, approximately after the three years of the patients follow up. However, the other two distributions (Gompertz & Marshall-Olkin Inverse Gaussian) do not fit well to the data set. It is seen that the respective coloured lines are not close to the black curve of the Kaplan Meier estimate, representing a lack of fit. Finally, we could state that although according to the AIC values the best fit is proposed by the Inverse Gaussian distribution, in **Figure 6.4** the best fit seems to be succeeded by the Marshall-Olkin Gompertz distribution.

6.4 Bayesian estimates: Applications

After the application of the maximum likelihood estimation procedure, in this section we are going to apply the four defective distributions to the two data sets, based on the Bayesian estimates of their parameters.

6.4.1 Melanoma data set

In the following table we can see the Bayesian estimates of the parameters, after applying the defective distributions to the melanoma data set.

Distribution	**Parameters**	**Estimate**	**Lower 95% C.I**	**Upper 95% C.I**
Gompertz	$\hat{\alpha}$	-0.63	-0.75	-0.50
	$\hat{\beta}$	0.77	0.62	0.93
	$\hat{\pi}$	29.12%	24.07%	34.56%
Inverse Gaussian	$\hat{\alpha}$	-0.32	-0.50	-0.17
	$\hat{\beta}$	2.57	2.11	3.12
	$\hat{\pi}$	22.09%	13.09%	30.77%
Marshall-Olkin Gompertz	$\hat{\alpha}$	-0.58	-0.76	-0.42
	$\hat{\beta}$	0.66	0.16	1.42
	$\hat{r}$	0.89	0.17	2.28
	$\hat{\pi}$	28.57%	24.32%	32.82%
Marshall-Olkin Inverse Gaussian	$\hat{\alpha}$	-0.26	-0.47	-0.06
	$\hat{\beta}$	2.93	2.11	4.19
	$\hat{r}$	1.28	0.75	2.11
	$\hat{\pi}$	19.30%	5.39%	29.70%

Table 6.3: **Bayesian estimates for the parameters of the defective distributions in the melanoma data set. 95% C.I represents the 95% credible interval for the parameters.**

Based on **Table 6.3** the conclusions which are drawn are very similar to those noted in the previous section. In addition, the two estimating methodologies do not seem to differ much, concerning the parameters' estimates as well as the respective confidence and credible intervals, yet they differ on the fact that, while under the maximum likelihood estimates the Gompertz distribution was significantly different from its extension, now they are statistically equal since the parameter r is not significantly different from unity. What is worth noting, is the fact that

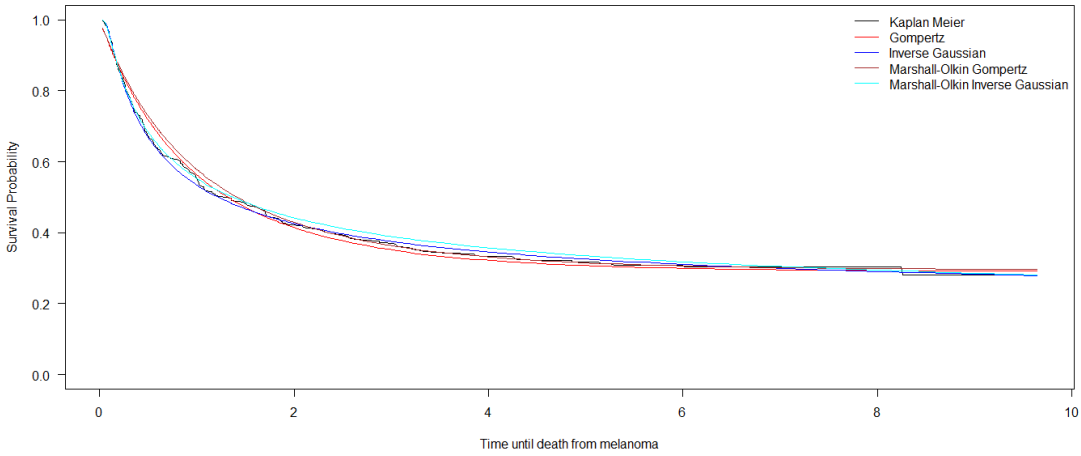

Figure 6.5: Fit of the defective distributions to the melanoma data set, based on the Bayesian estimates of their parameters.

the 95% credible intervals of the cure fraction are narrower compared to the 95% confidence intervals, except for the case of the Marshall-Olkin Inverse Gaussian distribution, where the credible interval is wider than the confidence interval obtained in the previous section. As far as the point estimates are concerned, we can see that the two estimating methods do not differ at all in the case of the defective Gompertz and the defective Inverse Gaussian distributions. However, a tiny difference is observed between the point estimates in the other two distributions. Finally, the point estimates of the cure fraction in all cases are almost the same. We should note that the choice of non-informative prior distribution for the parameters can justify the fact that there is no difference between the two estimation methods, since the inferential "weight" lies on the likelihood function, even in the context of the Bayesian inferential procedure.

In **Figure 6.5** we can see the fit of the defective distributions to the melanoma data set. As it can be seen, all distributions fit very well to the data and they capture the behaviour of the Kaplan Meier estimate at a very satisfactory degree. In addition, a small difference is observed between the baseline distributions and their extension, yet as mentioned before, this difference is not statistically significant.

6.4.2 Colon cancer data set

In **Table 6.4** we present the Bayesian estimates for the parameters of the defective distributions after applying them to the colon cancer data set. The conclusions drawn from **Table 6.4**

Distribution	**Parameters**	**Estimate**	**Lower 95% C.I**	**Upper 95% C.I**
Gompertz	$\hat{\alpha}$	-0.10	-0.15	-0.06
	$\hat{\beta}$	0.14	0.12	0.16
	$\hat{\pi}$	24.53%	10.89%	36.13%
Inverse Gaussian	$\hat{\alpha}$	-0.06	-0.09	-0.03
	$\hat{\beta}$	0.45	0.40	0.51
	$\hat{\pi}$	21.97%	11.52%	31.35%
Marshall-Olkin Gompertz	$\hat{\alpha}$	-0.51	-0.58	-0.44
	$\hat{\beta}$	1.94	1.48	2.47
	$\hat{r}$	40.21	21.68	69.28
	$\hat{\pi}$	46.91%	42.66%	50.24%
Marshall-Olkin Inverse Gaussian	$\hat{\alpha}$	-0.011	-0.014	-0.003
	$\hat{\beta}$	0.61	0.57	0.66
	$\hat{r}$	1.76	1.60	1.94
	$\hat{\pi}$	5.41%	0.30%	19.93%

Table 6.4: **Bayesian estimates for the parameters of the defective distributions in the colon cancer data set.**

are similar to those mentioned in section 6.3.2 for all the defective distributions, except for the Marshall-Olkin Inverse Gaussian. As it can be seen, based on the Bayesian inferential procedure, the Marshall-Olkin Inverse Gaussian distribution suggests the existence of a small proportion of long term survivors in the data set, whereas based on the maximum likelihood estimates, the same distribution estimated the parameter α with a positive value meaning that there are no cured patients in the data set.

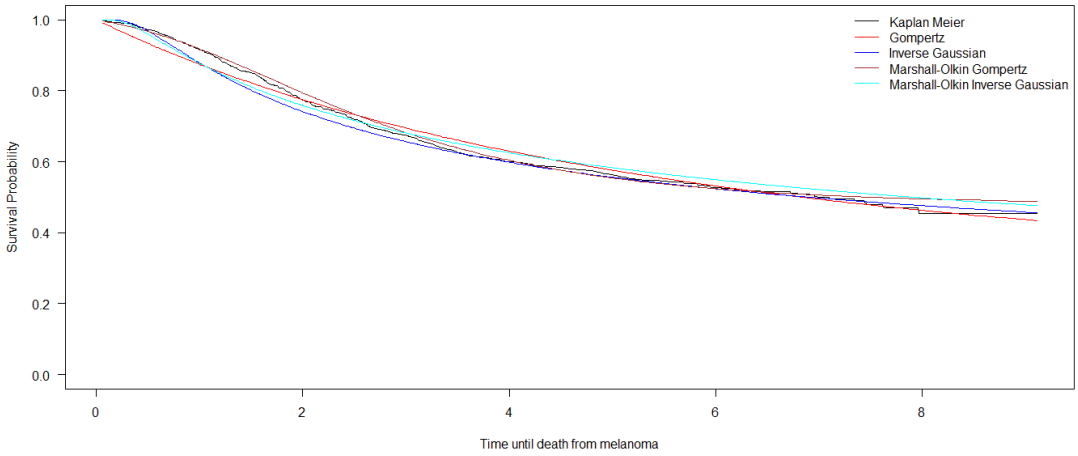

Figure 6.6: Fit of the defective distributions to the colon cancer data set, based on the Bayesian estimates of their parameters.

Regarding the point estimates for the parameters of the baseline distributions, we can see that there are no differences between the two estimating methods. However, there is a tiny difference in the estimates for the parameters of the Marshall-Olkin Gompertz distribution. Concerning the confidence and the credible intervals of the parameters, it is observed in the majority of the parameters, that the 95% credible intervals are narrower compared to the respective 95% confidence intervals, which means that more confidence is added to the results of the Bayesian estimation procedure.

Finally, in **Figure 6.6** the fit of the defective distributions can be seen. Based on this figure we can see that the Marshall-Olkin Gompertz distribution, as well as the Inverse Gaussian distribution, are closer to the Kaplan Meier curve providing a better fit compared to the other two defective distributions. As in the case of the maximum likelihood estimates, it is seen that the the Gompertz distribution and the Marshall-Olkin Inverse Gaussian distribution can not capture the behaviour of the Kaplan Meier estimate at a satisfactory degree, leading us to conclude that they do not fit well to the colon cancer data set. Finally it should be noted that again the Marshall-Olkin Gompertz distribution estimates the cure fraction in an expected range, whereas the other three distributions lead to much lower proportion of cured patients.

Chapter 7

Conclusions

In this thesis, we considered the Bayesian approach to inference for various defective distributions, recently introduced by Rocha (2016), in the field of cure rate modelling in survival analysis. In addition, the new Bayesian estimates for the parameters were compared to the respective maximum likelihood estimates. Specifically, we carried out Bayesian estimation for the parameters of the defective Gompertz distribution, which was already known in the literature. Furthermore, we also provided Bayesian estimates for the parameters of the defective Inverse Gaussian distribution, as well as for the Gompertz and the Inverse Gaussian distribution under the Marshall-Olkin family of distributions.

The present thesis started in a more general context, presenting in the 1^{st} chapter the general principles in the field of survival analysis. In this chapter, the basic functions used in survival analysis and the concepts of censoring and truncation were presented, as well as, the basic parametric, semi-parametric and non-parametric models which are widely known in the field of biostatistics.

Subsequently, in the 2^{nd} chapter, the general concepts of the Bayesian approach to inference were presented. More specifically, various ways of selecting prior distributions were presented, while the basic algorithms used for Bayesian inference were described, as well.

After discussing the general principles of survival analysis and the respective principles concerning the Bayesian approach to inference, we moved on to the 3^{rd} chapter, where the cure rate models were introduced. In this chapter, an extensive literature review was presented concerning the most widely used cure rate models, such as the mixture cure rate model, the proportional hazards and proportional odds cure rate models, as well as, the transformation cure rate models. Finally, we presented some of the most prominent works, based on which the whole concept of cure rate models has been extended.

The final step of the present thesis was the introduction of the defective models for cure rate modelling. More specifically, in the fourth chapter the defective Gompertz and the defective Inverse Gaussian distributions were presented. Through various simulation scenarios we evaluated the performance of the Bayesian estimates for their parameters and we compared them to the maximum likelihood estimates. In the fifth chapter, the extension of those distributions under the Marshall-Olkin family was discussed. As in the previous chapter, various simulation experiments were carried out in order to obtain new Bayesian estimates for the parameters of the defective distributions. The behaviour of these models, based both on maximum likelihood estimates and Bayesian estimates of the parameters, was evaluated by applying them to the melanoma and the colon cancer data sets, presented in the last chapter.

What the present thesis adds

The present thesis submitted in partial fulfilment of the requirements for the degree of Master in Biostatistics, is the first attempt to Bayesian inference for these distributions. First of all, a new approach was introduced for obtaining the Bayesian estimates for the parameters of the defective Gompertz distribution, by assuming the Uniform distribution as a prior distribution for the parameters, compared to the one presented by MR dos Santos et.al. (2017), who used the Gamma and the Inverse Gamma distribution as prior distributions of the parameters. In addition, we are the first who obtained the Bayesian estimates for the parameters of the defective Inverse Gaussian distribution, as well as for the parameters of the extension of those distributions under the Marshall-Olkin family.

Future work

Next, we intend to introduce several concepts in the framework of defective distributions for cure rate modelling, some of which are listed below:

- We are going to obtain new Bayesian estimates for the parameters of the defective Gompertz and the defective Inverse Gaussian distribution, under the Kumaraswamy family of distributions.

- The next step is to incorporate covariate information in the defective models considered here. By exploiting covariate information one can gain useful insights about each particular individual in the study, regarding its survival and cure rate.

- Furthermore, we will consider the problem of variable selection in the context of the regression defective cure rate models under consideration. Our main interest is to develop new, Bayesian methods for the variable selection problem in the context of defective cure rate models, not only when the number of parameters is lower than the number of observations, but also when the number of parameters is large.

Bibliography

Abrams, K., Ashby, D., and Errington, D. (1996). A bayesian approach to weibull survival models•application to a cancer clinical trial. *Lifetime Data Analysis*, 2(2):159–174.

Achcar, J. A., Coelho-Barros, E. A., and Mazucheli, J. (2012). Cure fraction models using mixture and non-mixture models. *Tatra Mountains Mathematical Publications*, 51(1):1–9.

AHMED, A. M. (2014). Markov chain monte carlo and lindley"s approximation for estimation weibull distribution based on right censored data. *International Journal*, 2(10):1069–1086.

Akerkar, R., Martino, S., and Rue, H. (2010). Implementing approximate bayesian inference for survival analysis using integrated nested laplace approximations. *Preprint Statistics, Norwegian University of Science and Technology*, 1:1–38.

Al-Jarallah, R. and Al-Hussaini, E. (2007). Bayes inference under a finite mixture of two-compound gompertz components model. *Journal of Statistical Computation and Simulation*, 77(11):915–927.

Al-Shomrani, A. A., Shawky, A., Arif, O. H., and Aslam, M. (2016). Log-logistic distribution for survival data analysis using mcmc. *SpringerPlus*, 5(1):1774.

Amico, M. and Van Keilegom, I. (2017). Cure models in survival analysis. *Annual Review of Statistics and Its Application*, (0).

Ananda, M. M., Dalpatadu, R. J., and Singh, A. K. (1996). Adaptive bayes estimators for parameters of the gompertz survival model. *Applied Mathematics and Computation*, 75(2-3):167–177.

Andrew, G. and Gao, J. (2007). Scalable training of l 1-regularized log-linear models. In *Proceedings of the 24th international conference on Machine learning*, pages 33–40. ACM.

Annest, A., Bumgarner, R. E., Raftery, A. E., and Yeung, K. Y. (2010). The iterative bayesian model averaging algorithm for survival analysis: an improved method for gene selection and survival analysis on microarray data.

ANTONIO MOALA, F., LUIZ RAMOS, P., and ALBERTO ACHCAR, J. (2013). Bayesian inference for two-parameter gamma distribution assuming different noninformative priors. *Revista Colombiana de Estadística*, 36(2):319–336.

Apolloni, B. and Bassis, S. (2009). Algorithmic inference of two-parameter gamma distribution. *Communications in Statistics-Simulation and Computation*, 38(9):1950–1968.

Arbel, J., Lijoi, A., and Nipoti, B. (2016). Full bayesian inference with hazard mixture models. *Computational Statistics & Data Analysis*, 93:359–372.

Arjas, E. and Gasbarra, D. (1994). Nonparametric bayesian inference from right censored survival data, using the gibbs sampler. *Statistica sinica*, pages 505–524.

Baghestani, A. R., Moghaddam, S. S., Majd, H. A., Akbari, M. E., Nafissi, N., and Gohari, K. (2015). Application of a non-mixture cure rate model for analyzing survival of patients with breast cancer. *Asian Pac J Cancer Prev*, 16(16):7359–63.

Balakrishnan, N., Barui, S., and Milienos, F. (2017). Proportional hazards under conway–maxwell-poisson cure rate model and associated inference. *Statistical methods in medical research*, 26(5):2055–2077.

Balakrishnan, N., Koutras, M., Milienos, F., and Pal, S. (2016). Piecewise linear approximations for cure rate models and associated inferential issues. *Methodology and Computing in Applied Probability*, 18(4):937–966.

Banerjee, S. and Carlin, B. P. (2004). Parametric spatial cure rate models for interval-censored time-to-relapse data. *Biometrics*, 60(1):268–275.

Basak, I. and Balakrishnan, N. (2012). Estimation for the three-parameter gamma distribution based on progressively censored data. *Statistical Methodology*, 9(3):305–319.

Bennett, S. (1983). Log-logistic regression models for survival data. *Applied Statistics*, pages 165–171.

Berger, J. O. and Bernardo, J. M. (1989). Estimating a product of means: Bayesian analysis with reference priors. *Journal of the American Statistical Association*, 84(405):200–207.

Berkson, J. and Gage, R. P. (1952). Survival curve for cancer patients following treatment. *Journal of the American Statistical Association*, 47(259):501–515.

Bernardo, J. M. (1979). Reference posterior distributions for bayesian inference. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 113–147.

Bloomfield, C. D., Shuma, C., Regal, L., Philip, P. P., Hossfeld, D. K., Hagemeijer, A. M., Garson, O. M., Peterson, B. A., Sakurai, M., Alimena, G., et al. (1997). Long-term survival of patients with acute myeloid leukemia. *Cancer*, 80(S11):2191–2198.

Boag, J. W. (1949). Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *Journal of the Royal Statistical Society. Series B (Methodological)*, 11(1):15–53.

Borges, P. (2017). Em algorithm-based likelihood estimation for a generalized gompertz regression model in presence of survival data with long-term survivors: an application to uterine cervical cancer data. *Journal of Statistical Computation and Simulation*, 87(9):1712–1722.

Box, G. E. and Tiao, G. C. (2011). *Bayesian inference in statistical analysis*, volume 40. John Wiley & Sons.

Broët, P., Rycke, Y., Tubert-Bitter, P., Lellouch, J., Asselain, B., and Moreau, T. (2001). A semiparametric approach for the two-sample comparison of survival times with long-term survivors. *Biometrics*, 57(3):844–852.

Brown, E. R. and Ibrahim, J. G. (2003). Bayesian approaches to joint cure-rate and longitudinal models with applications to cancer vaccine trials. *Biometrics*, 59(3):686–693.

Burnham, K. P. and Anderson, D. R. (2003). *Model selection and multimodel inference: a practical information-theoretic approach*. Springer Science & Business Media.

Byrd, R. H., Lu, P., Nocedal, J., and Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16(5):1190–1208.

Byrd, R. H., Nocedal, J., and Schnabel, R. B. (1994). Representations of quasi-newton matrices and their use in limited memory methods. *Mathematical Programming*, 63(1-3):129–156.

Cancho, V. G., Rodrigues, J., and de Castro, M. (2011). A flexible model for survival data with a cure rate: a bayesian approach. *Journal of Applied Statistics*, 38(1):57–70.

Casella, G. and George, E. I. (1992). Explaining the gibbs sampler. *The American Statistician*, 46(3):167–174.

Casellas, J. (2007). Bayesian inference in a piecewise weibull proportional hazards model with unknown change points. *Journal of Animal Breeding and Genetics*, 124(4):176–184.

Chen, M.-H., Harrington, D. P., and Ibrahim, J. G. (2002). Bayesian cure rate models for malignant melanoma: a case-study of eastern cooperative oncology group trial e1690. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 51(2):135–150.

Chen, M.-H., Ibrahim, J. G., and Sinha, D. (1999). A new bayesian model for survival data with a surviving fraction. *Journal of the American Statistical Association*, 94(447):909–919.

Chen, T. (2015). *Cure Rate Models with Nonparametric Form of Covariate Effects*. PhD thesis, Virginia Tech.

Clayton, D. G. (1978). 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.

Collett, D. (2015). *Modelling survival data in medical research*. CRC press.

Cooner, F., Banerjee, S., Carlin, B. P., and Sinha, D. (2007). Flexible cure rate modeling under latent activation schemes. *Journal of the American Statistical Association*, 102(478):560–572.

Corbière, F., Commenges, D., Taylor, J. M., and Joly, P. (2009). A penalized likelihood approach for mixture cure models. *Statistics in medicine*, 28(3):510–524.

Cordeiro, G. M., Ortega, E. M., and Silva, G. O. (2011). The exponentiated generalized gamma distribution with application to lifetime data. *Journal of statistical computation and simulation*, 81(7):827–842.

Cox, D. R. (2018). *Analysis of survival data*. Routledge.

da ROCHA, R. F., TOMAZELLA, V. L. D., and LOUZADA, F. (2014). Inferência classica e bayesiana para o modelo de frac ao de cura gompertz defeituoso. *Rev. Bras. Biom*, 32(1):104–114.

Danish, M. Y. and Aslam, M. (2014). Bayesian inference for the randomly censored weibull distribution. *Journal of Statistical Computation and Simulation*, 84(1):215–230.

Datta, A. (2013). *A Study of the Cure Rate Model with Case Weights and Time-Dependent Weights*. PhD thesis.

De Castro, M., Cancho, V. G., and Rodrigues, J. (2010). A hands-on approach for fitting long-term survival models under the gamlss framework. *computer methods and programs in biomedicine*, 97(2):168–177.

de Souza, D., Cancho, V. G., Rodrigues, J., and Balakrishnan, N. (2017). Bayesian cure rate models induced by frailty in survival analysis. *Statistical methods in medical research*, 26(5):2011–2028.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (methodological)*, pages 1–38.

Devroye, L. (1984). A simple algorithm for generating random variates with a log-concave density. *Computing*, 33(3-4):247–257.

Dobson, A. J. and Barnett, A. (2008). *An introduction to generalized linear models*. CRC press.

dos Santos, M. R., Achcar, J. A., and Martinez, E. Z. (2017). Bayesian and maximum likelihood inference for the defective gompertz cure rate model with covariates. *Ciência e Natura*, 39(2):244.

El-Din, M. M., Abdel-Aty, Y., and Abu-Moussa, M. (2017). Statistical inference for the gompertz distribution based on type-ii progressively hybrid censored data. *Communications in Statistics-Simulation and Computation*, 46(8):6242–6260.

ELANGOVAN, D. and JAYAKUMAR, B. Cure rate models.

Farewell, V. T. (1982). The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics*, pages 1041–1046.

Farewell, V. T. (1986). Mixture models in survival analysis: Are they worth the risk? *Canadian Journal of Statistics*, 14(3):257–262.

FerreirA, L. A. and SiLvA, J. L. (2017). Parameter estimation for weibull distribution with right censored data using em algorithm zastosowanie algorytmu maksymalizacji wartości oczekiwanej do estymacji parametrów rozkładu weibulla w przypadku danych obcietych prawostronnie. *EKSPLOATACJA I NIEZAWODNOSC*, 19(2):310.

Finkelstein, D. M. and Wolfe, R. A. (1985). A semiparametric model for regression analysis of interval-censored failure time data. *Biometrics*, pages 933–945.

Gallardo, D. I., Gómez, H. W., and Bolfarine, H. (2017). A new cure rate model based on the yule–simon distribution with application to a melanoma data set. *Journal of Applied Statistics*, 44(7):1153–1164.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian data analysis*. CRC press.

Geman, S. and Geman, D. (1987). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. In *Readings in Computer Vision*, pages 564–584. Elsevier.

George, D. and George, S. (2013). Marshall-olkin esscher transformed laplace distribution and processes. *Brazilian Journal of Probability and Statistics*, pages 162–184.

Ghitany, M., Al-Awadhi, F., and Alkhalfan, L. (2007). Marshall–olkin extended lomax distribution and its application to censored data. *Communications in Statistics•Theory and Methods*, 36(10):1855–1866.

Ghitany, M., Al-Hussaini, E., and Al-Jarallah, R. (2005). Marshall–olkin extended weibull distribution and its application to censored data. *Journal of Applied Statistics*, 32(10):1025–1034.

Ghitany, M., Al-Mutairi, D., Al-Awadhi, F., and Al-Burais, M. (2012). Marshall-olkin extended lindley distribution and its application. *International Journal of Applied Mathematics*, 25(5):709–721.

Ghitany, M., Maller, R. A., and Zhou, S. (1994). Exponential mixture models with long-term survivors and covariates. *Journal of multivariate Analysis*, 49(2):218–241.

Gilks, W. R. and Wild, P. (1992). Adaptive rejection sampling for gibbs sampling. *Applied Statistics*, pages 337–348.

Goldman, A. I. (1984). Survivorship analysis when cure is a possibility: a monte carlo study. *Statistics in Medicine*, 3(2):153–163.

Gu, Y., Sinha, D., and Banerjee, S. (2011). Analysis of cure rate survival data under proportional odds model. *Lifetime data analysis*, 17(1):123–134.

Gui, W. (2013). Marshall-olkin extended log-logistic distribution and its application in minification processes. *Appl Math Sci*, 7(80):3947–3961.

Guo, S. and Zeng, D. (2014). An overview of semiparametric models in survival analysis. *Journal of Statistical Planning and Inference*, 151:1–16.

Gupta, C. C. H. (2016). *Bayesian inference and novel models for survival data with cured fraction*. PhD thesis, The Florida State University.

Gupta, R. C., Akman, O., and Lvin, S. (1999). A study of log-logistic model in survival analysis. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 41(4):431–443.

Gupta, R. C., Kannan, N., and Raychaudhuri, A. (1997). Analysis of lognormal survival data. *Mathematical biosciences*, 139(2):103–115.

Guure, C. B., Ibrahim, N. A., and Ahmed, A. O. M. (2012). Bayesian estimation of two-parameter weibull distribution using extension of jeffreys' prior information with three loss functions. *Mathematical Problems in Engineering*, 2012.

Hanagal, D. D. and Sharma, R. (2015). Bayesian inference in marshall–olkin bivariate exponential shared gamma frailty regression model under random censoring. *Communications in Statistics-Theory and Methods*, 44(1):24–47.

Herring, A. H. and Ibrahim, J. G. (2002). Maximum likelihood estimation in random effects cure rate models with nonignorable missing covariates. *Biostatistics*, 3(3):387–405.

Ibrahim, J. G., Chen, M.-H., and Sinha, D. (2013). *Bayesian survival analysis*. Springer Science & Business Media.

Ismail, A. A. (2010). Bayes estimation of gompertz distribution parameters and acceleration factor under partially accelerated life tests with type-i censoring. *Journal of Statistical Computation and Simulation*, 80(11):1253–1264.

Jackson, R. (2015). *Some statistical methods for the analysis of survival data in cancer clinical trials*. PhD thesis, University of Liverpool.

Jaynes, E. T. (1968). Prior probabilities. *IEEE Transactions on systems science and cybernetics*, 4(3):227–241.

Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, pages 453–461.

Jeffreys 3rd, H. Theory of probability 3rd edn 1961 clarendon press.

Jose, K., Joseph, A., and Ristić, M. M. (2009). A marshall-olkin beta distribution and its applications. *Journal of Probability and Statistical Science*, 7(2):173–186.

Jose, K. and Krishna, E. (2011). Marshall-olkin extended uniform distribution. In *ProbStat Forum*, volume 4, pages 78–88.

Jose, K., Naik, S. R., and Ristić, M. M. (2010). Marshall–olkin q-weibull distribution and max–min processes. *Statistical papers*, 51(4):837–851.

Jose, K. K. and Sebastian, R. (2013). Marshall–olkin morgenstern–weibull distribution: generalisations and applications. *Economic Quality Control*, 28(2):105–116.

Kannan, N., Kundu, D., Nair, P., and Tripathi, R. (2010). The generalized exponential cure rate model with covariates. *Journal of Applied Statistics*, 37(10):1625–1636.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282):457–481.

Kim, H. T. and Gray, R. (2012). Three-component cure rate model for nonproportional hazards alternative in the design of randomized clinical trials. *Clinical Trials*, 9(2):155–163.

Kim, S., Chen, M.-H., Dey, D. K., and Gamerman, D. (2007). Bayesian dynamic models for survival data with a cure fraction. *Lifetime Data Analysis*, 13(1):17–35.

Kim, S., Xi, Y., Chen, M.-H., et al. (2009). A new latent cure rate marker model for survival data. *The Annals of Applied Statistics*, 3(3):1124–1146.

Kim, Y.-J. and Jhun, M. (2008). Cure rate model with interval censored data. *Statistics in medicine*, 27(1):3–14.

Klein, J. P. and Moeschberger, M. L. (2005). *Survival analysis: techniques for censored and truncated data*. Springer Science & Business Media.

Koutras, M. and Milienos, F. (2017). A flexible family of transformation cure rate models. *Statistics in medicine*, 36(16):2559–2575.

Krishna, E., Jose, K., Alice, T., and Ristić, M. M. (2013). The marshall-olkin fréchet distribution. *Communications in Statistics-Theory and Methods*, 42(22):4091–4107.

Krishna, E. and Jose, K. K. (2011). Marshall-olkin generalized asymmetric laplace distributions and processes. *Statistica*, 71(4):453–467.

Kuk, A. Y. and Chen, C.-H. (1992). A mixture model combining logistic regression with proportional hazards regression. *Biometrika*, 79(3):531–541.

Kundu, D. (2008). Bayesian inference and life testing plan for the weibull distribution in presence of progressive censoring. *Technometrics*, 50(2):144–154.

Kundu, D. and Gupta, A. K. (2013). Bayes estimation for the marshall–olkin bivariate weibull distribution. *Computational Statistics & Data Analysis*, 57(1):271–281.

Kundu, D. and Mitra, D. (2016). Bayesian inference of weibull distribution based on left truncated and right censored data. *Computational Statistics & Data Analysis*, 99:38–50.

Kundu, D. and Raqab, M. Z. (2012). Bayesian inference and prediction of order statistics for a type-ii censored weibull distribution. *Journal of statistical planning and inference*, 142(1):41–47.

Lagakos, S. (1979). General right censoring and its impact on the analysis of survival data. *Biometrics*, pages 139–156.

Lam, K., Fong, D. Y., and Tang, O. (2005). Estimating the proportion of cured patients in a censored sample. *Statistics in medicine*, 24(12):1865–1879.

Laurie, J. A., Moertel, C., Fleming, T., Wieand, H., Leigh, J., Rubin, J., McCormack, G., Gerstner, J., Krook, J., and Malliard, J. (1989). Surgical adjuvant therapy of large-bowel carcinoma: an evaluation of levamisole and the combination of levamisole and fluorouracil. the north central cancer treatment group and the mayo clinic. *Journal of Clinical Oncology*, 7(10):1447–1456.

Lawless, J. F. (2011). *Statistical models and methods for lifetime data*, volume 362. John Wiley & Sons.

Lee, D. (2013). Carbayes: An r package for bayesian spatial modeling with conditional autoregressive priors. *Journal of Statistical Software*, 55(13):1–24.

Lee, E. T. and Go, O. T. (1997). Survival analysis in public health research. *Annual review of public health*, 18(1):105–134.

Leung, K.-M., Elashoff, R. M., and Afifi, A. A. (1997). Censoring issues in survival analysis. *Annual review of public health*, 18(1):83–104.

Lin, D. (1994). Cox regression analysis of multivariate failure time data: the marginal approach. *Statistics in medicine*, 13(21):2233–2247.

Liu, D. C. and Nocedal, J. (1989). On the limited memory bfgs method for large scale optimization. *Mathematical programming*, 45(1-3):503–528.

Liu, J. S. (1994). The collapsed gibbs sampler in bayesian computations with applications to a gene regulation problem. *Journal of the American Statistical Association*, 89(427):958–966.

López-Cheda, A., Jácome, M. A., and Cao, R. (2017). Nonparametric latency estimation for mixture cure models. *TEST*, 26(2):353–376.

Louzada, F., Oliveira Jr, M. R. d., and Moreira, F. F. (2015). The zero-inflated promotion cure rate regression model applied to fraud propensity in bank loan applications. *arXiv preprint arXiv:1510.00443*.

Mahani, A. S. and Sharabiani, M. T. (2014). Multivariate-from-univariate mcmc sampler: R package mfusampler. *arXiv preprint arXiv:1412.7784*.

Maller, R. A. and Zhou, S. (1994). Testing for sufficient follow-up and outliers in survival data. *Journal of the American Statistical Association*, 89(428):1499–1506.

Malouf, R. (2002). A comparison of algorithms for maximum entropy parameter estimation. In *proceedings of the 6th conference on Natural language learning-Volume 20*, pages 1–7. Association for Computational Linguistics.

Mandouh, R. (2016). Bayesian inference from the kumaraswamy-weibull distribution with applications to real data. *International Journal of Contemporary Mathematical Sciences*, 11(3):119–129.

Marin, J.-M. and Robert, C. (2007). *Bayesian core: a practical approach to computational Bayesian statistics*. Springer Science & Business Media.

Marshall, A. W. and Olkin, I. (1997). A new method for adding a parameter to a family of distributions with application to the exponential and weibull families. *Biometrika*, 84(3):641–652.

Marshall, A. W. and Olkin, I. (2015). A bivariate gompertz–makeham life distribution. *Journal of Multivariate Analysis*, 139:219–226.

Martinez, E. Z. and Achcar, J. A. (2018). A new straightforward defective distribution for survival analysis in the presence of a cure fraction. *Journal of Statistical Theory and Practice*, (just-accepted).

Matthies, H. and Strang, G. (1979). The solution of nonlinear finite element equations. *International journal for numerical methods in engineering*, 14(11):1613–1626.

Missov, T. I. and Lenart, A. (2013). Gompertz–makeham life expectancies: Expressions and applications. *Theoretical population biology*, 90:29–35.

Mitra, D. (2013). *Likelihood inference for left truncated and right censored lifetime data*. PhD thesis.

Moala, F. (2010). Bayesian analysis for the weibull parameters by using noninformative prior distributions. *Advances and Applications in Statistics*, 14(2):117–143.

Moertel, C. G., Fleming, T. R., Macdonald, J. S., Haller, D. G., Laurie, J. A., Goodman, P. J., Ungerleider, J. S., Emerson, W. A., Tormey, D. C., Glick, J. H., et al. (1990). Levamisole and fluorouracil for adjuvant therapy of resected colon carcinoma. *New England Journal of Medicine*, 322(6):352–358.

Moertel, C. G., Fleming, T. R., Macdonald, J. S., Haller, D. G., Laurie, J. A., Tangen, C. M., Ungerleider, J. S., Emerson, W. A., Tormey, D. C., Glick, J. H., et al. (1995). Fluorouracil plus levamisole as effective adjuvant therapy after resection of stage iii colon carcinoma: a final report. *Annals of internal medicine*, 122(5):321–326.

Mohammed A, A. O. and Ibrahim, N. A. (2011). Bayesian survival estimator for weibull distribution with censored data. *Journal of applied Sciences*, 11:393–396.

Mokhtari, A. and Ribeiro, A. (2014). Res: Regularized stochastic bfgs algorithm. *IEEE Transactions on Signal Processing*, 62(23):6089–6104.

Mokhtari, A. and Ribeiro, A. (2015). Global convergence of online limited memory bfgs. *Journal of Machine Learning Research*, 16(1):3151–3181.

Moriña, D., Navarro, A., et al. (2014). The r package survsim for the simulation of simple and complex survival data. *Journal of Statistical Software*, 59(2):1–20.

Moslemi, A., Mahjub, H., Saidijam, M., Poorolajal, J., and Soltanian, A. R. (2016). Bayesian survival analysis of high-dimensional microarray data for mantle cell lymphoma patients. *Asian Pac J Cancer Prev*, 17(1):95–100.

Mullen, K. M., Ardia, D., Gil, D. L., Windover, D., and Cline, J. (2009). Deoptim: An r package for global optimization by differential evolution.

Nagatsuka, H. and Balakrishnan, N. (2012). A consistent parameter estimation in the three-parameter lognormal distribution. *Journal of Statistical Planning and Inference*, 142(7):2071–2086.

Nash, J. C. et al. (2014). On best practice optimization methods in r. *Journal of Statistical Software*, 60(2):1–14.

Nieto-Barajas, L. E. and Yin, G. (2009). Bayesian cure rate model accommodating multiplicative and additive covariates. *Statistics and Its Interface*, 2(4):513–521.

Nocedal, J. (1980). Updating quasi-newton matrices with limited storage. *Mathematics of computation*, 35(151):773–782.

O'Hara, R. B., Sillanpää, M. J., et al. (2009). A review of bayesian variable selection methods: what, how and which. *Bayesian analysis*, 4(1):85–117.

Ortega, E. M., Cordeiro, G. M., Suzuki, A. K., and Ramires, T. G. (2017). A new extended birnbaum-saunders model with cure fraction: classical and bayesian approach. *Communications for Statistical Applications and Methods*, 24(4):397–419.

Overduin, S. (2004). *Use of the lognormal distribution for survival data: inference and robustness*. PhD thesis, Department of Statistics and Actuarial Science-Simon Fraser University.

Ozel, G., Alizadeh, M., Cakmakyapan, S., Hamedani, G., Ortega, E. M., and Cancho, V. G. (2017). The odd log-logistic lindley poisson model for lifetime data. *Communications in Statistics-Simulation and Computation*, 46(8):6513–6537.

Pal, S. and Balakrishnan, N. (2015). Likelihood inference based on em algorithm for the destructive length-biased poisson cure rate model with weibull lifetime. *Communications in Statistics-Simulation and Computation*, (just-accepted).

Pal, S. and Balakrishnan, N. (2017a). An em type estimation procedure for the destructive exponentially weighted poisson regression cure model under generalized gamma lifetime. *Journal of Statistical Computation and Simulation*, 87(6):1107–1129.

Pal, S. and Balakrishnan, N. (2017b). Expectation maximization algorithm for box-cox transformation cure rate model and assessment of model mis-specication under weibull lifetimes. *IEEE journal of biomedical and health informatics*.

Patti, S., Biganzoli, E., and Boracchi, P. (2007). Review of the maximum likelihood functions for right censored data. a new elementary derivation. *COBRA Preprint Series*, page 21.

Peng, Y. and Taylor, J. M. (2011). Mixture cure model with random effects for the analysis of a multi-center tonsil cancer study. *Statistics in medicine*, 30(3):211–223.

Pérez-Casany, M. and Casellas, A. (2013). Marshall-olkin extended zipf distribution. *arXiv preprint arXiv:1304.4540*.

Perperoglou, A. (2011). Fitting survival data with penalized poisson regression. *Statistical Methods & Applications*, 20(4):451–462.

Perros, H. G. (2009). Computer simulation techniques: The definitive introduction!

Peto, R. (1973). Experimental survival curves for interval-censored data. *Applied Statistics*, pages 86–91.

Pradhan, B. and Kundu, D. (2011). Bayes estimation and prediction of the two-parameter gamma distribution. *Journal of Statistical Computation and Simulation*, 81(9):1187–1198.

Rebora, P., Salim, A., and Reilly, M. (2014). bshazard: A flexible tool for nonparametric smoothing of the hazard function. *The R Journal*, 6(2):114–122.

Renganathan, V. (2016). Overview of frequentist and bayesian approach to survival analysis. *Applied Medical Informatics*, 38(1):25.

Ristic, M. M., Jose, K., and Ancy, J. (2007). A marshall–olkin gamma distribution and minification process. *Stress Anxiety Res Soc*, 11:107–117.

Robinson, M. (2014). *Mixture cure models: Simulation comparisons of methods in R and SAS*. PhD thesis, University of South Carolina.

Rocha, R., Nadarajah, S., Tomazella, V., and Louzada, F. (2016a). Two new defective distributions based on the marshall–olkin extension. *Lifetime data analysis*, 22(2):216–240.

Rocha, R., Nadarajah, S., Tomazella, V., Louzada, F., and Eudes, A. (2017). New defective models based on the kumaraswamy family of distributions with application to cancer data sets. *Statistical methods in medical research*, 26(4):1737–1755.

Rocha, R. F. d. et al. (2016b). Defective models for cure rate modeling.

Rodrigues, J., Cancho, V. G., de Castro, M., and Balakrishnan, N. (2012). A bayesian destructive weighted poisson cure rate model and an application to a cutaneous melanoma data. *Statistical methods in medical research*, 21(6):585–597.

Rodrigues, J., Cancho, V. G., de Castro, M., and Louzada-Neto, F. (2009a). On the unification of long-term survival models. *Statistics & Probability Letters*, 79(6):753–759.

Rodrigues, J., Cordeiro, G. M., Cancho, V. G., and Balakrishnan, N. (2016). Relaxed poisson cure rate models. *Biometrical Journal*, 58(2):397–415.

Rodrigues, J., de Castro, M., Balakrishnan, N., and Cancho, V. G. (2011). Destructive weighted poisson cure rate models. *Lifetime data analysis*, 17(3):333–346.

Rodrigues, J., de Castro, M., Cancho, V. G., and Balakrishnan, N. (2009b). Com–poisson cure rate survival models and an application to a cutaneous melanoma data. *Journal of Statistical Planning and Inference*, 139(10):3605–3611.

Rodrıguez, G. (2005). Non-parametric estimation in survival models.

Rodrıguez, G. (2010). Parametric survival models. Technical report, Technical report. Princeton: Princeton University.

Royston, P. (2001). The lognormal distribution as a model for survival time in cancer, with an emphasis on prognostic factors. *Statistica Neerlandica*, 55(1):89–104.

Saleem, M. and Aslam, M. (2009). On bayesian analysis of the rayleigh survival time assuming the random censor time. *Pakistan Journal of Statistics*, 25(2).

Samartzis, L. (2006). Survival and censored data. Technical report.

Sandhya, E. and Prasanth, C. (2014). Marshall-olkin discrete uniform distribution. *Journal of probability*, 2014.

Santos-Neto, M., Bourguignon, M., Zea, L. M., Nascimento, A. D., and Cordeiro, G. M. (2014). The marshall-olkin extended weibull family of distributions. *Journal of Statistical Distributions and Applications*, 1(1):9.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. CRC press.

Schraudolph, N. N., Yu, J., and Günter, S. (2007). A stochastic quasi-newton method for online convex optimization. In *Artificial Intelligence and Statistics*, pages 436–443.

Sellers, K. F., Borle, S., and Shmueli, G. (2012). The com-poisson model for count data: a survey of methods and applications. *Applied Stochastic Models in Business and Industry*, 28(2):104–116.

Sellers, K. F., Morris, D. S., and Balakrishnan, N. (2016). Bivariate conway–maxwell–poisson distribution: formulation, properties, and inference. *Journal of Multivariate Analysis*, 150:152–168.

Shaby, B. and Wells, M. T. (2010). Exploring an adaptive metropolis algorithm. *Currently under review*, 1:1–17.

Shehla, R. and Khan, A. A. (2013). Bayesian analysis of gamma model with laplace approximation. *International Journal of Innovative Research in Science, Engineering and Technology*, 2(11).

Shen, P.-S. (2017). A semiparametric regression cure model for interval-censored and left-truncated data. *Communications in Statistics-Simulation and Computation*, 46(5):3690–3700.

Siannis, F., Copas, J., and Lu, G. (2005). Sensitivity analysis for informative censoring in parametric survival models. *Biostatistics*, 6(1):77–91.

Sinha, D., Chen, M.-H., and Ibrahim, J. G. (2003). Bayesian inference for survival data with a surviving fraction. *Lecture Notes-Monograph Series*, pages 117–138.

Soland, R. M. (1969). Bayesian analysis of the weibull process with unknown scale and shape parameters. *IEEE Transactions on Reliability*, 18(4):181–184.

Son, Y. S. and Oh, M. (2006). Bayesian estimation of the two-parameter gamma distribution. *Communications in Statistics•Simulation and Computation®*, 35(2):285–293.

Spiegelhalter, D. J., Abrams, K. R., and Myles, J. P. (2004). *Bayesian approaches to clinical trials and health-care evaluation*, volume 13. John Wiley & Sons.

Steelman, M. (2015). Survival analysis for truncated data and competing risks.

Stevenson, M. and EpiCentre, I. (2009). An introduction to survival analysis. *EpiCentre, IVABS, Massey University*.

Sumathi, K. and Rao, A. (2008). Cure rate models–a partial review with an application to recurrent event or count data. *in INTERSTAT*.

Swain, P. K., Grover, G., and Goel, K. (2016). Mixture and non-mixture cure fraction models based on generalized gompertz distribution under bayesian approach. *Tatra Mountains Mathematical Publications*, 66(1):121–135.

Sy, J. P. and Taylor, J. M. (2000). Estimation in a cox proportional hazards cure model. *Biometrics*, 56(1):227–236.

Tai, P., Tonita, J., Yu, E., and Skarsgard, D. (2003). Twenty-year follow-up study of long-term survival of limited-stage small-cell lung cancer and overview of prognostic and treatment factors. *International Journal of Radiation Oncology• Biology• Physics*, 56(3):626–633.

Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American statistical Association*, 82(398):528–540.

Taylor, J. M. (1995). Semi-parametric estimation in failure time mixture models. *Biometrics*, pages 899–907.

Thamrin, S. A., McGree, J. M., and Mengersen, K. L. (2013). Bayesian weibull survival model for gene expression data. *Case studies in Bayesian statistical modelling and analysis*, pages 171–185.

Tsodikov, A. (1998). A proportional hazards model taking account of long-term survivors. *Biometrics*, pages 1508–1516.

Tsodikov, A., Ibrahim, J., and Yakovlev, A. (2003). Estimating cure rates from survival data: an alternative to two-component mixture models. *Journal of the American Statistical Association*, 98(464):1063–1078.

Tutkun, N. A. and Demirhan, H. (2016). A bayesian approach to cox-gompertz model. *Hacettepe Journal of Mathematics and Statistics*, 45(5):1621–1639.

Uddin, M., Islam, M., and Ibrahim, Q. (2006). An analytical approach on cure rate estimation based on uncensored data. *Journal of Applied Sciences*, 6(3):548–552.

Uddin, M. T. (2004). *Cure rate estimation based on uncensored and censored data*. PhD thesis, Concordia University.

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

Wan, F. (2017). Simulating survival data with predefined censoring rates for proportional hazards models. *Statistics in medicine*, 36(5):838–854.

Wang, X. (2017). *Inferences on Gamma Distributions: Uncensored and Censored Cases*. University of Louisiana at Lafayette.

Welch, B. and Peers, H. (1963). On formulae for confidence points based on integrals of weighted likelihoods. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 318–329.

Wiangnak, P. et al. (2016). *Expectation Maximization-based Likelihood Inference for COM-Poisson Cure Rate Model with Interval Censored Data*. PhD thesis.

Wienke, A. (2010). *Frailty models in survival analysis*. CRC Press.

Wileyto, E. P., Li, Y., Chen, J., and Heitjan, D. F. (2012). Assessing the fit of parametric cure models. *Biostatistics*, 14(2):340–350.

Witten, M. and Satzer, W. (1992). Gompertz survival model parameters: Estimation and sensitivity. *Applied mathematics letters*, 5(1):7–12.

Wu, Y. (2010). *Extension of cure rate model when cured is partially known*. Rutgers The State University of New Jersey-New Brunswick and University of Medicine and Dentistry of New Jersey.

Xu, J. (2012). *Statistical Modeling and Analysis for Survival Data with a Cure Fraction*. PhD thesis.

Yin, G. and Ibrahim, J. G. (2005). Cure rate models: a unified approach. *Canadian Journal of Statistics*, 33(4):559–570.

Yiqi, B., Cancho, V. G., and Louzada, F. (2017). On the bayesian estimation and influence diagnostics for the weibull-negative-binomial regression model with cure rate under latent failure causes. *Communications in Statistics-Theory and Methods*, 46(3):1462–1489.

Yiqi, B., Maria Russo, C., Cancho, V. G., and Louzada, F. (2016). Influence diagnostics for the weibull-negative-binomial regression model with cure rate under latent failure causes. *Journal of Applied Statistics*, 43(6):1027–1060.

Yu, H. (2017). *Use of Generalized Gamma Distribution in Modeling Lifetime Data*. PhD thesis, University of Texas at Arlington.

Yusuf, M. U. and Bakar, M. R. A. (2016). Cure models based on weibull distribution with and without covariates using right censored data. *Indian Journal of Science and Technology*, 9(28).

Zeigler, B. P., Praehofer, H., and Kim, T. G. (2000). *Theory of modeling and simulation: integrating discrete event and continuous complex dynamic systems*. Academic press.

Zeng, D., Yin, G., and Ibrahim, J. G. (2006). Semiparametric transformation models for survival data with a cure fraction. *Journal of the American Statistical Association*, 101(474):670–684.

Zhang, J. and Peng, Y. (2009). Accelerated hazards mixture cure model. *Lifetime data analysis*, 15(4):455.

Zhang, Z. and Sun, J. (2010). Interval censoring. *Statistical Methods in Medical Research*, 19(1):53–70.

Zhu, C., Byrd, R. H., Lu, P., and Nocedal, J. (1997). Algorithm 778: L-bfgs-b: Fortran subroutines for large-scale bound-constrained optimization. *ACM Transactions on Mathematical Software (TOMS)*, 23(4):550–560.