PARAMETER ESTIMATION FOR STAGE-DURATION MODELS

A dissertation submitted for the degree of Doctor of Philosophy (Statistics)

to the School of Computer Science, Engineering and Mathematics,

Division of Science and Engineering Faculty, Flinders University.

by

Hoa Thi Thu Pham

December 2016

Contents

List of Figures

List of Tables

Summary

Multi-stage time evolving models, so called stage-duration models, have been studied in various biological contexts. We consider stage-duration models that describe single cohort stage-frequency data with destructive samples. These models can give an understanding of the maturation of biological systems, industrial processes or the progression of disease. The main goal of this thesis is to estimate the stage-dependent maturation parameters and hazard rate parameters of the models. The contributions of the thesis are as follows:

First, we obtain novel methods for estimating maturation parameters in models with stage-wise constant hazard rates and with linear time-dependent hazard rates. We use Laplace transform methods with the assumption of constant scale parameters or constant shape parameters. The key result is the exploration of the relationships between the stage-dependent maturation parameters in each stage.

Second, we obtain methods for estimating maturation parameters and hazard rate parameters without imposing unrealistic conditions as in previous studies. In particular, by using a Bayesian approach, we derive estimators of the maturation parameters and the hazard rate parameters in each stage simultaneously, without initial knowledge about maturation parameters. The Metropolis-Hastings (MH) algorithm based on deterministic transformations is applied in order to accelerate the convergence of the Markov process. We embed the relationships of the stage-dependent maturation parameters within the deterministic MH algorithms. The number of sampling times for the deterministic MH methods is reduced compared to the Laplace transform methods.

Third, the application of the methodology in the models is evaluated using both simulated data and case studies including cattle parasitic data and breast development data of New Zealander schoolgirls. From the simulated data, results show that the proposed methods are able to estimate parameters in situations where non-trivial hazard rates apply. The methods also work well when the assumptions of maturation parameters are relaxed.

From the case studies, the results show that parameter estimation is better using these methods in comparison to Laplace transform methods in previous studies.

Declaration

I declare that:

This thesis is my own work and does not incorporate any material that has been submitted previously, in whole or in part, for the award of any other academic degree or diploma except where referenced or acknowledged.

To the best of my knowledge, this thesis does not contain, without acknowledgement, any material previously published or written by another person.

Hoa Thi Thu Pham

December 2016

Publications

This thesis was written under the supervision of Associate Professor Alan Branford, Associate Professor Murk Bottema, Dr Darfiana Nur, and Professor Jerzy Filar. Some parts of the thesis have been published as follows:

1. A shortened version of Chapter 3 was published in Australian & New Zealand Journal of Statistics:

H. Pham and A. Branford. Exploring parameter relations for multi-stage models in stage-wise constant and time dependent hazard rates. Australian & New Zealand Journal of Statistics, 58(3):357–376, 2016.

2. A shortened version of Chapter 4 is presented in the New Zealand Statistical Association conference on 27^{th} November 2016:

H. Pham, D. Nur, H.T.T. Pham and A. Branford. A Bayesian approach for parameter estimation in multi-stage models.

Acknowledgements

I remember clearly all the difficulties from when I commenced my PhD at Flinders University. The difficulties were both from the research and the new life of an international student. The knowledge and experience I have gained from these difficulties has been due to the love and support I received from the people around me. The kindness, guidance, honesty, friendship and caring of the people around me made me feel lucky to be a PhD student at Flinders University and to be their friend.

First of all, I would like to express my very great appreciation to my supervisors Associate Professor Alan J. Branford and Professor Jerzy A. Filar for their guidance, ideas and wise advice and Professor Murk Bottema and Dr. Darfiana Nur for their help when I truly needed it. I would like to thank Dr. Joshua V. Ross for his patience and new ideas for my models and Professor Terrence P. Speed and Professor Richard Boys for their valuable comments. I would also like to thank the school administrative staff for supporting a friendly working environment.

I would like to offer my special thanks to the staff of the ISSU for their support. The way that they have supported international students including my friends and me has been very inspiring. It is clearly beyond their duty; it is their love and commitment to what they do. In addition, my special thanks are extended to the Vietnamese student association at Flinders University for activities connecting students, helping new students, organizing food events and Vietnamese student festivals. These activities have helped me balance research and everyday life.

This project would have been impossible without the generous financial support from the Department of Foreign Affairs and Trade under the Australia Awards Scholarships. My postgraduate study years were quite comfortable and enjoyable.

Finally, I would like to thank my parents and my friends for their support and encouragement throughout my study.

Chapter 1

Introduction

Multi-stage time-evolving models are fundamental for many biological systems in which time-evolving progression moves through distinct stages ([47]; [51]; [39]; [12]; [31]). Generally, there are three classes of multi-stage models: matrix models, stage-duration models and delay-differential equation models ([19]). This thesis focuses on stage-duration models for single cohort stage-frequency data, which are usually applied to unmarked cohort data. In these models, individuals are assumed to enter the study population at the same time. In addition, individuals are not identified due to destructive sampling at different times. The data are collected by assessing the stages reached by individuals as a stage-structured time series. In particular, the numbers of alive and dead individuals in each stage are counted at each sampling time. These stage-duration data are fundamental to biology and have been studied in various biological contexts ([47]; [58]; [26]; [51]). Exploring the development of the population in each stage gives an understanding of maturation of individuals through their life cycle or industrial processes or the progression of a disease. Thus these models are also important for developing strategies to treat diseases.

In stage-duration models, life cycles of individuals are divided into stages. The distributions of stage duration and survival time in each stage are commonly modelled using exponential, Weibull, Erlangian and gamma distributions. The main challenge of these models is to develop methods for estimating stage-dependent maturation parameters and hazard rates in each stage. Previous studies lacked flexibility in the assumptions needed to estimate the stage-dependent maturation parameters and the hazard rate parameters in each stage. When the stage durations were modelled using gamma distributions, the shape or the rate parameters of the distributions were assumed to be known and the same. This assumption is frequently unrealistic, but it simplifies estimations. They also assumed that individuals could not die but rather would eventually move to the next

stage, or that individuals could die but that the hazard rate was the same for each stage ([51]; [25]; [19]; [33]).

In this thesis, the assumptions on the hazard rate parameters are relaxed and no assumptions are made on the maturation parameters. The objectives are to propose novel methods for estimating parameters in the following situations:

- To estimate parameters for models with stage-wise constant hazard rates by constant shape rate and constant scale cases. This means that shape parameters or rate parameters are known and constant ([51]).
- To estimate parameters for models with linear time-dependent hazard rates by constant shape rate and constant scale cases.
- To estimate parameters without initial knowledge of shape and rate parameters for models with no hazard rates.
- To estimate parameters without initial knowledge of shape and rate parameters for models with stage-wise constant hazard rates.
- To estimate parameters without initial knowledge of shape and rate parameters for models with linear time-dependent hazard rates.

This thesis is organized into six chapters including the present introductory chapter. Chapter 2 introduces stage-duration models, the Laplace transform method, Bayesian analysis and previous studies of stage-duration models. In Chapter 3, we propose two new stage-duration models where non-trivial hazard rates apply. The first model considers hazard rates that are constant within each stage but vary between stages. The second model considers linear time-dependent hazard rates within stages. We use the Laplace transform method to estimate stage-dependent maturation parameters for these models. In addition, we propose methods for estimating these non-trivial hazard rates. A key result in this chapter is to explore the relationships among the stage-dependent maturation parameters in each stage. Those relationships are then embedded within the Markov chain Monte Carlo (MCMC) algorithm proposed in the subsequent chapter for better convergence of Markov processes. Each model is applied to simulated datasets in order to evaluate the accuracy. Chapter 4 uses a Bayesian approach to estimate parameters in the models with no hazard rates, with stage-wise constant hazard rates and with linear time-dependent hazard rates. Metropolis-Hastings (MH) algorithm based on deterministic

transformations is used to improve the mixing of Markov chains. The main aim of this chapter is to relax assumptions about maturation parameters in the proposed models. By allowing uncertainties through prior distributions, parameters are estimated using a Bayesian approach. These methods are implemented in simulated datasets. In Chapters 5 and 6, we apply the techniques from Chapter 4 to simulation studies and the case studies of cattle parasite data and data related to breast development of New Zealander schoolgirls. Finally, Chapter 7 draws conclusions from the major results in the thesis and discusses remaining problems and future research directions.

Chapter 2

Literature Review

2.1 Stage-duration data and models

2.1.1 Stage-duration data

Stage-duration data, sometimes called stage-frequency data ([39]; [37]), is the information relating to the life cycle of individuals having distinct life stages and is obtained by counting the number of individuals in different stages at different sampling times ([39]; [31]; [12]). In biological contexts, the starting point is controlled and destructive sampling is required to assess the stage reached by individuals ([47]; [51]; [25]; [19]; [33]). Stage-duration models are often used for the analysis of populations of insects such as Callosobruchus chinensis ([6]), cattle parasitic nematodes ([58]), copepod ([32]) and grasshopper Chorthippus ([46]). Moreover, some individuals' progress can be categorised by size as in the case of the study on breast development of New Zealander schoolgirls ([39], p. 98).

Such data have the following three specific properties:

• Initially, we consider single cohort stage-frequency data. This means that the starting point is controlled. We suppose that at the first sampling time, all individuals commence at stage 1;

• An individual's stage cannot be observed without harvesting; thus destructive samples are conducted. Different individuals are counted at each sampling time;

• Finally, we consider the effect of mortality on stage duration. The mortality rate varies from stage to stage. At each sampling time, the number of dead organisms in each stage can be counted. An individual spending a long time in a stage with high hazard rate is less likely to survive than an individual only spending a short time in the stage.

Stage-duration data are often used to model insect populations. For example, an insect may pass through several stages: from the egg stage to many interim stages before finally reaching the adult stage. The only information that can be collected is the number of organisms in each stage. This is gathered from samples taken at different sampling times.

Suppose that the life cycle of an individual in a population consists of $(I + 1)$ stages, where stage $(I + 1)$ is some final stage (for example, the adult stage or death stage). We consider here the case where an individual must pass from one stage to the next without missing a stage. However, death can occur in any stage, from stage 1 to stage I, before progressing to the next stage.

$$\begin{array}{ccccccccc} 1 & \rightarrow & 2 & \rightarrow & ... & \rightarrow & I \\ \searrow & & \downarrow & \downarrow & & & \swarrow \\ & & & Death & & & \end{array} \tag{2.1.1}$$

Possible transitions of individuals at each stage.

An example of stage-duration data is an experiment for cattle parasitic nematode ([58]). The data for this study (Table 2.1) comprises four stages of the parasite life cycle; including stage 1 (eggs), stage 2 (first stage larvae), stage 3 (second stage larvae) and stage 4 (third stage larvae). Samples were collected at roughly the same stage of development. Numbers of cattle parasites in the four different stages of the life cycle were counted at 10 sampling times. The samples were destroyed after determining the number of parasites in each of the four stages. The data are proportions of individuals at each stage out of approximately 606 eggs at each sample time.

2.1.2 Models

The stage time and death time (if death occurs) of each individual are unknown. Rather, there are sampling times at which a sample of organisms are drawn from the population and the numbers of alive and dead organisms in each stage are observed. In each stage j, the lifetime of an individual is influenced by the death density function $\mu_j(t)$ and the stage time in stage j which has the density function $g_j(t)$ depending on the parameter

Table 2.1: Data for cattle parasitic nematode example includes the proportions of observations at each stage of Ostertagia ostertagi, expressed as a percentage. Stage 1 = eggs (unembryonated embryonated), stage 2 = 1st stage larvae, stage 3 = 2nd stage larvae, stage 4 = 3rd stage larve. Time is in hours.

Sampling number	**Time**	**Stage 1**	**Stage 2**	**Stage 3**	**Stage 4**
1	20	87	0	3	0
2	45	12	81	0	0
3	65	11	13	74	2
4	90	6	7	65	14
5	115	14	2	54	21
6	140	10	2	24	61
7	160	0	0	17	67
8	185	0	0	33	51
9	210	0	0	10	70
10	260	0	0	8	70

θ_j. Define $N_j(T_k) = N_{kj},\ j = 1, ..., I,\ k = 1, ..., K$ to be the number of organisms alive in stage j at time T_k and $D_j(T_k) = D_{kj}$ to be the number of organisms that are observed to be dead in stage j at time T_k. Let N_k be the number sampled at time T_k, which in the case of zero hazard rate is given by

$$N_k = \sum_{j=1}^{I+1} N_j(T_k)\,.$$

In the case of stage-wise constant hazard rates, N_k is given by

$$N_k = \sum_{j=1}^{I+1} N_j(T_k) + \sum_{j=1}^{I+1} D_j(T_k)\,.$$

The probability of the life time of each particular independent organism is defined in [47] as

$$\begin{aligned} p_j(t) &= \mathrm{P}(\text{an organism is alive in stage } j \text{ at time t} \mid \text{the organism starts stage 1 at time 0}) \\ &= h_1 * h_2 * ... * h_{j-1} * H_j(t)\,, \quad j = 1, ..., I, \end{aligned} \tag{2.1.2}$$

where

$$h_i(t) = g_i(t)S_i(t) = g_i(t)\exp\left(-\int_0^t \mu_i(x)dx\right), i = 1, ..., j-1 \tag{2.1.3}$$

is the density function of an organism being alive in stage i and the notation $*$ denotes convolution. Furthermore, $S_i(t)$ is the probability of an organism being alive in stage i for longer than time t and $\mu_i(t)$ is the hazard function of the survival time in stage i at time t, and

$$H_j(t) = S_j(t) \int_t^\infty g_j(x)dx = \exp\left(-\int_0^t \mu_j(x)dx\right)\left[\int_t^\infty g_j(x)dx\right] .$$

The maximum-likelihood (ML) method is difficult to implement in this context. Compared to ML methods, methods based on Laplace transform matching are very efficient in multi-stage models because of the ease of computation ([51]). This is illustrated in Section 3.2.1, where the stage time densities $g_j(t)$ take the forms of the exponential, Erlangian and gamma distributions. Because we cannot observe the time of individuals transiting from stage to stage, or their time of death, stage-duration data are totally censored. The convolution in (2.1.2) shows that the maximum likelihood method is difficult to apply in the stage-duration models, especially when the number of stages increase. In the next sections, we present approaches to estimate parameters for these models. These approaches include a frequentist approach (Section 2.2) and a Bayesian approach (Section 2.3).

2.2 Frequentist inference

In this section, we introduce Laplace transform methods applied in the stage-duration models in order to estimate parameters. These methods were first proposed in [51]. The method of moments based on Laplace transform was exploited in order to derive robust estimates and calculate the variance of these estimates ([25]).

In general, let the probability of an event occurring at time t be $p_\theta(t)$ and let $n(t)$ represent the population proportion at time t. Let T be a random sampling time with exponential density function $\alpha(t) = s\exp(-st)$. The principle of Laplace transform matching may be illustrated as follows.

By considering random sampling times, instead of estimating $\theta = (\theta_1, \theta_2, \ldots, \theta_I)$ by matching $p_\theta(t)$ to $n(t)$, $\theta = (\theta_1, \theta_2, \ldots, \theta_I)$ is estimated by matching $\psi_\theta(s) = s\int_0^\infty p_\theta(t)\exp(-st)\,dt$

to the population proportion as

$$\hat{\psi}_\theta(s) = \sum_{j=1}^{K} I_j n(T_j) \,, \tag{2.2.1}$$

where I_j is defined in Equation (2.2.2) below.

In particular, stratified sampling is used to increase the efficiency of estimates, especially when fixed sampling times dominate most experiments ([51]). Let $\mathcal{C} = \{c_0 = 0, c_1, ..., c_K = \infty\}$ be a set of cutpoints between K sampling times $T_1, T_2, ..., T_K$. The variables $T_1, T_2, ..., T_K$ are independent and identically distributed as

$$\alpha_j(t) = \begin{cases} \frac{\alpha(t)}{I_j} & t \in [c_j, c_{j+1}] \\ 0 & t \notin [c_j, c_{j+1}] \end{cases} , \tag{2.2.2}$$

where $I_j := \int_{c_{j-1}}^{c_j} \alpha(t)dt = \exp(-sc_{j-1}) - \exp(-sc_j)$.

The iterative method of Schuh and Tweedie ([51]):

The choice of the sampling rate s for the Laplace transform is not obvious. In some models, an iterative method ([51]) is used to determine s. The mean sampling time s^{-1} is chosen to equal the mean of each stage time density function $g_j(t)$, $j = 1, ..., I$. Details of the iterative steps are as follows. In stage j, an arbitrary value s_1^{-1} is chosen and parameters θ_j of stage j are estimated through Laplace transform matching with sampling time s_1^{-1}. Then $g_j(t)$ is updated according to the new parameters θ_j. Taking s_2^{-1} = mean of $g_j(t)$, the parameters are estimated through Laplace transform matching with sampling time s_2^{-1}. This iteration is repeated until the value s^{-1} converges to the estimated mean of $g_j(t)$.

Then the sampling time s multiplying the Laplace transform of $p_\theta(t)$ and empirical data can be matched as

$$\psi_j(s) = s \int_0^\infty p_j(t) \exp(-st)\, dt \approx \sum_{k=1}^{K} I_k \frac{N_j(T_k)}{N_k} \,, \quad j = 1, ..., I \,. \tag{2.2.3}$$

2.3 Bayesian inference

This section reviews basic theory underlying Markov chain Monte Carlo (MCMC) methods based on [49], [48], [9], [24] and [16].

2.3.1 Bayesian perspective

The conditional probability of observing event B given that event A has ocurred or must occur has the mathematical form

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)},$$

where $P(A) > 0$, (e.g., see ([48])).

The concept underlying the so-called Bayesian approach to parameter estimation is to first treat the parameter θ as a random variable with a probability density function $p(\theta)$, known as the prior distribution. Then, after carrying out an experiment where a data vector y is observed, we revise the latter distribution in view of the obseved data and call it the posterior distribution denoted by $\pi(\theta|y)$.

More precisely, we consider a general case where the data vector $y = (y_1, y_2, ..., y_n)$ is a realization of n sampling points of independent, identically distributed, random variables $Y_1, Y_2, ..., Y_n$. Thus $y \in \mathcal{Y} \subset R^n$ and $y_i \in \mathcal{Y}_i$, $i = 1, ..., n$. Let $\theta = (\theta_1, \theta_2, ..., \theta_d)$ be a vector of parameters of the model having prior density $p(\theta)$, where $\theta \in \Theta \subset R^d$. Recall that parameters θ_i, $i = 1, ..., d$ are unknown and treated as random variables. Then Bayes's theorem extended to conditional density functions implies

$$\pi(\theta|y) = \frac{f(y|\theta)p(\theta)}{f(y)} = \frac{f(y|\theta)p(\theta)}{\int_\Theta f(y|\theta)p(\theta)d\theta}, \tag{2.3.1}$$

where $\pi(\theta|y)$ is the posterior probability density function of θ given the observations y, $f(y|\theta)$ is the likelihood function and $f(y)$ is the marginal probability density function of the data. The methods of choosing a prior distribution and deriving the posterior distribution are presented in the following section.

2.3.2 Links between posterior and prior distributions of the parameters

All aspects of the Bayesian inference are based on the posterior density (2.3.1). The numerator of (2.3.1) is computable while the denominator, which is a marginal likelihood, is usually difficult to compute. However, the marginal likelihood is merely a normalizing

constant and does not include any information about the parameter vector θ. Hence, ignoring the marginal likelihood, we note that posterior density is proportional to the likelihood multiplied by the prior density ([9]; [24]), as follows:

$$\begin{aligned} \pi(\theta|y) &= \frac{f(y|\theta)p(\theta)}{f(y)} = \frac{f(y|\theta)p(\theta)}{\int_\Theta f(y|\theta)p(\theta)d\theta} \\ &\propto f(y|\theta)p(\theta) \\ &\propto \text{Likelihood} \times \text{Prior}\,. \end{aligned} \tag{2.3.2}$$

Our interest is in the simulated values of parameter θ from the posterior distribution. If the value of the marginal likelihood is computable, inferences via the posterior distribution can be based on the posterior density or the mass function directly. When the value of the marginal likelihood is not analytically available, the inferences based on the posterior can be approximated via simulation from the posterior as described in Section 2.3.3.

An important aspect of the Bayesian approach is the need to specify prior distributions for the unknown parameters that would express one's beliefs about the parameters of interest before some evidence is taken into account. For complex models, the multiplication of the joint likelihood and the joint prior distribution in (2.3.2) usually results in a non-standard joint posterior distribution of the parameters of interest such that an MCMC algorithm is needed for parameter estimation.

The common types of priors are non-informative priors, partially informative priors and informative priors ([9]; [24]). Non-informative priors are chosen when we have very little information about the parameters of interest. The uniform distribution is usually chosen as a non-informative prior distribution. Partially informative priors are used when the priors' distribution is built on using history matching. We keep the priors having reasonable summary statistics from generated data. Informative priors are used when we have specific information about the parameters of interest which incorporated in the prior density.

2.3.3 Markov chain Monte Carlo methods

When the posterior density $\pi(\theta|y)$ is complicated and θ is high dimensional, the traditional Monte Carlo method cannot be implemented. Markov chain Monte Carlo (MCMC) algorithms are commonly used to solve statistical computation problems related to Bayesian

inference. Of greatest interest are the posterior means of a function $g : \Theta \to R$ ([16]), defined by

$$I(y) := \int_{\Theta} g(\theta)\pi(\theta|y)d\theta = \frac{\int_{\Theta} g(\theta)f(y|\theta)p(\theta)d\theta}{\int_{\Theta} f(y|\theta)p(\theta)d\theta} \,. \tag{2.3.3}$$

MCMC methods are constructed by specifying a set of transition probabilities for an associated Markov chain. MCMC methods generate a proposal $\theta^{(t+1)}$ once we know a current state of the chain $\theta^{(t)}$. The basic idea of MCMC methods for a given probability distribution π such that $\pi : \Theta \to [0, 1]$ is generating random elements of Θ with distribution π. The MCMC methods do this by constructing a Markov chain with a stationary distribution π and simulating that chain.

The function π is called a stationary distribution of an irreducible Markov chain having transition matrix $P = [p(i, j)]_{i,j=1}^{m}$, where $m$is the number of states and $p(i, j) := p\left(\theta^{(t+1)} = j|\theta^{(t)} = i\right)$. Recall that $\sum_{j\in\Theta} \pi(j) = 1$ and

$$(\pi P)(j) = \pi(j), \text{ for every } j \,. \tag{2.3.4}$$

More specifically, let $\mathcal{S}$ be a finite state space and π be any probability distribution on Θ such that $\pi(j) > 0$ (the target distribution or the posterior distribution). We can define a new Markov chain $\{\theta^{(t)}\}$ such that its stationary distribution is π. The MCMC methods use the realizations $\left(\theta^{(1)}, ..., \theta^{(M)}\right)$ obtained from the Markov chain as the Monte Carlo sample. Note that $\theta^{(t)}$ are constructed with the help of the observed data vector y, however, to simplify the already complex notation, the argument y is supressed. Using (2.3.3), the estimator of the mean of function g is

$$\widehat{I}_M(y) = \frac{1}{M - B} \sum_{t=B+1}^{M} g(\theta^{(t)}) \,, \tag{2.3.5}$$

where B is a fixed nonnegative integer that indicates the number of initial sample values that will be discarded. The number of initial values B is chosen at the point when a Marlov chain settles into equilibrium distribution in order to avoid biases toward arbitrary initial values. It can be shown that $\widehat{I}_M(y) \to I(y)$ with probability 1 as $M \to \infty$ (Brooks et al. [9]).

2.3.4 Overview of the Metropolis-Hastings algorithm

The Metropolis-Hastings (MH) algorithm ([49]; [24]) proposes a new state of the Markov chain that is either accepted with probability α, or rejected with probability $1-\alpha$. If the state is accepted, the Markov chain moves to the new state. If the proposed state is rejected, the Markov chain remains in the same state. By choosing the acceptance probability α correctly, we create a Markov chain which has π as its stationary distribution. Given such a state space $\mathcal{S}$, a stationary distribution π and a proposal transition matrix $Q = [q(i|j)]_{ij}$, the MH algorithm is constructed as follows

The MH algorithm

1: Initialise $\theta^{(0)}$
2: for $t = 1$ to T, do
3: Given the current state $\theta = \theta^{(t)}$, propose $\theta^{(*)} \sim q\left(\theta^{(*)}|\theta^{(t)}\right)$
4: Calculate the acceptance probability $\alpha = \min\left(1, \frac{\pi\left(\theta^{(*)}\right)q\left(\theta^{(t)}|\theta^{(*)}\right)}{\pi\left(\theta^{(t)}\right)q\left(\theta^{(*)}|\theta^{(t)}\right)}\right)$
5: Set $\theta^{(t+1)} = \theta^{(*)}$ with probability α, otherwise set $\theta^{(t+1)} = \theta^{(t)}$
6: end for

2.3.5 Choice for the proposal distribution

For the MH algorithm, there are two popular choices for the proposal probability distribution $q(\theta^{(*)}|\theta^{(t)})$: (i) a random walk and (ii) an independent proposal ([24]).

(i) A random walk is used in situations for which we have little idea about the shape of the target distribution. Therefore, we need to ensure the entire parameter space is explored.

In the MH algorithm, the chain of random variables satisfies

$$\theta^{(t+1)} = \theta^{(t)} + \varepsilon_t \,, \tag{2.3.6}$$

where ε_t is generated independently of θ and y. The random walk generated is symmetric or non-symmetric if the distribution of ε is symmetric or non-symmetric, respectively. When the transition matrix P is symmetric, then we have the acceptance probability

$$\alpha = \min\left(1, \frac{\pi(\theta^{(*)})}{\pi(\theta_t)}\right).$$

For example, $\varepsilon_t \sim N(0, \boldsymbol{V}) \Rightarrow q(\theta^{(*)}|\theta^{(t)}) \sim N(\theta^{(t)}, \boldsymbol{V})$ is a symmetric random walk. We usually choose the covariance matrix as $\boldsymbol{V} = \sigma^2 I_d$, where d is the dimension of $\theta \in \Theta$. In this case, the covariance matrix $\boldsymbol{V}$ plays a crucial role in the performance of the sampling algorithm. The value of the covariance matrix $\boldsymbol{V}$ should be chosen in order to lead to the best performance of the MH algorithm.

If the elements on the diagonal of the covariance matrix $\boldsymbol{V}$ are very small, then the proposal parameter $\theta^{(*)}$ is close to the current state $\theta^{(t)}$. Thus, the proposals will usually be accepted, but the chain will hardly move, which is clearly suboptimal.

If the elements on the diagonal of the covariance matrix $\boldsymbol{V}$ are very large, then by (2.3.6) the proposal parameter $\theta^{(*)}$ will usually be very far from the current state $\theta^{(t)}$. Thus, the values $\pi(\theta^{(*)})$ are likely to be very small. This implies that they will almost always be rejected, which is again clearly suboptimal.

Therefore, the optimal value of the covariance matrix $\boldsymbol{V}$ should be chosen to be proportional to the true covariance matrix of the target distribution π (we usually do not know this covariance). It has been shown in [9] that the optimal covariance matrix has the form

$$\boldsymbol{V}_d = \frac{(2.38)^2}{d}\boldsymbol{V}\,.$$

Tuning variances from the normal random walk can be tedious, but the desirable acceptance rate is between 20% and 50% ([9]).

(ii) An independent sampler is used in situations for which we have a pretty good idea about the target distribution π. Choosing the proposed parameter $\theta^{(*)}$ as

$$q\left(\theta^{(*)}|\theta^{(t)}\right) = f\left(\theta^{(*)}\right)\,, \tag{2.3.7}$$

where $\theta^{(*)}$ and the density function f do not depend on the current state of the chain $\theta^{(t)}$, the acceptance rate is denoted as

$$\alpha = \min\left(1, \frac{\pi\left(\theta^{(*)}\right) f\left(\theta^{(t)}\right)}{\pi\left(\theta^{(t)}\right) f\left(\theta^{(*)}\right)}\right)\,. \tag{2.3.8}$$

2.3.6 Reversibility and stationarity of Markov chain for the MH algorithm

The convergence in various senses of the distribution of a Markov chain $\left\{\theta^{(t)}\right\}$ to the target distribution $\pi(.)$ is discussed in [40], [50] and [57].

An irreducible Markov chain is one that has a positive probability of eventually reaching any one location from any other (it is possible to reach any state from any state).

A stationary Markov chain is reversible if the transition matrix $P = \left[p(\theta^{(i)}, \theta^{(j)})\right]_{i,j=1}^{m}$, where m is the number of states, and the stationary distribution π satisfy the balance property

$$\pi(\theta^{(i)})p(\theta^{(i)}, \theta^{(j)}) = \pi(\theta^{(j)})p(\theta^{(j)}, \theta^{(i)}), \text{ for } i, j = 1, ..., m\,. \tag{2.3.9}$$

If, for an irreducible Markov chain with a transition matrix P, there exists a probability solution π satisfying the balance property for all pairs of states (i, j), then the chain is time-reversible and the solution π is the unique stationary distribution.

Using (2.3.8), the transition matrix $P = \left[p(\theta^{(i)}, \theta^{(j)})\right]_{i,j=1}^{m}$ of the Markov chain in the MH algorithm has the form

$$p(\theta^{(i)}, \theta^{(j)}) = q(\theta^{(j)}|\theta^{(i)})\alpha = q(\theta^{(j)}|\theta^{(i)})\min\left(1, \frac{\pi(\theta^{(j)})q(\theta^{(i)}|\theta^{(j)})}{\pi(\theta^{(i)})q(\theta^{(j)}|\theta^{(i)})}\right), \text{ for } i, j = 1, ..., m\,. \tag{2.3.10}$$

Because the transition matrix of the Markov chain $P = \left[p(\theta^{(i)}, \theta^{(j)})\right]_{i,j=1}^{m}$ from the MH algorithm satisfies the balance property (2.3.9) , the new Markov chain $\{\theta^{(t)}\}$ has a stationary distribution $\pi(.)$.

2.3.7 Overview of Metropolis-Hastings algorithm based on deterministic transformations

In order to simplify the Metropolis-Hastings algorithm based on deterministic transformations ([20]; [4]; [53]; [35]), the parameter θ is seperated into two parts $\theta = (\theta_1, \theta_2)$. The algorithm implies that one parameter is a deterministic proposal conditional on the other parameter as

$$\theta_2 = \boldsymbol{f}(\theta_1) \text{ or } \theta_1 = \boldsymbol{f}^{-1}(\theta_2)\,.$$

This relationship is embedded in MH algorithm. The acceptance probability is calculated as

$$\begin{aligned}
\alpha &= \min\left(1, \frac{\pi\left(\theta_1^{(*)}, \theta_2^{(*)}|y\right) q\left(\theta_1^{(t)}|\theta_1^{(*)}\right)}{\pi\left(\theta_1^{(t)}, \theta_2^{(t)}|y\right) q\left(\theta_1^{(*)}|\theta_1^{(t)}\right)}\right) \\
&= \min\left(1, \frac{f\left(y|\theta_1^{(*)}, \theta_2^{(*)}\right) p\left(\theta_1^{(*)}, \theta_2^{(*)}\right) q\left(\theta_1^{(t)}|\theta_1^{(*)}\right)}{f\left(y|\theta_1^{(t)}, \theta_2^{(t)}\right) p\left(\theta_1^{(t)}, \theta_2^{(t)}\right) q\left(\theta_1^{(*)}|\theta_1^{(t)}\right)}\right) \\
&= \min\left(1, \frac{f\left(y|\theta_1^{(*)}, \theta_2^{(*)}\right) p(\theta_1^{(*)})^2 \left|\left(\boldsymbol{f}^{-1}\right)'_{\theta_2^{(*)}|\theta_1^{(*)}}\right| q\left(\theta_1^{(t)}|\theta_1^{(*)}\right)}{f\left(y|\theta_1^{(t)}, \theta_2^{(t)}\right) p(\theta_1^{(t)})^2 \left|\left(\boldsymbol{f}^{-1}\right)'_{\theta_2^{(t)}|\theta_1^{(t)}}\right| q\left(\theta_1^{(*)}|\theta_1^{(t)}\right)}\right)\,.
\end{aligned} \tag{2.3.11}$$

The Metropolis-Hastings algorithm based on deterministic transformations is presented as follows:

The MH algorithm

1: Initialise $\theta^{(0)} = \left(\theta_1^{(0)}, \theta_2^{(0)}\right)$
2: for $t = 1$ to T, do
3: Given the current state $\theta_1 = \theta_1^{(t)}$, propose $\theta_1^{(*)} \sim q\left(\theta_1^{(*)}|\theta_1^{(t)}\right)$
$\theta_2^{(t+1)} = \boldsymbol{f}\left(\theta_1^{(t+1)}\right)$ and $\theta_2^{(*)} = \boldsymbol{f}\left(\theta_1^{(*)}\right)$
4: Calculate the acceptance probability α
5: Set $\left(\theta_1^{(t+1)}, \theta_2^{(t+1)}\right) = \left(\theta_1^{(*)}, \theta_2^{(*)}\right)$ with probability α,
otherwise set $\left(\theta_1^{(t+1)}, \theta_2^{(t+1)}\right) = \left(\theta_1^{(t)}, \theta_2^{(t)}\right)$
6: end for

2.3.8 Gelman and Rubin and Geweke convergence diagnostic tests

Gelman and Rubin (1992) proposed a convergence diagnostic test to monitoring convergence of MCMC output **([22]; [49]; [9]; [24])**. The convergence diagnostic test is based on calculating a potential scale reduction factor, R. If R values are substantially larger than one, it indicates that the chains do not converge. In this case, we should run chains

longer to improve the convergence of the chains. Functions for calculating and plotting the potential scale reduction factor R (gelman.diag and gelman.plot) are provided in the coda package in the language R ([45]).

Geweke (1991) proposed a convergence diagnostic test based on a test of equality of means between the first 10% and last 50% of a Markov chain **([23]; [49])**. The convergence diagnostic test is based on a standard Z-score for the test. Functions for calculating and plotting Z-scores (geweke.diag and geweke.plot) are provided in the coda package in the language R ([45]).

2.4 Research on stage-duration models

There are models for repeated censuses of cohort stage structure data which we do not consider ([7]; [17]; [34]; [52]). In these models, although individuals cannot be defined and their life cycle divided into stages, there is only one sample. At each sampling time, the number of individuals in each stage can be assessed repeatedly. The models that we address in this thesis are distinguished from these other studies by single cohort data collected through destructive samplings ([46]; [47]; [30]; [51]; [21]; [39]; [25]; [27]; [32]; [43]; [19]; [18]; [52]). A different sample is determined at each different sampling time.

There have been many methods for estimating parameters for stage-duration data including maximum likelihood estimators, matching empirical data and theoretical Laplace transforms and Bayesian approaches. These methods provide equations for estimating the parameters of stage duration in each stage. These methods had been studied in many biological contexts as indicated below.

First, the maximum likelihood method was applied to estimate the distribution of periods spent in each stage of the stage-duration models ([47]; [30]). The method was applied to study worm infections in mice ([30]; [26]). The researchers tried to resolve whether the rejection of worms had a physiological or immunological basis. Sixteen groups of mice were infected with worms on day zero. The mice in different groups were subsequently killed at different sampling times to count the number of mice infected. A characteristic of this type of survival data is right censoring which means that the times of infection of some individual mice were not observed. The maximum likelihood method was also applied to

estimation of the life cycle of the grasshopper ([6]). In this study, the assumption was that the growth rate of the grasshopper had an exponential distribution and the distribution of the periods spent in each stage had an Erlangian distribution. They divided the life cycle of the grasshopper into six stages and then estimated the two parameters of Erlangian distributions in each stage. Because the probability of an individual in each stage was complicated, simplified forms for the densities of duration time in each stage were chosen for computational convenience. They used exponentially distributed maturation times. However, usually the exponential maturation times do not fit well the stage-duration data. Instead of using the exponential distributions, gamma and Erlangian distributions are typically used to model the stage-duration data.

The second method is the Laplace transform methods and methods of moments in order to estimate maturation parameters and overall death rate for stage-frequency data ([51]; [21]; [25]). The Laplace transform methods were used under assumptions about shape parameters or rate parameters. Laplace transform techniques were applied to the hatching of eggs of larvae of the parasitic nematode Ostertagia circumcincta ([51]). Moment estimators were used in the above experiment to compare with Laplace transform estimators and maximum likelihood estimators based on the sum of squares ([21]). They also assumed that maturation distributions in each stage were Erlangian distributions. These authors found that moment estimators gave better residual sum of squares compared to Laplace transform estimators. In order to develop the methods in [21], empirical transforms were constructed and two alternative ways of choosing sampling rates were introduced to minimise mean square error ([36]). Another approach is moment methods as introduced by [25]. The methods of moments are based on Laplace transform methods from [51] and used to find the variances of estimates. These methods compared well with the maximum likelihood methods on simulated data, studying multi-stage growth of the parasitic nematode data and the grasshopper life cycle data ([25]).

Recently, many approaches have been taken in order to address the theoretical and statistical aspects of this problem. A major difficulty is the lack of general computational methods to estimate the maturation parameters and stage specific mortality ([41]). Models of repeated censuses of cohort stage structure data without destruction were presented in [19]. Bayesian approaches were used in this paper in order to estimate stage duration parameters and mortality rates. Other approaches were used to estimate separate distri-

butions by combining stages instead of separating stage duration distributions ([32]; [43]). De Valpine and Knape ([18]; [33]) presented computational methods for smoothed maximum likelihood estimation of general multi-stage models from cohort data. The Markov chain Monte Carlo (MCMC) method was mentioned as an extension to solve statistical and computational issues in these papers.

This thesis is an extension of the work in [47], [51] and [25]. In Chapter 3, by considering the stage-duration models with stage-wise hazard rates or linear time-dependent hazard rates, the Laplace transform methods are applied in order to estimate maturation parameters. Because of the non-trivial form of hazard rates, computational methods in the estimation become complex compared with models having no hazard rate ([51]; [25]). Fortunately, the Laplace transform of the convolution form (2.1.2) of the probability of the life time of an individual exposes an elegant recursive structure that facilitates iterative estimation of parameters of interest at a given stage, from their estimates at previous stages (e.g., see (3.2.14)). This insight makes the computational problem much more tractable.

Furthermore, in Chapter 4, we relax assumptions in [25] and estimate parameters simultaneously using a Bayesian approach. The relationships between maturation parameters in each stage, investigated in Chapter 3, are embedded into the MH algorithm based on deterministic transformations. Hence, maturation parameters in each stage are estimated through Bayesian analysis without any initial information about shape or rate parameters. In Chapter 5 we apply the Metropolis-Hastings type algorithms discussed in Chapter 4 to demonstrate their effectiveness with the help of synthetic data. In the process, we compare their relative performance in producing reliable estimates of parameters. In Chapter 6 we extend this numerical testing and validation phase to data from two independent case studies, the first concerning a cattle parasite and the second concerning breast development of New Zealander school girls.

Chapter 3

Parameter Estimation in Multi-stage Models: A Classical Approach

Single cohort stage-frequency data (stage-duration data) are the result of using destructive sampling to identify the stage reached by individuals. For this type of data, when all hazard rates are assumed constant and equal, Laplace transform methods have been applied in the past to estimate the parameters of each stage-duration distribution and the overall hazard rates. If hazard rates are not all equal, estimating stage-duration parameters using Laplace transform methods becomes complex. In this chapter, two new models are proposed to estimate stage-dependent maturation parameters using Laplace transform methods where non-trivial hazard rates apply. The first model considers hazard rates that are constant within each stage but vary between stages. The second model considers linear time-dependent hazard rates within stages. Moreover, this chapter introduces a method for estimating the hazard rate in each stage for the stage-wise constant hazard rates model. This chapter presents methods that could be used in specific types of laboratory studies, but the main motivation is to explore the relationships between stage maturation parameters that, in the next chapter, will be exploited in applying deterministic Bayesian approaches. The application of the methodology in each model is evaluated using simulated data in order to illustrate the models' structure.

3.1 Introduction

Parameter estimation in multi-stage time-evolving models (stage-duration models) has been studied in [47], [5], Manly [39] and [27]. These models are applied in many biological contexts in which a starting point is controlled and destructive samples are required in

order to assess the stages as the system evolves. Such models have been considered in [51] and [25].

In previous studies, in order to estimate stage-dependent maturation parameters, the model either assumed that an individual could not die but rather would eventually move to the next stage, or that individuals could die but that the hazard rate was the same for each stage. In cases where dead individuals could not be counted, the stages were ignored and an overall hazard rate was estimated ([51]). In fact, there are many situations in which the number of dead individuals may be counted in distinct stages. Examples include monitoring the effect of diseases which progress through well-defined stages and the study of organisms in various stages of their life cycles. The hazard rates in the stages may not be the same and may depend on time.

Suppose there are $(I+1)$ stages, where stage $(I+1)$ is the final stage and the vector of stage-dependent maturation parameters that needs to be estimated has the form $\theta = (\theta_1, \theta_2, \ldots, \theta_I)$. For each individual, let $S_j, j = 1, 2, ..., I$ be the length of time spent in stage j, where we interpret $S_{I+1} = \infty$. Note that S_j is a random variable. Let $g_j(t),\ j = 1, ..., I$ be the probability density function of the time spent in stage j. In stage j, $g_j(t)$ is called the stage time distribution and is parametrised by the maturation parameters denoted by

$$\theta^{(j)} = \left(\theta_1^{(j)}, \theta_2^{(j)}, ..., \theta_{n_j}^{(j)}\right),\ j = 1, ..., I. \tag{3.1.1}$$

The main goal of this chapter is to estimate the stage-dependent maturation parameters $\theta^{(j)}, j = 1, ..., I$ in situations where non-trivial hazard rates apply. Specifically, two models for the death process will be considered. The first is the case of stage-wise constant hazard rates, and the second is the case of linear, time-dependent, hazard rates. This expands on the work of [51] and [25]. In their models, either stage-specific mortality did not occur or hazard rates were the same for each stage. This limitation is mentioned in [19]. Here, techniques based on Laplace transform matching ([3]; [36]; [51]) will be employed to estimate maturation parameters in non-trivial hazard rate models. In addition, we propose methods for estimating these non-trivial hazard rates.

Estimation techniques using Laplace transforms will also be discussed. In Section 3.2, the estimators of maturation parameters will be investigated under the assumption that the

death processes are known, both for stage-wise constant hazard rates and for linear time-dependent hazard rates. Estimators of the hazard rates in stage-wise constant hazard rates will be introduced in Section 3.3. In order to estimate the hazard rates separately from the maturation rates, some further properties need to be added to the model. By using estimators of the hazard rates in Section 3.3, we obtain estimators of the maturation process parameters. In Section 3.4, simulations are conducted to evaluate the methodology described in Sections 3.2 and 3.3.

The long-range goal is to use Markov chain Monte Carlo (MCMC) methods (Chapter 4) to estimate stage-dependent maturation parameters and hazard rate parameters at each stage. The convergence of MCMC in this context is poor unless reasonable estimators of the relationship of the stage-dependent maturation parameters in each stage are used as input. The main motivation for this chapter is to provide estimators for input into an MCMC of sufficient quality to improve convergence. The implementation of MCMC is deferred until later in this thesis.

There are several practical applications of estimating the stage-dependent maturation parameters. In the first model, exploring the development of the individual in each stage can give an understanding of the progress of the individual through its life cycle. For instance, multi-stage models have been used to study mice ([26]), parasites ([21]), Anguillicola crassus ([56]) and loggerhead sea turtles ([12]). The second model will allow for comparisons of treatments at various stages: for example, when monitoring the progression of disease stages. The effects of covariates could be investigated with the view of improving development in the stage of interest according to context-specific criteria.

3.2 Estimating stage parameters when hazard rates are known

3.2.1 Stage-wise constant hazard rates

In earlier research ([51]; [25]), it was assumed that the hazard rate in any given stage was a constant: $\mu_j(t) \equiv \mu$ for all t and $j = 1, ..., I$. In this section, we will assume that the $\mu_j(t) = \mu_j,\ j = 1, ..., I$ have been estimated, where μ_j need not be equal to μ_i when

$j \neq i$. We will seek to estimate the maturation parameters in the constant shape case and in the constant rate case. A method for estimating the hazard rates is proposed in Section 3.3.1. The expression for $p_j(t)$, the probability of an individual being alive in stage j at time t, in (2.1.2) leads to a difficult computation for estimating parameters if the maximum-likelihood method is used. The probability $p_j(t)$ has complicated forms when $g_j(t)$ has either exponential, Erlangian, or gamma distributions. On the other hand, the Laplace transforms $\psi_j(s)/s = \int_0^\infty p_j(t) \exp(-st)\, dt$ have quite simple and explicit forms, enabling the estimation of the parameters in $g_j(t)$.

Firstly, if each density functions $g_j(t)$ has an exponential distribution $g_j(t) = \lambda_j \exp(-\lambda_j t)$, $p_j(t)$ is calculated in [5] as

$$p_j(t) = \prod_{r=1}^{j-1} \lambda_r \sum_{h=1}^{j} \exp(-\rho_h t) \prod_{m=1, m\neq h}^{j} (\rho_m - \rho_h)^{-1}\,, \tag{3.2.1}$$

where it is assumed that $\rho_j = \mu_j + \lambda_j$ are all distinct, for $j = 1, ..., I$.

In addition to this, if the density functions $g_j(t)$, $j = 1, ..., I$ have Erlangian distributions with rate parameter λ_j and positive integer constant shape parameter $a_j \equiv a, j = 1, ..., I$, $g_j(t) = t^{a-1} \exp(-\lambda_j t)\, \lambda_j^a/(a-1)!$, then $p_j(t)$ is calculated in [51] as

$$\begin{aligned} p_j(t) = \prod_{l=1}^{j-1} \lambda_l^a \sum_{i=0}^{a-1} \Bigg\{ \lambda_j^i \Bigg[\frac{(-1)^{a-1}}{(a-1)!} \sum_{l=1}^{j-1} \frac{d^{a-1}}{d\rho^{a-1}} \left(\exp(-\rho t) \prod_{m=1, m\neq l}^{j-1} (\rho_m - \rho)^{-a} (\rho_i - \rho)^{-j-1} \right)_{\rho=\rho_l} \Bigg] \\ + \left[\frac{(-1)^i}{i!} \frac{d^i}{d\rho^i} \left(\exp(-\rho t) \prod_{m=1}^{j-1} (\rho_m - \rho)^{-a} \right)_{\rho=\rho_j} \right] \Bigg\}\,. \end{aligned} \tag{3.2.2}$$

Finally, if the density functions $g_j(t), j = 1, ..., I$ have gamma distributions with constant, positive rate parameter $\lambda_j \equiv \lambda$ and shape parameter a_j, $g_j(t) = t^{a_j - 1} \exp(-\lambda t)\, \lambda^{a_j}/\Gamma(a_j)$ then $p_j(t)$ is calculated by Schuh and Tweedie ([51]) as

$$p_j(t) = A_{j-1}(t) - A_j(t)\,, \tag{3.2.3}$$

where $A_j(t) = \exp(-\mu t) \int_0^t \left[\lambda (x\lambda)^{b(j)-1} \exp(-\lambda x) \right] dx/\Gamma(b(j))$ and $b(j) = \sum_{i=1}^{j} a_i$.

It is clear that even when the density functions $g_j(t)$ has Erlangian distribution, the probability $p_j(t)$ still has a complicated form. As a consequence, estimating the parameters in $g_j(t)$ is non-trivial.

By contrast, consider the case where the density functions $g_j(t)$, $j = 1, ..., I$ have gamma distributions with rate parameters λ_j and shape parameters a_j. Define

$$\begin{aligned}\beta_i(s) = \mathcal{L}\{h_i\}(s) = \mathcal{L}\{g_i(t)S_i(t)\}(s) = \int_0^\infty g_i(t)\exp(-s-\mu_i)tdt \\ = \left(\frac{\lambda_i}{\lambda_i+\mu_i+s}\right)^{a_i},\ i = 1, ..., j-1\,,\end{aligned} \tag{3.2.4}$$

where $S_i(t)$ is the probability of an organism being alive in stage i longer than time t ([15]).

The Laplace transform of the function $H_j(t)$, $j = 1, ..., I$ is given by

$$\begin{aligned}\mathcal{L}\{H_j(t)\}(s) = \mathcal{L}\left\{\exp(-\mu_j t)\int_t^\infty g_j(x)dx\right\}(s) = \int_0^\infty \exp(-\mu_j - s)t\left[1 - \int_0^t g_j(x)dx\right]dt \\ = \frac{1-\beta_j(s)}{s+\mu_j}\,.\end{aligned} \tag{3.2.5}$$

Then the Laplace transforms, $\psi_j(s)/s, j = 1, ..., I$ have explicit forms

$$\begin{aligned}\psi_j(s) &= s\int_0^\infty p_j(t)\exp(-st)\,dt \\ &= s\int_0^\infty h_1 * h_2 * ... * h_{j-1} * H_j(t)\exp(-st)\,dt \\ &= s\mathcal{L}\{h_1\}(s)\mathcal{L}\{h_2\}(s)...\mathcal{L}\{h_{j-1}\}(s)\mathcal{L}\{H_j\}(s) \\ &= s\beta_1(s)\beta_2(s)...\beta_{j-1}(s)\frac{1-\beta_j(s)}{s+\mu_j}\,.\end{aligned} \tag{3.2.6}$$

The estimation of the stage-dependent maturation parameters, $\theta = (\theta_1, \theta_2, \ldots, \theta_I)$ is introduced in [51] in which the death rates are assumed constant and the same throughout the stages. However, we will estimate the stage-dependent maturation parameters in the case when the hazard rates in each stage are constant but not all equal.

Theorem 1. *Consider single cohort stage-frequency data comprising the numbers of organisms in each stage $N_j(T_k)$, $j = 1, 2, .., I, k = 1, 2, ..., K$ at each random sampling time T_k. Assume that the probability density functions $g_j(t)$ of the time spent in stages $j = 1, 2, .., I$ are gamma distributions and that the hazard rates within stages, μ_j, $j = 1, 2, ..., I$ are known, constant and may vary across stages. Then estimators of the other parameters*

of the density functions $g_j(t)$ can be derived in the constant shape case and the constant rate case respectively as

$$\begin{aligned}\beta_1(s) &= 1 - \frac{(\mu_1 + s)}{s}\psi_1(s)\,, \\ \widehat{\beta_j}(s) &= 1 - \frac{(\mu_j + s)}{s}\frac{\widehat{\psi}_j(s)}{\widehat{\beta}_1(s).\widehat{\beta_2}(s)...\widehat{\beta_{j-1}}(s)}\,, \quad j = 2, ..., I\,,\end{aligned} \tag{3.2.7}$$

$$\widehat{\lambda}_j(s) = \frac{\mu_j + s}{\left[\widehat{\beta}_j(s)\right]^{-1/a} - 1}\,, \quad j = 1, ..., I\,, \tag{3.2.8}$$

and

$$\widehat{a}_j(s) = \frac{\log \widehat{\beta}_j(s)}{\log\left[\lambda/(\lambda + s + \mu_j)\right]}\,, \quad j = 1, ..., I\,. \tag{3.2.9}$$

Proof. Using the expression in (3.2.6), the function $\beta_1(s + \mu_1)$ can be written as

$$\beta_1(s) = 1 - \frac{(\mu_1 + s)}{s}\psi_1(s)\,. \tag{3.2.10}$$

Recalling the estimator of $\psi_j(s)$ in (2.2.3), we have

$$\widehat{\psi_j}(s) = \sum_{k=1}^{K} I_k \frac{N_j(T_k)}{N_k}\,, \quad j = 1, ..., I. \tag{3.2.11}$$

Then the function $\beta_1(s)$ can be estimated as

$$\begin{aligned}\widehat{\beta}_1(s) &= 1 - (\mu_1 + s)\widehat{\psi}_1(s)/s \\ &= 1 - \frac{(\mu_1 + s)}{s}\sum_{k=1}^{K} I_k \frac{N_1(T_k)}{N_k}\,.\end{aligned} \tag{3.2.12}$$

Similarly, from (3.2.6) $\beta_2(s + \mu_1)$ can be estimated as

$$\widehat{\beta}_2(s) = 1 - \frac{(\mu_2 + s)}{s}\frac{\widehat{\psi}_2(s)}{\widehat{\beta}_1(s)} \quad . \tag{3.2.13}$$

In general, the estimators of $\beta_j(s),\ j = 1, ..., I$ are given

$$\widehat{\beta_j}(s) = 1 - \frac{(\mu_j + s)}{s}\frac{\widehat{\psi}_j(s)}{\widehat{\beta}_1(s).\widehat{\beta_2}(s)...\widehat{\beta_{j-1}}(s)}\,,\ j = 1, ..., I \quad . \tag{3.2.14}$$

Using (3.2.14) and (3.2.4), the other parameters of the density functions $g_j(t)$ can be derived in constant shape case and constant rate case as follows

i) The constant shape case $a_j \equiv a, j = 1, ..., I$.

Rearrangement of (3.2.4), an expression for $\lambda_j(s)$ in terms of the constant shape a is

$$\begin{aligned} \beta_j(s) &= \left(\frac{\lambda_j}{\lambda_j + \mu_j + s}\right)^{a_j}, \ j = 1, ..., I \\ \iff \lambda_j &= [\beta_j(s)]^{-1/a} (\lambda_j + \mu_j + s) \\ \iff \lambda_j &= \frac{\mu_j + s}{[\beta_j(s)]^{-1/a} - 1} . \end{aligned} \tag{3.2.15}$$

A complete set of explicit estimators in (3.2.8) are derived as

$$\widehat{\lambda}_j(s) = \frac{\mu_j + s}{\left[\widehat{\beta}_j(s)\right]^{-1/a} - 1}, \quad j = 1, ..., I .$$

ii) Similarly in the constant rate case $\lambda_j \equiv \lambda, j = 1, ..., I$.

Rearrangement of (3.2.4) leads to an expression for $a_j(s)$ in terms of the constant rate λ is as

$$a_j = \frac{\log \beta_j(s)}{\log [\lambda/(\lambda + s + \mu_j)]} . \tag{3.2.16}$$

A complete set of explicit estimators in (3.2.9) are derived as

$$\widehat{a}_j(s) = \frac{\log \widehat{\beta}_j(s)}{\log [\lambda/(\lambda + s + \mu_j)]}, \quad j = 1, ..., I.$$

□

Next, we introduce a method to estimate the stage-dependent maturation parameters when the hazard rate in each stage has a linear relation with time t.

3.2.2 Linear time-dependent hazard rates

In the case of non-trivial hazard rates, [34] proposed a model in which hazard rates were dependent on the stage. The number of individuals in each stage was assessed repeatedly from a small cohort. This expanded upon an assumption of a previous model ([51]; [25]). However, our model requires destructive sampling in order to assess the stage reached by each individual. The hazard rate in each stage is assumed to depend on the

starting time as in [25]. In this model, the hazard rate in each stage is modelled as $\mu_j(t) = \gamma_{j0} + \gamma_{j1}t, \gamma_{j1} \neq 0, j = 1, ..., I$. We assume that γ_{j0} and γ_{j1} are both positive since $\mu_j(t) > 0, \forall t \geq 0$. The Laplace transform method is applied in order to estimate the stage parameters. In some stages, the hazard rates may be constant, while in other stages the hazard rates may depend on time. Such cases will be mentioned in Section 3.4.2.

The probability that an independent organism is alive in stage j at time t is given in (2.1.2) by

$$p_j(t) = h_1 * h_2 * ... * h_{j-1} * H_j(t) \,,$$

where

$$\begin{aligned} h_i(t) &= g_i(t)S_i(t) = g_i(t) \exp\left(-\int_0^t \mu_i(x)dx\right) \\ &= g_i(t) \exp\left(-\gamma_{i0}t - \gamma_{i1}\frac{t^2}{2}\right), i = 1, ..., j-1 \,, \\ H_j(t) &= S_j(t) \int_t^\infty g_j(x)dx \\ &= \exp\left(-\gamma_{j0}t - \gamma_{j1}\frac{t^2}{2}\right) \int_t^\infty g_j(x)dx \,. \end{aligned} \tag{3.2.17}$$

In order to find the relation between the rate parameter and the shape parameter in each stage, let us consider an approximation to the complementary error function.

An approximation to the complementary error function is given by Chiani *et al* ([13]) as follows:

$$\text{erfc}(x) \approx \frac{1}{6}\exp\left(-x^2\right) + \frac{1}{2}\exp\left(-\frac{4}{3}x^2\right), \quad x > 0.5 \quad . \tag{3.2.18}$$

The right side of (3.2.18) is still difficult to use in the present context. This expression can be further approximated with respect to $\exp(-x^2)$ as (3.2.19).

Away from $x = 0$, the complementary error function (erfc(x)) is well approximated by a function of the form $A\exp(-x^2)$

$$\text{erfc}(x) \approx 0.6\exp\left(-x^2\right), \quad x > 0.5 \,. \tag{3.2.19}$$

To find the best value of A on the interval of interest, namely $[0.5, 6]$, the absolute difference ($\|\text{erfc}(x) - A\exp(-x^2)\|_1$) was computed numerically. We used piece-wise constant integration on 1000 points with the values of A ranging from 0.01 to 2. The result showed

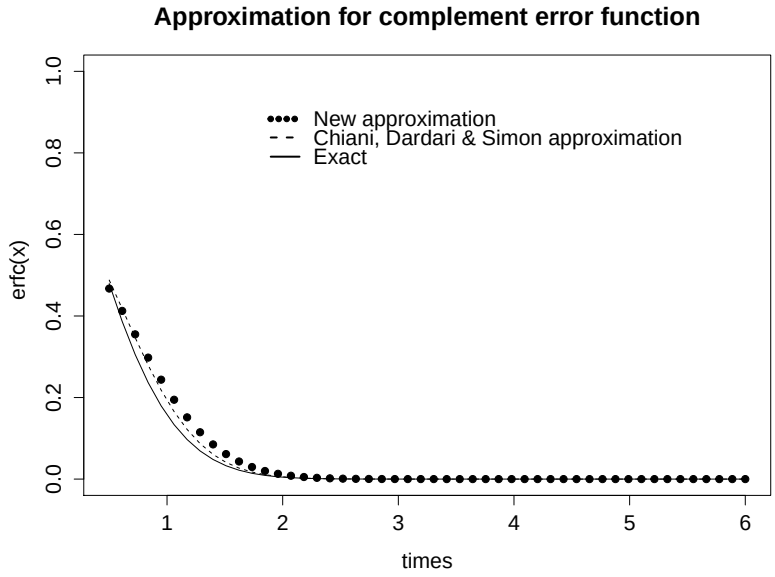

Figure 3.1: Comparison among exponential bounds on the $erfc(x), x > 0.5$

that $A \approx 0.6$ gave the best fit. This approximation is illustrated graphically in Figure 3.1.

This leads us to estimate the Laplace transform $(\psi_j(s)/s)$ of $p_j(t)$ as follows

Lemma 1. *The Laplace transform $\psi_j(s)/s$ of $p_j(t)$ is approximately given by*

$$\psi_j(s) \approx s\beta_1(s)\beta_2(s)...\beta_{j-1}(s)\left\{exp\left(\frac{(s+\gamma_{j0})^2}{2\gamma_{j1}}\right)\sqrt{\frac{\pi}{2\gamma_{j1}}}\left[1-erf\left(\sqrt{\frac{\gamma_{j1}}{2}}\frac{(s+\gamma_{j0})}{\gamma_{j1}}\right)\right]\right.$$
$$\left.-0.6\sqrt{\frac{\pi}{2\gamma_{j1}}}\beta_j(s)\right\}, \quad j=1,...,I\,. \tag{3.2.20}$$

Proof. From (3.2.6), the form of $\psi_j(s)/s$ can be written as

$$\begin{aligned}\psi_j(s) &= s\int_0^\infty p_j(t)\exp(-st)\,dt \\ &= s\int_0^\infty h_1 * h_2 * ... * h_{j-1} * H_j(t)\exp(-st)\,dt \\ &= s\mathcal{L}\{h_1\}(s)\mathcal{L}\{h_2\}(s)...\mathcal{L}\{h_{j-1}\}(s)\mathcal{L}\{H_j(t)\}(s) \\ &= s\beta_1(s)\beta_2(s)...\beta_{j-1}(s)\mathcal{L}\{H_j(t)\}(s)\,.\end{aligned} \tag{3.2.21}$$

Recalling definition (3.2.4), in this case the functions $\beta_j(s), j = 1, ..., I$ are

$$\begin{aligned}\beta_j(s) &= \mathcal{L}\left\{h_j\right\}(s) = \mathcal{L}\left\{g_j(t)S_j(t)\right\}(s) \\ &= \int_0^\infty g_j(t)\exp\left(-\gamma_{j0}t - \gamma_{j1}\frac{t^2}{2}\right)\exp\left(-st\right)dt\,, \quad j = 1, ..., I.\end{aligned} \tag{3.2.22}$$

The Laplace transform $\mathcal{L}\left\{H_j(t)\right\}(s)$ in (3.2.21) is calculated as

$$\begin{aligned}\mathcal{L}\left\{H_j(t)\right\}(s) &= \int_0^\infty \exp\left(-\gamma_{j0}t - \gamma_{j1}\frac{t^2}{2}\right)\left(\int_t^\infty g_j(x)dx\right)\exp\left(-st\right)dt \\ &= \int_0^\infty \exp\left(-\gamma_{j0}t - \gamma_{j1}\frac{t^2}{2}\right)\left(1 - \int_0^t g_j(x)dx\right)\exp\left(-st\right)dt \\ &= \int_0^\infty \exp\left(-\gamma_{j0}t - \gamma_{j1}\frac{t^2}{2}\right)\exp\left(-st\right)dt \\ &\quad - \int_0^\infty \exp\left(-\gamma_{j0}t - \gamma_{j1}\frac{t^2}{2}\right)\left(\int_0^t g_j(x)dx\right)\exp\left(-st\right)dt\,.\end{aligned} \tag{3.2.23}$$

Integrating of the first elements in the right hand side (3.2.23) reveals that

$$\begin{aligned}&\int_0^\infty \exp\left(-\gamma_{j0}t - \gamma_{j1}\frac{t^2}{2}\right)\exp\left(-st\right)dt \\ &= \int_0^\infty \exp\left(-(s+\gamma_{j0})t - \gamma_{j1}\frac{t^2}{2}\right)dt \\ &= \int_0^\infty \exp\left[-\frac{\gamma_{j1}}{2}\left(t + \frac{(s+\gamma_{j0})}{\gamma_{j1}}\right)^2 + \frac{(s+\gamma_{j0})^2}{2\gamma_{j1}}\right]dt\end{aligned} \tag{3.2.24}$$

Changing the variable $u = t + u_{j0} = t + (s + \gamma_{j0})/\gamma_{j1}$ in (3.2.24), we have

$$\begin{aligned}&\int_0^\infty \exp\left(-\gamma_{j0}t - \gamma_{j1}\frac{t^2}{2}\right)\exp\left(-st\right)dt \\ &= \int_{u_{j0}}^\infty \exp\left[-\frac{\gamma_{j1}}{2}u^2 + \frac{(s+\gamma_{j0})^2}{2\gamma_{j1}}\right]du \\ &= \exp\left(\frac{\gamma_{j1}}{2}u_{j0}^2\right)\int_{u_{j0}}^\infty \exp\left(-\frac{\gamma_{j1}}{2}u^2\right)du\end{aligned} \tag{3.2.25}$$

$$
\begin{aligned}
&= \sqrt{\frac{2}{\gamma_{j1}}} \exp\left(\frac{\gamma_{j1}}{2} u_{j0}^2\right) \int_{\sqrt{\gamma_{j1}/2}u_{j0}}^{\infty} \exp\left(-v^2\right) dv, \quad v = \sqrt{\frac{\gamma_{j1}}{2}} u \\
&= \sqrt{\frac{\pi}{2\gamma_{j1}}} \exp\left(\frac{\gamma_{j1}}{2} u_{j0}^2\right) \left[1 - \operatorname{erf}\left(\sqrt{\frac{\gamma_{j1}}{2}} u_{j0}\right)\right] .
\end{aligned}
$$

Integrating of the second elements in the right hand side (3.2.23), we obtain

$$
\begin{aligned}
&\int_0^{\infty} \exp\left(-\gamma_{j0} t - \gamma_{j1} \frac{t^2}{2}\right) \left(\int_0^t g_j(x) dx\right) \exp\left(-st\right) dt \\
&= \mathcal{L}\left(\exp\left(-\gamma_{j0} t - \gamma j1 \frac{t^2}{2}\right) \int_0^t g_j(x) dx\right) \\
&= \int_0^{\infty} \int_x^{\infty} \exp\left(-st\right) \exp\left(-\gamma_{j0} t - \gamma_{j1} \frac{t^2}{2}\right) g_j(x) dt dx .
\end{aligned}
\tag{3.2.26}
$$

Rewriting the double integrations (3.2.26), we obtain

$$
\begin{aligned}
&\int_0^{\infty} \exp\left(-\gamma_{j0} t - \gamma_{j1} \frac{t^2}{2}\right) \left(\int_0^t g_j(x) dx\right) \exp\left(-st\right) dt \\
&= \int_0^{\infty} g_j(x) \int_x^{\infty} \exp\left(-st\right) \exp\left(-\gamma_{j0} t - \gamma_{j1} \frac{t^2}{2}\right) dt dx \\
&= \int_0^{\infty} g_j(x) \left(\int_x^{\infty} \exp\left(-st - \gamma_{j0} t - \gamma_{j1} \frac{t^2}{2}\right) dt\right) dx \\
&= \int_0^{\infty} g_j(x) \left(\int_x^{\infty} \exp\left(-\frac{1}{2}\left(\left(\sqrt{\gamma_{j1}} t + \frac{\gamma j0 + s}{\sqrt{\gamma_{j1}}}\right)^2 - \frac{(\gamma_{j0} + s)^2}{\gamma j1}\right)\right) dt\right) dx \\
&= \int_0^{\infty} g_j(x) \exp\left(\frac{(\gamma_{j0} + s)^2}{2\gamma_{j1}}\right) \left(\int_x^{\infty} \exp\left(-\frac{1}{2}(\sqrt{\gamma_{j1}} t + \frac{\gamma_{j0} + s}{\sqrt{\gamma_{j1}}})^2\right) dt\right) dx .
\end{aligned}
\tag{3.2.27}
$$

Changing the variable $u = \sqrt{\gamma_{j1}} t + (\gamma_{j0} + s) / \sqrt{\gamma_{j1}}$ in (3.2.27), we have

$$
\begin{aligned}
&\int_0^{\infty} \exp\left(-\gamma_{j0} t - \gamma_{j1} \frac{t^2}{2}\right) \left(\int_0^t g_j(x) dx\right) \exp\left(-st\right) dt \\
&= \int_0^{\infty} g_j(x) \exp\left(\frac{(\gamma_{j0} + s)^2}{2\gamma_{j1}}\right) \frac{1}{\sqrt{\gamma_{j1}}} \left(\int_{\sqrt{\gamma_{j1}} x + (\gamma_{j0} + s)/\sqrt{\gamma_{j1}}}^{\infty} \exp\left(-\frac{1}{2} u^2\right) du\right) dx
\end{aligned}
\tag{3.2.28}
$$

$$= \frac{1}{\sqrt{\gamma_{j1}}} \exp\left(\frac{(\gamma_{j0}+s)^2}{2\gamma_{j1}}\right) \int_0^\infty g_j(x) \left(\int_{\sqrt{\gamma_{j1}}x+(\gamma_{j0}+s)/\sqrt{\gamma_{j1}}}^\infty \exp\left(-\frac{1}{2}u^2\right) du \right) dx$$

$$= \exp\left(\frac{(\gamma_{j0}+s)^2}{2\gamma_{j1}}\right) \sqrt{\frac{\pi}{2\gamma_{j1}}} \int_0^\infty g_j(x) \left(1 - \text{erf}\left(\frac{1}{\sqrt{2}}\left(\sqrt{\gamma_{j1}}x + \frac{\gamma_{j0}+s}{\sqrt{\gamma_{j1}}}\right)\right)\right) dx$$

$$= \exp\left(\frac{(\gamma_{j0}+s)^2}{2\gamma_{j1}}\right) \sqrt{\frac{\pi}{2\gamma_{j1}}} \int_0^\infty g_j(x) \text{erfc}\left(\frac{1}{\sqrt{2}}\left(\sqrt{\gamma_{j1}}x + \frac{\gamma_{j0}+s}{\sqrt{\gamma_{j1}}}\right)\right) dx \,.$$

Following the approximation in (3.2.18) for the complementary error function

$$\text{erfc}\left[\frac{1}{\sqrt{2}}\left(\sqrt{\gamma_{j1}}x + \frac{\gamma_{j0}+s}{\sqrt{\gamma_{j1}}}\right)\right] ,$$

for $x \in [0,\infty)$ and γ_{j0} and γ_{j1} assumed both positive, when

$$\left(\sqrt{\gamma_{j1}}x + \frac{(\gamma_{j0}+s)}{\sqrt{\gamma_{j1}}}\right)/\sqrt{2} \leq 0.5$$

has negligible effect on the approximation. Equation (3.2.28) gives

$$\begin{aligned}
&\int_0^\infty \exp\left(-\gamma_{j0}t - \gamma_{j1}\frac{t^2}{2}\right)\left(\int_0^t g_j(x)dx\right)\exp(-st)\,dt \\
&\approx \exp\left(\frac{(\gamma_{j0}+s)^2}{2\gamma_{j1}}\right)\sqrt{\frac{\pi}{2\gamma_{j1}}}\int_0^\infty g_j(x)\left[0.6\exp\left(-\frac{1}{2}\left(\sqrt{\gamma_{j1}}x+\frac{\gamma_{j0}+s}{\sqrt{\gamma_{j1}}}\right)^2\right)\right]dx \\
&\approx \exp\left(\frac{(\gamma_{j0}+s)^2}{2\gamma_{j1}}\right)\sqrt{\frac{\pi}{2\gamma_{j1}}}\times \\
&\int_0^\infty g_j(x)\left[0.6\exp\left(-\frac{\gamma_{j1}}{2}x^2-(\gamma_{j0}+s)x-\frac{(\gamma_{j0}+s)^2}{2\gamma_{j1}}\right)\right]dx \\
&\approx 0.6\sqrt{\frac{\pi}{2\gamma_{j1}}}\int_0^\infty g_j(x)\exp\left(-\frac{\gamma_{j1}}{2}x^2-(\gamma_{j0}+s)x\right)dx \\
&\approx 0.6\sqrt{\frac{\pi}{2\gamma_{j1}}}\beta_j(s)\,.
\end{aligned} \tag{3.2.29}$$

Substituting (3.2.25) and (3.2.29) in (3.2.23), the assertion is proved. □

Lemma 2. *The estimators of the functions $\beta_j(s), j = 1,...,I$, in (3.2.22) are given by*

$$\widehat{\beta_j}(s) = \frac{1}{0.6s}\sqrt{\frac{2\gamma_{j1}}{\pi}}\left(s\widehat{b}_j(s) - \frac{\widehat{\psi}_j(s)}{\widehat{\beta}_1(s)\widehat{\beta}_2(s)...\widehat{\beta}_{j-1}(s)}\right), \quad j = 1,...,I, \tag{3.2.30}$$

where

$$\widehat{b}_j(s) = \exp\left(u_{jo}^2\right)\sqrt{\frac{\pi}{2\gamma_{j1}}}\left[1 - erf(u_{jo})\right] , \quad j = 1, ..., I ,$$

and $u_{jo} = (s + \gamma_{j0})/\sqrt{2\gamma_{j1}} , \quad j = 1, ..., I .$

Proof. Recalling (2.2.3), the consistent estimators of $\psi_j(s), j = 1, ..., I$ are given by

$$\widehat{\psi_j}(s) = \sum_{k=1}^{K} I_k \frac{N_j(T_k)}{N_k} , \quad j = 1, ..., I .$$

Using (3.2.20) and Lemma 1 in order to obtain the form of $\psi_j(s) , j = 1, ..., I$, we have

$$\begin{aligned}
\psi_1(s) &= s\left\{\exp\left(u_{1o}^2\right)\sqrt{\frac{\pi}{2\gamma_{11}}}\left[1 - \operatorname{erf}(u_{1o})\right] - 0.6\sqrt{\frac{\pi}{2\gamma_{11}}}\beta_1(s)\right\} \\
&= s\left(b_1(s) - 0.6\sqrt{\frac{\pi}{2\gamma_{11}}}\beta_1(s)\right), \\
b_1(s) &= \exp\left(\frac{(s+\gamma_{10})^2}{2\gamma_{11}}\right)\sqrt{\frac{\pi}{2\gamma_{11}}}\left[1 - \operatorname{erf}\left(\sqrt{\frac{\gamma_{11}}{2}}\frac{(s+\gamma_{10})}{\gamma_{11}}\right)\right] \\
\iff \beta_1(s) &= \frac{1}{0.6s}\sqrt{\frac{2\gamma_{j1}}{\pi}}\left(sb_1(s) - \psi_1(s)\right) .
\end{aligned} \tag{3.2.31}$$

Then, the estimator of $\beta_1(s)$ is given as

$$\begin{aligned}
\widehat{\beta}_1(s) &= \frac{1}{0.6s}\sqrt{\frac{2\gamma_{j1}}{\pi}}\left(s\widehat{b}_1(s) - \widehat{\psi}_1(s)\right) \\
&= \frac{1}{0.6s}\sqrt{\frac{2\gamma_{j1}}{\pi}}\left(s\widehat{b}_1(s) - \sum_{k=1}^{K} I_j \frac{N_j(T_k)}{N_k}\right) ,
\end{aligned} \tag{3.2.32}$$

where

$$\widehat{b}_1(s) = \exp\left(u_{1o}^2\right)\sqrt{\frac{\pi}{2\gamma_{11}}}\left[1 - \operatorname{erf}(u_{1o})\right] . \tag{3.2.33}$$

Similarly,

$$\begin{aligned}
\psi_2(s) &= s\beta_1(s)\left\{\exp\left(u_{2o}^2\right)\sqrt{\frac{\pi}{2\gamma_{21}}}\left[1 - \operatorname{erf}(u_{2o})\right] - 0.6\sqrt{\frac{\pi}{2\gamma_{21}}}\beta_2(s)\right\} \\
&= s\beta_1(s)\left\{b_2(s) - 0.6\sqrt{\frac{\pi}{2\gamma_{21}}}\beta_2(s)\right\}, \; b_2(s) = \exp\left(u_{2o}^2\right)\sqrt{\frac{\pi}{2\gamma_{21}}}\left[1 - \operatorname{erf}(u_{2o})\right] ,
\end{aligned} \tag{3.2.34}$$

and so $\beta_2(s)$ is estimated as

$$\widehat{\beta}_2(s) = \frac{1}{0.6s}\sqrt{\frac{2\gamma_{j1}}{\pi}}\left(s\widehat{b}_2(s) - \frac{\widehat{\psi}_2(s)}{\widehat{\beta}_1(s)}\right) , \tag{3.2.35}$$

where

$$\widehat{b}_2(s) = \exp\left(u_{2o}^2\right)\sqrt{\frac{\pi}{2\gamma_{21}}}\left[1-\operatorname{erf}\left(u_{2o}\right)\right] . \tag{3.2.36}$$

Inductively, the general estimators of the $\beta_j(s)$ are given by

$$\widehat{\beta_j}(s) = \frac{1}{0.6s}\sqrt{\frac{2\gamma_{j1}}{\pi}}\left(s\widehat{b}_j(s) - \frac{\widehat{\psi}_j(s)}{\widehat{\beta}_1(s)\widehat{\beta}_2(s)...\widehat{\beta}_{j-1}(s)}\right) , \quad j = 1, ..., I ,$$

where

$$\widehat{b}_j(s) = \exp\left(u_{jo}^2\right)\sqrt{\frac{\pi}{2\gamma_{j1}}}\left[1-\operatorname{erf}\left(u_{jo}\right)\right] , \quad j = 1, ..., I .$$

□

Theorem 2. *Assume that the probability density function $g_j(t)$ of the time spent in stage j , $j = 1, 2, .., I$ is a gamma distribution with a positive integer shape parameter and the hazard rate in each stage depends on time t and is modelled as*

$$\mu_j(t) = \gamma_{j0} + \gamma_{j1}t , \quad j = 1, ..., I, \gamma_{j0} \geq 0 \text{ and } \gamma_{j1} > 0 ,$$

where $\gamma_{j0},\ j = 1, ..., I$ and $\gamma_{j1},\ j = 1, ..., I$ are given. Then, the relation between the rate parameter and the integer shape parameter in each stage is given by

$$\begin{aligned}\beta_j(s) = \exp\left(\frac{(\gamma_{j0}+\lambda_j+s)^2}{2\gamma_{j1}}\right)\frac{\lambda_j^{a_j}}{(a_j-1)!}\sum_{k=0}^{a_j-1}(-1)^{a_j-1-k}\binom{a_j-1}{k}\frac{1}{2}\left(\frac{2}{\gamma_{j1}}\right)^{(k+1)/2} \\ \left[\frac{(\gamma_{j0}+\lambda_j+s)}{\gamma_{j1}}\right]^{a_j-1-k}\Gamma\left(\frac{\gamma_{j1}}{2}\left[\frac{(\gamma_{j0}+\lambda_j+s)}{\gamma_{j1}}\right]^2, \frac{k+1}{2}\right) , \quad j = 1, 2, ..., I ,\end{aligned} \tag{3.2.37}$$

where $\Gamma(x, a)$ is the upper incomplete gamma function

$$\Gamma(x, a) = \int_x^\infty \exp\left(-t\right)t^{a-1}dt .$$

Proof. Because shape parameter a_j is a positive integer, (3.2.22) may be used to write

$$\begin{aligned}\beta_j(s) &= \int_0^\infty g_j(t)\exp\left(-\gamma_{j0}t - \gamma_{j1}\frac{t^2}{2}\right)\exp\left(-st\right)dt \\ &= \frac{\lambda_i^{a_j}}{(a_j-1)!}\int_0^\infty \exp\left(-(\gamma_{j0}+\lambda_j+s)t - \gamma_{j1}\frac{t^2}{2}\right)t^{a_j-1}dt \\ &= \frac{\lambda_j^{a_j}}{(a_j-1)!}\int_0^\infty \exp\left[-\frac{\gamma_{j1}}{2}\left(t+\frac{(\gamma_{j0}+\lambda_j+s)}{\gamma_{i1}}\right)^2 + \frac{(\gamma_{j0}+\lambda_j+s)^2}{2\gamma_{j1}}\right]t^{a_j-1}dt\end{aligned} \tag{3.2.38}$$

$$= \exp\left(\frac{(\gamma_{j0}+\lambda_j+s)^2}{2\gamma_{j1}}\right)\frac{\lambda_j^{a_j}}{(a_j-1)!}\times$$
$$\int_0^\infty \exp\left[-\frac{\gamma_{j1}}{2}\left(t+\frac{(\gamma_{j0}+\lambda_j+s)}{\gamma_{j1}}\right)^2\right]t^{a_j-1}dt,\ j=1,2,...,I\ .$$

By the change of variable $u = t + u_0 = t + (\gamma_{j0}+\lambda_j+s)/\gamma_{j1}$ in (3.2.38), we have

$$\begin{aligned}
\beta_j(s) &= \exp\left(\frac{\gamma_{j1}}{2}u_0^2\right)\frac{\lambda_j^{a_j}}{(a_j-1)!}\int_{u_0}^{\infty}\exp\left(-\frac{\gamma j1}{2}u^2\right)(u-u_0)^{a_j-1}du \\
&= \exp\left(\frac{\gamma_{j1}}{2}u_0^2\right)\frac{\lambda_j^{a_j}}{(a_j-1)!}\int_{u_0}^{\infty}\exp\left(-\frac{\gamma_{j1}}{2}u^2\right)\sum_{k=0}^{a_j-1}(-1)^{a_j-1-k}\binom{a_j-1}{k}u^k u_0^{a_j-1-k}du \\
&= \exp\left(\frac{\gamma_{j1}}{2}u_0^2\right)\frac{\lambda_j^{a_j}}{(a_j-1)!}\times \\
&\sum_{k=0}^{a_j-1}(-1)^{a_j-1-k}\binom{a_j-1}{k}u_0^{a_j-1-k}\int_{u_0}^{\infty}\exp\left(-\frac{\gamma_{j1}}{2}u^2\right)u^k du\ .
\end{aligned} \tag{3.2.39}$$

By the change of variable $v = \gamma_{j1}u^2/2$ in (3.2.39), we have

$$\begin{aligned}
\beta_j(s) &= \exp\left(\frac{\gamma_{j1}}{2}u_0^2\right)\frac{\lambda_j^{a_j}}{(a_j-1)!}\times \\
&\sum_{k=0}^{a_j-1}(-1)^{a_j-1-k}\binom{a_j-1}{k}u_0^{a_j-1-k}\frac{1}{2}(\frac{2}{\gamma_{j1}})^{(k+1)/2}\int_{\gamma_{j1}u_0^2/2}^{\infty}\exp(-v)\,v^{(k-1)/2}dt \\
&= \exp\left(\frac{\gamma_{j1}}{2}u_0^2\right)\frac{\lambda_j^{a_j}}{(a_j-1)!}\sum_{k=0}^{a_j-1}(-1)^{a_j-1-k}\binom{a_j-1}{k}u_0^{a_j-1-k}\frac{1}{2}\times \\
&\Gamma\left(\frac{\gamma_{j1}}{2}u_0^2,\frac{k+1}{2}\right)\left(\frac{2}{\gamma_{j1}}\right)^{(k+1)/2},\ j=1,2,...,I\ ,
\end{aligned} \tag{3.2.40}$$

which give us (3.2.37). □

By Lemma 2 and (3.2.40), the relationship between the rate parameters and the shape parameters in stage j is shown to be

$$\widehat{\beta}_j(s) = \exp\left(\frac{(\gamma_{j0}+\hat{\lambda}_j+s)^2}{2\gamma_{j1}}\right)\frac{\hat{\lambda}_j^{\hat{a}_j}}{(\hat{a}_j-1)!}\sum_{k=0}^{\hat{a}_j-1}(-1)^{\hat{a}_j-1-k}\binom{\hat{a}_j-1}{k}\frac{1}{2}\left(\frac{2}{\gamma_{j1}}\right)^{(k+1)/2}\times$$
$$\left[\frac{(\gamma_{j0}+\hat{\lambda}_j+s)}{\gamma_{j1}}\right]^{\hat{a}_j-1-k}\Gamma\left(\frac{\gamma_{j1}}{2}\left[\frac{(\gamma_{j0}+\hat{\lambda}_j+s)}{\gamma_{j1}}\right]^2,\frac{k+1}{2}\right)\ .$$

In the constant shape case in which $a_j \equiv a,\ j = 1, ..., I$ are a known positive integer, the rate parameters $\lambda_j,\ j = 1, ..., I$ can be estimated by solving (3.2.37) numerically. Conversely, in the constant rate case in which $\lambda_j \equiv \lambda,\ j = 1, ..., I$ are known, then the shape parameters $a_j,\ j = 1, ..., I$ can be estimated by solving (3.2.37) numerically.

Lemma 3. *The shape parameters $a_j, j = 1, ..., I$ of the maturation distributions $g_j(t), j = 1, ..., I$ in Theorem 2 are positive integers. If the shape parameters are positive numbers but not necessarily integers, the relation between the rate parameter and the shape parameter in each stage is given by*

$$\begin{aligned}\beta_j(s) \approx \exp\left(\frac{(\gamma_{j0}+\lambda_j+s)^2}{2\gamma_{i1}}\right)\frac{\lambda_j^{a_j}}{\Gamma(a_j)}\left[\frac{1}{\gamma_{j1}}\left(\frac{2}{\gamma_{j1}}\right)^{(a_j-2)/2}\Gamma\left(\frac{\gamma_{j1}}{2}\left[\frac{(\gamma_{j0}+\lambda_j+s)}{\gamma_{j1}}\right]^2, \frac{a_j}{2}\right)\right.\\ \left.-\frac{1}{\gamma_{j1}}\left(\frac{2}{\gamma_{j1}}\right)^{(a_j-3)/2}(a_j-1)\frac{(\gamma_{j0}+\lambda_j+s)}{\gamma_{j1}}\Gamma\left(\frac{\gamma_{j1}}{2}\left[\frac{(\gamma_{j0}+\lambda_j+s)}{\gamma_{j1}}\right]^2, \frac{a_j-1}{2}\right)\right].\end{aligned} \tag{3.2.41}$$

Proof. When a_j is a real number, we apply the approximation $(u-u_0)^{a_j-1} \approx u^{a_j-1} - u^{a_j-2}(a_j-1)u_0$, where $u_0 = (\gamma_{j0}+\lambda_j+s)/\gamma_{j1}$ for (3.2.39) to obtain

$$\begin{aligned}\beta_j(s) &= \exp\left(\frac{\gamma_{j1}}{2}u_0^2\right)\frac{\lambda_j^{a_j}}{\Gamma(a_j)}\int_{u_0}^{\infty}\exp\left(-\frac{\gamma j1}{2}u^2\right)(u-u_0)^{a_j-1}du\\ &\approx \exp\left(\frac{\gamma_{j1}}{2}u_0^2\right)\frac{\lambda_j^{a_j}}{\Gamma(a_j)}\int_{u_0}^{\infty}\left[\exp\left(-\frac{\gamma_{j1}}{2}u^2\right)u^{a_j-1}\right.\\ &\left.-\exp\left(-\frac{\gamma_{j1}}{2}u^2\right)u^{a_j-2}(a_j-1)u_0\right]du.\end{aligned} \tag{3.2.42}$$

By the change of variable $v = \gamma_{j1}u^2/2$ in (3.2.42), we have

$$\begin{aligned}\beta_j(s) \approx \exp\left(\frac{\gamma_{j1}}{2}u_0^2\right)\frac{\lambda_j^{a_j}}{\Gamma(a_j)}\left(\frac{1}{\gamma_{j1}}\left(\frac{2}{\gamma_{j1}}\right)^{(a_j-2)/2}\int_{\frac{\gamma_{j1}}{2}u_0^2}^{\infty}\exp(-v)\,v^{(a_j-2)/2}\right.\\ \left.-\frac{1}{\gamma_{j1}}\left(\frac{2}{\gamma_{j1}}\right)^{(a_j-3)/2}(a_j-1)u_0\int_{\frac{\gamma_{j1}}{2}u_0^2}^{\infty}\exp(-v)v^{(a_j-3)/2}\right)dt\end{aligned} \tag{3.2.43}$$

$$\approx \exp\left(\frac{\gamma_{j1}}{2}u_0^2\right)\frac{\lambda_j^{a_j}}{\Gamma(a_j)}\left[\frac{1}{\gamma_{j1}}\left(\frac{2}{\gamma_{j1}}\right)^{(a_j-2)/2}\Gamma\left(\frac{\gamma_{j1}}{2}u_0^2,\frac{a_j}{2}\right)\right.$$
$$\left.-\frac{1}{\gamma_{j1}}\left(\frac{2}{\gamma_{j1}}\right)^{(a_j-3)/2}(a_j-1)u_0\Gamma\left(\frac{\gamma_{j1}}{2}u_0^2,\frac{a_j-1}{2}\right)\right],$$

which give us (3.2.41). □

By Lemma 2 and (3.2.41), the relationship between the rate parameters and the shape parameters is shown to be

$$\widehat{\beta}_j(s) \approx \exp\left(\frac{(\gamma_{j0}+\hat{\lambda}_j+s)^2}{2\gamma_{i1}}\right)\frac{\hat{\lambda}_j^{\hat{a}_j}}{\Gamma(a_j)}\left[\frac{1}{\gamma_{j1}}\left(\frac{2}{\gamma_{j1}}\right)^{(\hat{a}_j-2)/2}\Gamma\left(\frac{\gamma_{j1}}{2}\left[\frac{(\gamma_{j0}+\hat{\lambda}_j+s)}{\gamma_{j1}}\right]^2,\frac{\hat{a}_j}{2}\right)\right.$$
$$\left.-\frac{1}{\gamma_{j1}}\left(\frac{2}{\gamma_{j1}}\right)^{(\hat{a}_j-3)/2}(\hat{a}_j-1)\frac{(\gamma_{j0}+\hat{\lambda}_j+s)}{\gamma_{j1}}\Gamma\left(\frac{\gamma_{j1}}{2}\left[\frac{(\gamma_{j0}+\hat{\lambda}_j+s)}{\gamma_{j1}}\right]^2,\frac{\hat{a}_j-1}{2}\right)\right].$$

When $a_j,\ j=1,...,I$ are considered as positive numbers , then the rate parameters $\lambda_j,\ j=1,...,I$ can be estimated by solving (3.2.41) numerically. Conversely, in the constant rate case in which $\lambda_j \equiv \lambda,\ j=1,...,I$ are known, then the shape parameters $a_j,\ j=1,...,I$ can be estimated by solving (3.2.41) numerically.

3.3 Estimating hazard rates in each stage

As mentioned in Section 3.1, previous studies assumed either that an individual could not die but rather would eventually move to the next stage, or that individuals could die but that the hazard rate was the same for each stage. Hence, individual stages could be ignored and an overall hazard rate estimated. However, in many common situations such as disease progression and the life-cycle of organisms, the hazard rates in the stages are not all the same. Because of the nature of the estimation method, the estimators of the hazard rates and the estimators of the maturation rates are no longer independent. In order to estimate the hazard rate separately from estimating stage parameters, the multi-stage models in Section 2.1 need to have additional properties. Moreover, the estimation of the hazard rate in each stage requires more frequent sampling times. This is illustrated by simulations in Section 3.4.

3.3.1 Stage-wise constant hazard rates case

In order to estimate the maturation parameters in Section 3.2.1, the hazard rates were assumed to be known. This section proposes methods for estimating the hazard rate in each stage separately from estimating maturation parameters. Although this step will not be needed in a Bayesian approach, it will be used to obtain initial information regarding prior distributions of hazard rate variables (Section 4.3).

In an experiment described in [56], it was remarked that dead larvae of nematode are easily detected. The viability of motionless larvae was checked by touching with a fine needle. Thus, dead organisms, as well as living organisms, were able to be detected, and their stage identified.

Assume that $T_1, T_2, ..., T_K$ are independent and identically distributed random sampling times having an exponential distribution with density $s\exp(-st)$, $s > 0$. Let $D_j(T_k)$, $j = 1, ..., I$, denote the number of organisms that are observed to be dead in stage j at time T_k. Since the sampling is destructive, each of the observations $D := \{D_j(T_k),\ j = 1, ..., I\}$ is obtained from a separate population, $k = 1, ..., K$. Hence

$$N_k = \sum_{j=1}^{I+1} \left(N_j(T_k) + D_j(T_k)\right) \ . \tag{3.3.1}$$

Let $\tau_j = E(S_j)$, $j = 1, ..., I$ denote the mean duration time in stage j. Let

$$M_j(T_k) = N_k - \sum_{i=1}^{j-1} N_i(T_k) - \sum_{i=1}^{j-1} D_i(T_k) \tag{3.3.2}$$

be the number of individuals at time T_k, alive or dead, who have achieved at least stage j.

Because the hazard rate in each stage is assumed to be constant but varies from stage to stage, the failure time in each stage has an exponential distribution. Hence, the method of maximum likelihood (ML) can be applied to estimate the hazard rate. In order to estimate the hazard rate in each stage, the death times of individuals in each stage need to be known exactly. In a certain stage, the individuals either pass to the next stage or die in this stage. The number of individuals observed to be dead at time T_k is the cumulative number of individuals who died from the starting point to T_k. In a given stage, therefore,

we shall deem the difference between the number of dead organisms at the sampling time T_k and T_{k-1} to be the number of organisms which died exactly at sampling time T_k.

In stage 1, as a consequence of not knowing the exact number of dead individuals in each interval (T_{k-1}, T_k), we deem that $(D_1(T_k) - D_1(T_{k-1}))$ is the number of organisms that died in stage 1 at time T_k and hence the number of organisms moving to stage 2 at time T_k is defined as

$$\mathcal{M}_2(T_k) := \left(\sum_{i=2}^{I} N_i(T_k) + \sum_{i=2}^{I} D_i(T_k) - \sum_{i=2}^{I} N_i(T_{k-1}) - \sum_{i=2}^{I} D_i(T_{k-1}) \right) . \tag{3.3.3}$$

The likelihood function in stage 1 for assuming that $D_1(T_k), k = 1, ..., K$ individuals are observed to be dead is simply

$$\begin{aligned} L(D) &= \prod_{k=1}^{K} \left(\mu_1 \exp(-\mu_1 T_k)\right)^{(D_1(T_k) - D_1(T_{k-1}))} \left(\exp(-\mu_1 T_k)\right)^{\mathcal{M}_2(T_k)} , \\ &= \prod_{k=1}^{K} \left(\mu_1 \exp(-\mu_1 T_k)\right)^{(D_1(T_k) - D_1(T_{k-1}))} \times \\ &\quad \left(\exp(-\mu_1 T_k)\right)^{\left(\sum_{i=2}^{I} N_i(T_k) + \sum_{i=2}^{I} D_i(T_k) - \sum_{i=2}^{I} N_i(T_{k-1}) - \sum_{i=2}^{I} D_i(T_{k-1})\right)} , \end{aligned} \tag{3.3.4}$$

and, the log likelihood can be expressed by

$$\begin{aligned} l(D) &= \left[D_1(T_1) + \sum_{k=2}^{K} \left(D_1(T_k) - D_1(T_{k-1})\right) \right] \log(\mu_1) \\ &\quad - \mu_1 \left[D_1(T_1).T_1 + \sum_{k=2}^{K} \left(D_1(T_k) - D_1(T_{k-1})\right) T_k \right] - \mu_1 \left[\left(\sum_{i=2}^{I} N_i(T_k) + \sum_{i=2}^{I} D_i(T_k) \right) T_1 \right. \\ &\quad \left. + \sum_{k=2}^{K} \left(\sum_{i=2}^{I} N_i(T_k) + \sum_{i=2}^{I} D_i(T_k) - \sum_{i=2}^{I} N_i(T_{k-1}) - \sum_{i=2}^{I} D_i(T_{k-1}) \right) T_k \right] \\ &= \left[D_1(T_1) + \sum_{k=2}^{K} \left(D_1(T_k) - D_1(T_{k-1})\right) \right] \log(\mu_1) - \mu_1 A_1 - \mu_1 B_1 , \end{aligned} \tag{3.3.5}$$

where

$$\begin{aligned} A_1 &= D_1(T_1).T_1 + \sum_{k=2}^{K} \left(D_1(T_k) - D_1(T_{k-1})\right) T_k , \\ B_1 &= \left(\sum_{i=2}^{I} N_i(T_1) + \sum_{i=2}^{I} D_i(T_1) \right) T_1 \\ &\quad + \sum_{k=2}^{K} \left(\sum_{i=2}^{I} N_i(T_k) + \sum_{i=2}^{I} D_i(T_k) - \sum_{i=2}^{I} N_i(T_{k-1}) - \sum_{i=2}^{I} D_i(T_{k-1}) \right) T_k . \end{aligned} \tag{3.3.6}$$

The hazard rate in stage 1 is estimated by

$$\begin{aligned}\widehat{\mu}_1 &= \frac{D_1(T_1) + \sum_{k=2}^{K} \left(D_1(T_k) - D_1(T_{k-1})\right)}{A_1 + B_1} \\ &= \frac{D_1(T_K)}{A_1 + B_1}\,.\end{aligned} \tag{3.3.7}$$

Similarly in stage 2, using the estimated hazard rate in stage 1, the shape and rate parameters in stage 1 are estimated respectively via (3.2.8) or (3.2.9) as appropriate. Furthermore, we obtain $\widehat{\tau}_1$ the estimated mean duration in stage 1.

Because we assume that an individual must past from one stage to the next without missing a stage, an individual in stage 2 must pass from stage 1. Therefore, we shall deem the difference between the number of dead organisms at the sampling time T_k and T_{k-1} to be the number of dead organisms exactly at the sampling time $(T_k - \widehat{\tau}_1)$. We assume that $(D_2(T_k) - D_2(T_{k-1})M_2(T_k)/M_2(T_{k-1}))$ is the number of stage 2 organisms that died at time $(T_k - \widehat{\tau}_1)$ and

$$\mathcal{M}_3(T_k - \widehat{\tau}_1) := \left[\sum_{i=3}^{I} N_i(T_k) + \sum_{i=3}^{I} D_i(T_k) - \left(\sum_{i=3}^{I} N_i(T_{k-1}) + \sum_{i=3}^{I} D_i(T_{k-1})\right) M_2(T_k)/M_2(T_{k-1})\right] \tag{3.3.8}$$

is the number of stage 2 organisms moving to stage 3 at time $(T_k - \widehat{\tau}_1)$. Using the maximum likelihood method in a similar way that used to derive $\widehat{\mu}_1$, the hazard rate in stage 2 is estimated as

$$\widehat{\mu}_2 = \frac{D_2(T_1) + \sum_{k=2}^{K} \left(D_2(T_k) - D_2(T_{k-1})M_2(T_k)/M_2(T_{k-1})\right)}{A_2 + B_2}\,, \tag{3.3.9}$$

where

$$\begin{aligned} A_2 &= D_2(T_1)(T_1 - \widehat{\tau}_1) + \sum_{k=2}^{K} \left(D_2(T_k) - D_2(T_{k-1}).M_2(T_k)/M_2(T_{k-1})\right)(T_k - \widehat{\tau}_1)\,, \\ B_2 &= \left(\sum_{i=3}^{I} N_i(T_1) + \sum_{i=3}^{I} D_i(T_1)\right)(T_1 - \widehat{\tau}_1) + \sum_{k=2}^{K} \left[\sum_{i=3}^{I} N_i(T_k) + \sum_{i=3}^{I} D_i(T_k)\right. \\ &\quad \left. - \left(\sum_{i=3}^{I} N_i(T_{k-1}) + \sum_{i=3}^{I} D_i(T_{k-1})\right) M_2(T_k)/M_2(T_{k-1})\right](T_k - \widehat{\tau}_1)\,. \end{aligned} \tag{3.3.10}$$

By induction, estimators of the hazard rate in stage j, $j = 3, 4..., I$ are given by

$$\widehat{\mu}_j = \frac{D_j(T_1) + \sum_{k=2}^{K} \left(D_j(T_k) - D_j(T_{k-1}).M_j(T_k)/M_j(T_{k-1})\right)}{A_j + B_j}\,, j = 3, ..., I, \tag{3.3.11}$$

where

$$A_j = D_j(T_1).(T_1 - \sum_{i=1}^{j-1} \widehat{\tau}_i) + \sum_{k=2}^{K} \left(D_j(T_k) - D_j(T_{k-1}).M_j(T_k)/M_j(T_{k-1})\right) (T_k - \sum_{i=1}^{j-1} \widehat{\tau}_i) \,,$$

$$B_j = \left(\sum_{i=j+1}^{I} N_i(T_1) + \sum_{i=j+1}^{I} D_i(T_1) \right) (T_1 - \widehat{\tau}_1) + \sum_{k=2}^{K} \left[\sum_{i=j+1}^{I} N_i(T_k) + \sum_{i=j+1}^{I} D_i(T_k) - \right.$$
$$\left. - \left(\sum_{i=j+1}^{I} N_i(T_{k-1}) + \sum_{i=j+1}^{I} D_i(T_{k-1}) \right) M_j(T_k)/M_j(T_{k-1}) \right] (T_k - \widehat{\tau}_1) \,, \tag{3.3.12}$$

where $\widehat{\tau}_i$, $i = 1, ..., I$ is the estimated mean stage time in stage i, obtained from estimated stage maturation parameter in stage i.

In practice, the quantity $D_j(T_k) - D_j(T_{k-1})M_j(T_k)/M_j(T_{k-1})$ may be negative or undefined since destructive sampling requires different individuals to be counted at every stage. When this happens, the sampling time T_k may be simply ignored. This does not affect the estimates too much when the frequency of the sampling times are increased. In addition to this, some initial sampling times could be smaller than the sum of previous mean stage times. This makes the sampling time subtraction in stage j negative. This may be overcome by replacing the subtraction in stage j by the subtraction in the previous stage ($T_j - \sum_{i=1}^{j-1} \widehat{\tau}_i$ is replaced by $T_j - \sum_{i=1}^{j-2} \widehat{\tau}_i$) and repeating as necessary to arrive at a positive difference.

3.3.2 Linear time-dependent hazard rates

In order to estimate the linear time-dependent hazard rate in each stage separately from stage transition, the number of individuals dying in each stage and their death times are supposed to be known. Therefore, we develop a multi-stage model for sampling designs in which individuals are required to be identified. Hence, our model is extended as follows.

Single cohort stage-frequency data are collected in a group of identified individuals that have a life history consisting of $(I + 1)$ stages. An individual must pass from one stage to the next without missing a stage and destructive sampling is required in order to assess the stages reached as the system evolves. Because the individual's death time is known, covariates assessing the stages of each individual at their death time are recorded. A competing risk model is used to estimate the hazard rate in each stage. From the

estimated hazard rates in each stage, these estimated hazard rates are then approximated by linear functions of time, $\widehat{\mu}_j(t) = \widehat{\beta}_{j0} + \widehat{\beta}_{j1}t, j = 1, ..., I$. Estimating type-specific hazard functions through a competing risk model ([55]; [29]; [38]; [1]) is introduced briefly as follows.

Relative risk regression models are used for modelling type-specific hazard functions. Our interest is in identifying the stage at which an individual is going to die. Therefore, competing risks models may be applied to estimate hazard rate in each stage or the relationship between covariates and the hazard rates with some specific statistical methods. In our model, several failure types are distinguished by the death occurring in each stage. There are I stages and so there are I distinct failure types and I type-specific hazard functions

$$\mu_j[t, X(t)] = \lim_{h \to 0} h^{-1} P[t \leq T < t + h, J = j | T \geq t, X(t)] \, , \quad j = 1, ..., I \, , \tag{3.3.13}$$

where T is failure time, $X(t) = \{x(u), 0 \leq u < t\}$ where $x(u)$ is a vector of possible time-dependent covariates and J is the stage at which the event of interest occurs.

Because only one of the failure types can occur, the overall hazard function of the time to failure equals to sum of type-specific hazard functions

$$\mu(t, X) = \sum_{j=1}^{I+1} \mu_j(t, X) \, . \tag{3.3.14}$$

In Cox regression models ([1]), the type-specific hazard function at time t of stage j for a subject with covariate $Z(t)$ is modelled as

$$\mu_j(t, Z(t)) = \mu_{0j}(t) \exp[Z(t)' \beta_j] \, , \quad j = 1, ..., I \, . \tag{3.3.15}$$

Let $t_{j1} < ... < t_{jk_j}, j = 1, ..., I$ denote the k_j failure time of type j and Z_{ji} is the covariate of the individual that fails at t_{ji}. The corresponding partial likelihood is

$$L(\beta_1, \beta_2, ..., \beta_I) = \prod_{j=1}^{I} \prod_{i=1}^{k_j} \frac{\exp[Z_{ji}(t_{ji})' \beta_j)]}{\sum_{j \in R(t_{ji})} \exp[Z_l(t_{ji})' \beta_j]} \, , \tag{3.3.16}$$

where $R(t_{ji})$ is the risk set at time t_{ji}. The maximum partial likelihood estimator $(\widehat{\beta}_1, \widehat{\beta}_2, ..., \widehat{\beta}_I)$ can be obtained by simultaneously solving the equations

$$\partial log L(\beta_1, \beta_2, ..., \beta_I) / \partial \beta_{ji} = 0, j = 1, ..., I, i = 1, ..., K \, ,$$

where K is the number of elements of a covariate vector. The covariance matrix for this estimation is estimated as $I(\widehat{\beta})^{-1}$ where the observed information matrix is

$$I(\beta) = \left[-\frac{\partial^2}{\partial \beta j \partial \beta_i} \log L(\beta_1, \beta_2, ..., \beta_I) \right]_{i,j=1}^{I} . \tag{3.3.17}$$

The Breslow estimator ([1]) is applied to obtain an estimator for the type-specific cumulative baseline hazard $H_{0j}(t) = \int_0^t \lambda_{0j}(u)du$

$$\widehat{H}_{0j}(t) = \sum_{t_{ji}<t} \frac{1}{\sum_{j \in R(t_{ji})} \exp[Z_l(t_{ji})'\widehat{\beta}_j]} , \quad j = 1, ..., I. \tag{3.3.18}$$

The type-specific hazard functions are estimated by smoothing the Breslow estimator,

$$\widehat{\mu}_{0j}(t) = \frac{1}{b} \sum_{t_{ji}} K\left(\frac{t - t_{ji}}{b} \right) \triangle \widehat{H}_{0j}(t) , \quad j = 1, ..., I, \tag{3.3.19}$$

where b is a bandwidth, t_{ji} are times of events that occurred in interval $[t - b, t + b]$ and $K(x)$ is a bounded function that vanishes outside $[-1, 1]$ and has integral 1 (a weighted average of the $\triangle \widehat{H}_{0j}(t)$) and

$$\triangle \widehat{H}_{0j}(t) = \frac{1}{\sum_{j \in R(t_{ji})} \exp[Z_l(t_{ji})'\widehat{\beta}_j]} . \tag{3.3.20}$$

3.4 Simulation studies

The Laplace transform method for stage frequency data is efficient in the case of random sampling times. In practice, fixed sampling times dominate most experiments and stratified sampling may be used to increase the efficiency of the estimation as explained in [51]. Let $\mathcal{C} = \{c_0 = 0, c_1, ..., c_K = \infty\}$ be a set of cutpoints between K fixed sampling times $T_1, T_2, ..., T_K$.

Recalling (2.2.2), $T_1, T_2, ..., T_K$ are independent variables having density functions as

$$\alpha_j(t) = \begin{cases} \frac{\alpha(t)}{I_j} & t \in [c_j, c_{j+1}] \\ 0 & t \notin [c_j, c_{j+1}], \end{cases}$$

where $I_j = \int_{c_j}^{c_{j+1}} \alpha(t)dt = \int_{c_j}^{c_{j+1}} s \exp(-st)dt = \exp(-sc_j) - \exp(-sc_{j+1}) , \ j = 1, ..., K,$ where the dependence of I_j on s is suppressed.

An unbiased and strongly consistent estimator for $\int_0^\infty p_\theta(t)\alpha(t)dt$ is $\sum_{i=1}^K I_j n(T_j)$ (Section 2.2), where $I_j,\ j = 1, ..., K$, are called exponential weights.

In this section, datasets were simulated with known parameters. The method of Section 3.2 was used to estimate maturation parameters from the simulated data. In order to estimate maturation parameters in the constant shape case (Section 3.2), the estimation technique from Section 3.3 was applied to estimate hazard rate parameters. The estimation techniques in Sections 3.2 and 3.3 are evaluated by the accuracy of the estimations from simulated data.

3.4.1 Stage-wise constant hazard rates

The estimation of hazard rate in stage 1 is now obtained from empirical data. It should be noted that the inductive estimations of stage-wise hazard rates in Section 3.3.1 do not need the maturation parameters. Two simulations were conducted: one for the zero hazard rate case and another for the stage-wise constant hazard rate case. The former repeats the simulation of [25] [p. 19]. This was done to compare the structures of the model with zero hazard rate ([25]) and the model with stage-wise constant hazard rates. In particular, the model with stage-wise constant hazard rates required larger sample size and more frequent sampling than the model with zero hazard rate in order to estimate parameters.

Firstly, 10 stage times at each sampling time were generated from Erlangian distributions with constant shape $a = 2$ and rates $\lambda_1 = 1.5,\ \lambda_2 = 1.5$ and $\lambda_3 = 1.5$ for stage 1, stage 2 and stage 3, respectively. Fifteen sampling time points were taken between 0.1 and 6. In this simulation, the cutpoints between the 15 fixed sampling times were

$$\mathcal{C} = \{c_0 = 0, c_1 = \frac{T_1 + T_2}{2}, ..., c_{14} = \frac{T_{14} + T_{15}}{2}, c_{15} = \infty\}\,.$$

The numbers of observations in each stage appear in Table A.1, Appendix A. The unknown parameter s was determined using the iterative method in Section 2.2. We chose the initial random sampling rate s to be 0.5 for each of the three stages. The iterative scheme converged quickly and so the values of s obtained after three iterations were adopted as the final value in each stage. These were 0.74, 0.69 and 0.76 for the three stages

respectively. The number of steps was chosen as in [51] by inspection of the convergence. However, in practice there may not be an obvious method for choosing the number of steps. Some authors have suggested using certain optimality criteria for these choices ([21]; [36]) including minimum variance methods ([25]).

Means and standard deviations of the estimates of the stage parameters were taken from 100 simulations (Table 3.1). In this case the number of individuals in each sampling time was just 10, and we note that the standard deviations were not sufficiently similar to the corresponding results in the stage-wise constant hazard rate case reported in Table 3.2 where there were 1000 individuals in each sampling time. This is because of the effect of hazard rates in each stage of the model, the stage-wise constant hazard rate case requires an even larger sample size and more frequent sampling. The probabilities across time of a particular independent organism in each stage, conditional on the individuals starting at stage 1, are shown in Figure 3.2. This figure shows the empirical observations, the estimated probability curves and the true probability curves.

Table 3.1: Mean and standard deviation of estimated scale parameters $\lambda_i, i = 1, 2, 3$ in three stages in the constant shape rate case from 100 simulated data sets in which stage-specific mortality does not occur.

Parameter	λ_1	λ_2	λ_3
True values	1.5	1.5	1.5
Mean estimated values	1.51	1.51	1.59
Standard deviation	0.24	0.32	0.60

Secondly, data in the case of stage-wise constant hazard rates case were simulated. Compared with the simulation in [25], this simulation needs more sampling time points and stage times. Here, 1000 stage times at each sampling time were generated from Erlangian distributions with constant shape $a = 2$ and rates $\lambda_1 = 1.5$, $\lambda_2 = 1.5$ and $\lambda_3 = 1.5$ for stage 1, stage 2 and stage 3, respectively. Fifty sampling time points were taken between 0.1 and 6. This sampling time range extension was used to assess the hazard rate in each stage. From the initial number in each stage, the number of deaths was generated from exponential distributions with rates $\mu_1 = 0.3$, $\mu_2 = 0.5$ and $\mu_3 = 0.7$ in stage 1, stage 2

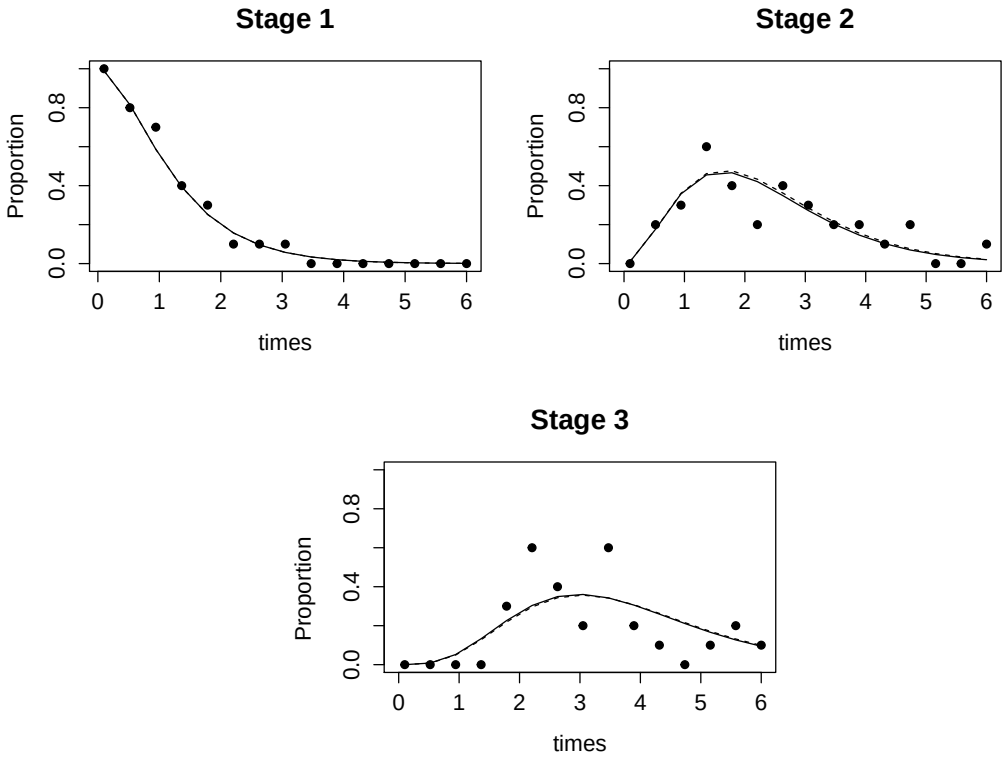

Figure 3.2: The empirical proportion (dotted line), the true probability curve (solid line) and the estimated probability curve (dashed line) of 10 sampled individuals at 15 sample times in the case that the mortality does not occur. The top left figure shows the curves in stage 1, the top right figure shows the curves in stage 2 and the bottom figure shows the curves in stage 3. The true probability curves and the estimated probability curves in the first stages are visually indistinguishable.

and stage 3 respectively. The number of individuals alive and the number of deaths in each stage were counted (Table A.2, Appendix A).

The unknown parameter s was derived using the iterative method described in Section 2.2. We chose an initial random sampling rate s to be 0.5 for each of the three stages. After two iterations, 0.74, 0.73 and 0.72 were determined as s values for stage 1, stage 2 and stage 3, respectively. Mean and standard deviation of the estimates of the stage parameters were taken from 100 simulations (Table 3.2). The true probability function and the estimated probability function in each stage are shown in Figure 3.3. These are visually indistinguishable between the true probability curves and the estimated probability curves in three stages. We used 1000 sampled individuals at 50 sample times in order to estimate hazard rates. With the large sample size, the approach produces good estimates of the

empirical proportion.

The simulations indicate that our methods can produce small biases for the fixed shape parameter in the stage-wise constant hazard rates case (Table 3.2). Because of a higher sampling rate in the stage-wise constant hazard rates case, the standard deviations of hazard rate parameters from 100 simulated data sets are small. Likewise, the standard deviations of rate parameters (Table 3.2) are similar to the case in which stage-specific mortality does not occur (Table 3.1). Overall, estimation errors from later stages were bigger due to additive estimation errors from previous stages.

Table 3.2: Mean and standard deviation of estimated scale and hazard rate parameters λ_i and $\mu_i, i = 1, 2, 3$ in three stages in the constant shape rate case from 100 simulated data sets in case of stage-wise constant hazard rates.

Parameter	λ_1	λ_2	λ_3	μ_1	μ_2	μ_3
True values	1.5	1.5	1.5	0.3	0.5	0.7
Mean estimated values	1.47	1.46	1.34	0.32	0.44	0.70
Standard deviation	0.21	0.53	1.40	0.03	0.04	0.11

3.4.2 Linear time-dependent death rates

A particularly easy case is one in which we combine stage-wise constant hazard rates in stage 1 and stage 2 and linear time-dependent hazard rate in stage 3. Data were simulated for organisms with three stages having Erlangian distributions with constant shape $a = 2$ and rates $\lambda_1 = 1.5$, $\lambda_2 = 1.5$ and $\lambda_3 = 1.5$ for stage 1, stage 2 and stage 3, respectively. The hazard rates in stage 1, 2 and 3 were set to $\mu_1 = 0$, $\mu_2 = 0.3$ and $\mu_3 = t$, respectively. We used the true values of the hazard rate in each stage as input to techniques in Section 3.2 for estimating the maturation parameters. The method used for estimating the maturation parameters for stages 1 and 2 was the same as in Section 3.4.1, but the method of Section 3.2.2 was used for stage 3.

Recalling (2.1.2), the probability of an organism being alive in stage j, for $j = 1, 2, 3$ at

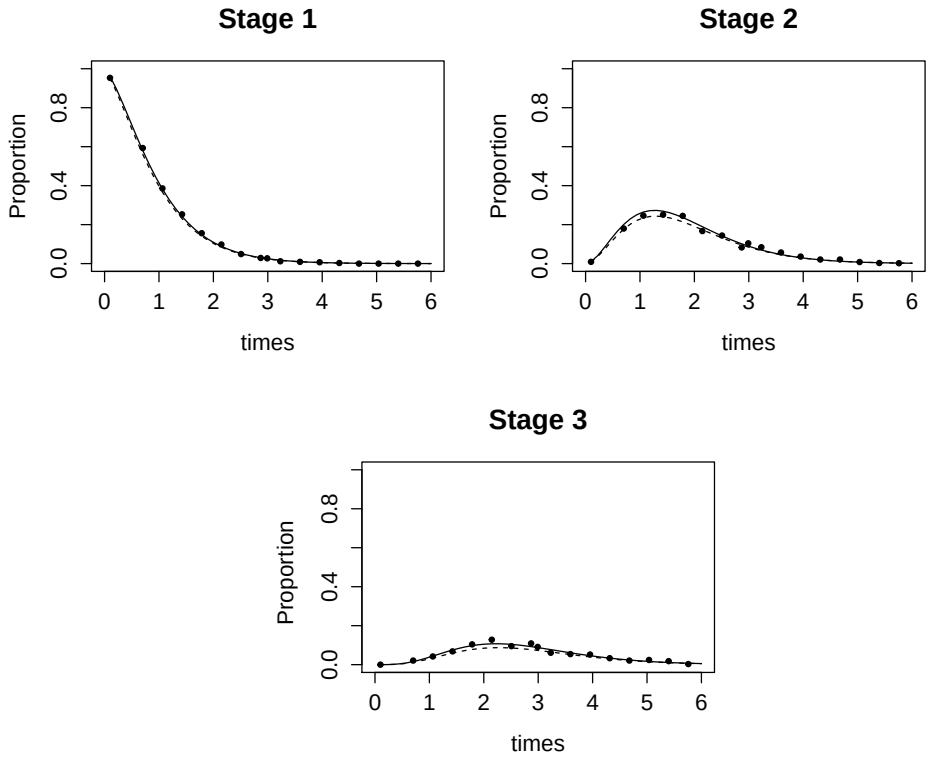

Figure 3.3: The empirical proportion (dotted line), the true probability curve (solid line) and the estimated probability curve (dashed line) of 1,000 sampled individuals at 50 sample times in stage-wise constant hazard rates case. The top left figure shows the curves in stage 1, the top right figure shows the curves in stage 2 and the bottom figure shows the curves in stage 3. The true probability curves and the estimated probability curves in the first stages are visually indistinguishable.

time t is calculated as

$$
\begin{aligned}
p_1(t) = H_1(t) &= \int_t^\infty g_1(x)dx = 1 - \int_0^t g_1(x)dx = 1 - \int_0^t \lambda_1^2 x \exp\left(-\lambda_1 x\right) dx \\
p_2(t) = h_1(t) * H_2(t) &= g_1(t) * \exp\left(-\mu_2 t\right) \int_t^\infty g_2(x)dx \\
p_3(t) = h_1(t) * h_2(t) * H_3(t) &= g_1(t) * \exp^{-\mu_2 t} g_2(t) * \exp\left(-\frac{t^2}{2}\right) \int_t^\infty g_3(x)dx \,.
\end{aligned}
\tag{3.4.1}
$$

The Laplace transforms $\psi_j(s)/s, j = 1, 2, 3$ in (3.2.21) become

$$
\begin{aligned}
\psi_1(s) &= s\int_0^\infty H_1(t)\exp(-st)\,dt = s\mathcal{L}(H_1(t)) = 1-\beta_1(s)\,,\\
\psi_2(s) &= s\int_0^\infty h_1 * H_2(t)\exp(-st)\,dt\\
&= s\mathcal{L}\{h_1\}(s)\mathcal{L}\{H_2(t)\}(s)\\
&= s\beta_1(s)\frac{1-\beta_2(s+\mu_2)}{s+\mu_2}\,,\\
\psi_3(s) &= s\int_0^\infty h_1 * h_2 * H_3(t)\exp(-st)\,dt\\
&\approx s\beta_1(s)\beta_2(s)\left\{\exp\left(\frac{s^2}{2}\right)\frac{\sqrt{\pi}}{2}\left[1-\operatorname{erf}\left(\sqrt{\frac{1}{2}}s\right)\right]-0.6\sqrt{\frac{\pi}{2\beta_{j1}}}\beta_3(s)\right\},
\end{aligned}
\tag{3.4.2}
$$

where

$$
\begin{aligned}
\beta_1(s) &= \mathcal{L}\{h_1\}(s) = \int_0^\infty g_1(x)\exp(-st)\,dx = \left(\frac{\lambda_1}{\lambda_1+s}\right)^a\\
\beta_2(s) &= \mathcal{L}\{h_2\})(s) = \int_0^\infty g_2(x)\exp(-s-\mu_2)t dx = \left(\frac{\lambda_2}{\lambda_2+\mu_2+s}\right)^a\\
\beta_3(s) &= \mathcal{L}\{h_3\}(s) = \int_0^\infty g_3(x)\exp\left(-st-\frac{t^2}{2}\right)dx\\
&= \lambda_3^a\exp\left(\frac{(\lambda_3+s)^2}{2}\right)\left[\Gamma\left(\frac{1}{2}(\lambda_3+s)^2,1\right)-\frac{(\lambda_3+s)}{\sqrt{2}}\Gamma\left(\frac{1}{2}(\lambda_3+s)^2,\frac{1}{2}\right)\right]
\end{aligned}
\tag{3.4.3}
$$

Using (3.2.14) and (3.2.30), the estimators of $\beta_j(s), j=1,2,3$ are given by

$$
\begin{aligned}
\widehat{\beta}_1(s) &= 1-\widehat{\psi}_1(s)\\
\widehat{\beta}_2(s) &= 1-\frac{s+\mu_2}{s}\frac{\widehat{\psi}_2(s)}{\widehat{\beta}_1(s)}\\
\widehat{\beta}_3(s) &\approx \frac{1}{0.6s}\sqrt{\frac{2}{\pi}}\left(s\exp\left(\frac{s^2}{2}\right)\frac{\sqrt{\pi}}{2}\left[1-\operatorname{erf}\left(\sqrt{\frac{1}{2}}s\right)\right]-\frac{\widehat{\psi}_3(s)}{\widehat{\beta}_1(s)\widehat{\beta}_2(s)}\right)\\
&= \frac{1}{0.6}\left(\exp\left(\frac{s^2}{2}\right)\left[1-\operatorname{erf}\left(\sqrt{\frac{1}{2}}s\right)\right]-\sqrt{\frac{2}{\pi}}\frac{\widehat{\psi}_3(s)}{s\widehat{\beta}_1(s)\widehat{\beta}_2(s)}\right)\,.
\end{aligned}
\tag{3.4.4}
$$

Following (3.2.8) and (3.2.37), the estimators for rate in each stage are given as

$$\begin{aligned}
\widehat{\lambda}_1(s) &= \frac{s}{[\widehat{\beta}_1(s)]^{-1/a} - 1}\,, \\
\widehat{\lambda}_2(s) &= \frac{s + \mu_2}{[\widehat{\beta}_2(s)]^{-1/a} - 1}\,, \\
\widehat{\beta}_3(s) &= \widehat{\lambda}_3^a \exp\left(\frac{(\lambda_3 + s)^2}{2}\right)\left[\Gamma\left(\frac{1}{2}(\widehat{\lambda}_3 + s)^2, 1\right) - \frac{(\widehat{\lambda}_3 + s)}{\sqrt{2}}\Gamma\left(\frac{1}{2}(\widehat{\lambda}_3 + s)^2, \frac{1}{2}\right)\right].
\end{aligned} \tag{3.4.5}$$

One hundred stage times at each sampling time were generated from Erlangian distributions with constant shape $a = 2$ and rates $\lambda_1 = 1.5$, $\lambda_2 = 1.5$ and $\lambda_3 = 1.5$ for stage 1, stage 2 and stage 3 respectively. Fifteen sampling time points were taken between 0.1 and 6. The number of observations and the number of deaths in each stage was recorded (Table A.3, Appendix A). The unknown variable s was determined by using the iterative method from Section 2.2. For each of the three stages, the initial sampling rate was set at 0.5. After five steps, s values were determined as 0.68, 0.65 and 0.78 for stage 1, stage 2 and stage 3 respectively. Mean and standard deviation of the estimated stage parameters were taken from 100 simulations (Table 3.3). The probabilities across time of a particular independent organism in each stage, conditional on starting at stage 1, are shown in Figure 3.4.

The estimated rate parameters, calculated using the assumptions of the estimated hazard rates and the fixed shape parameter, showed some bias (Table 3.3). The standard deviations from 100 simulations in the linear time-dependent hazard rates case are comparable to the standard deviations in the non-observed hazard rate case. However, there are 100 individuals at each sampling time, compared with only 10 individuals at each sampling time in the non-observed hazard rate case.

Table 3.3: Mean and standard deviation of estimated scale parameters $\lambda_i, i = 1, 2, 3$ in three stages in the constant shape rate case from 100 simulated data sets in the case of linear time-dependent hazard rates.

Parameter	λ_1	λ_2	λ_3
True values	1.5	1.5	1.5
Mean estimated values	1.51	1.57	1.60
Standard deviation	0.09	0.32	0.56

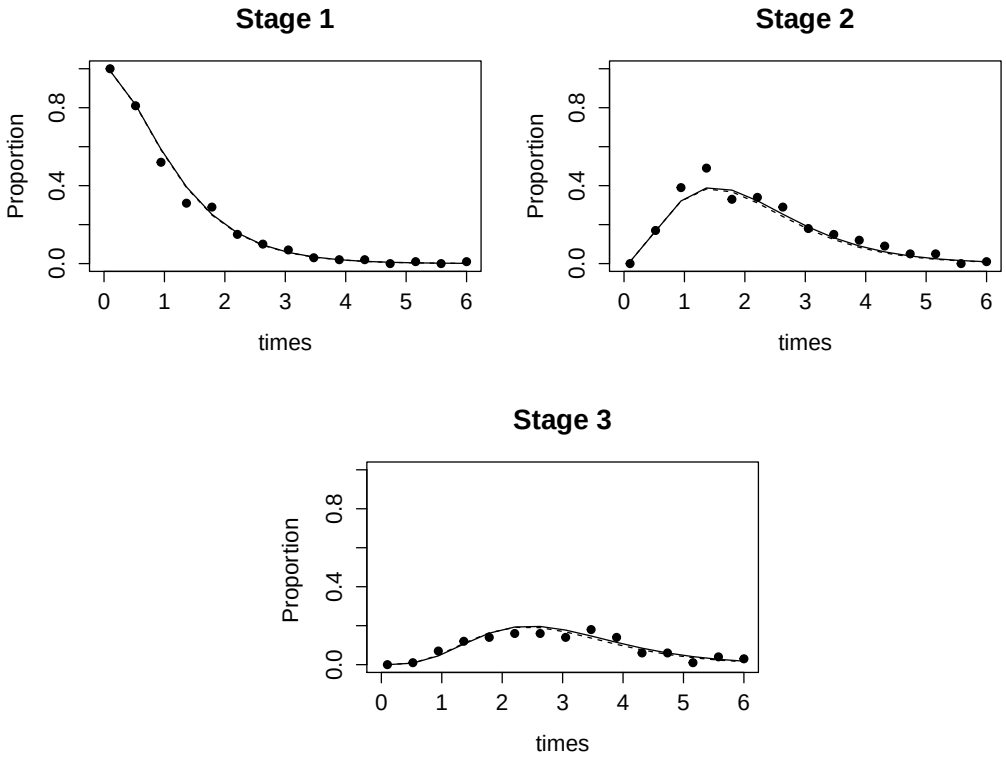

Figure 3.4: The empirical proportion (dotted line), the true probability curve (solid line) and the estimated probability curve (dashed line) of 100 sampled individuals at 15 sample times in the linear time-dependent hazard rates case. The top left figure shows the curves in stage 1, the top right figure shows the curves in stage 2 and the bottom figure shows the curves in stage 3. The true probability curves and the estimated probability curves in the first stages are visually indistinguishable.

3.5 Discussion

If, in a particular stage, the mean survival time is small compared to the mean maturation time (the sampling rate s^{-1}) the models are not acceptable. In such a stage, most of the individuals die before they move to the next stage. Therefore, the iterative method of [51] for choosing the sampling rate in Section 2.1 does not converge to the mean maturation time. In this case, the number of dead individuals in the stage is larger than the sum of the number of organisms in the next stages.

The models apply when assessing the stage reached by each individual through destructive sampling in situations where the hazard rates in each stage are non-trivial. The proposed

methods for estimating the hazard rates and the stage-dependent maturation parameters in each stage are extensions of previous stage-duration models of [51] and [25]. We believe that these methods could be useful in laboratory studies in which single cohort stage-frequency data are considered. The models (Section 3.2) apply when, at each sampling time, it is possible to assess the stage of living individuals and count the number of dead individuals.

The main contribution of this chapter is the exploration of the relationship between parameters in each stage that could be used in estimating parameters in the course of implementing Markov chain Monte Carlo (MCMC) methods. In this chapter, we estimate the maturation parameters and hazard rates at each stage under the assumption that the stage time durations are gamma distributed with fixed shape or rate parameters. We will use MCMC methods to overcome this limitation in the next chapters. There will be no need for any assumptions about the specific form of the stage-dependent maturation parameters. In particular, when the probability density functions of the time spent in each stage have gamma distribution, rate and shape parameters will be estimated at the same time. Thus, the rate parameters or shape parameters do not need to be assumed constant and known. Moreover, the survival time in each stage will not be limited to any parametric survival distribution and hence the hazard rate in each stage will not need to be a linear function of time t.

The estimation techniques of this chapter will be used as input to an MCMC algorithm in order to increase the convergence rate of MCMC methods. Running MCMC code to estimate the stage-dependent maturation parameters is non-trivial work, especially when the number of stages is large. Preliminary tests (reported in the next chapters) indicate that using the parameters estimated by the methods of this chapter will improve the convergence properties of the Markov process by applying the deterministic proposal of the MH algorithm. Thus, this chapter may be seen, in part, as an intermediate step toward applying MCMC methods. Once we are in a position to use MCMC methods, the full method may be applied to the data in [25]. At the same time, we are working to reduce the number of sampling times required in the data.

Chapter 4

Parameter Estimation in Multi-stage Models: A Bayesian Approach

Multi-stage time evolving models (stage-duration models) are common statistical models for biological systems, especially insect populations. In stage-duration distribution models, most approaches use Laplace transform method to estimate parameters ([51]; [25]). This method involves assumptions such as known constant shapes, known constant rates or the same overall hazard rate for all stages. These assumptions are strong, and restrictive. The main aim of this chapter is to weaken these assumptions by using a Bayesian approach. In particular, a Metropolis-Hastings (MH) algorithm based on deterministic transformations ([20]; [4]; [53]; [35]) is used to estimate parameters. We will use three models, one which has no hazard rates, the second has stage-wise constant hazard rates and the third has linear time-dependent hazard rates. These methods are validated in simulation studies followed by case studies of cattle parasites and breast development of New Zealander schoolgirls in Chapter 5.

4.1 Introduction

Recalling the models in Section 2.1, multi-stage time-evolving models are considered to be stage-duration models. The models are often used to model insect populations. An example of stage-structured data (Table 2.1) was an experiment analysing cattle parasitic nematode development ([58]). This parasite's life cycle has four stages including stage 1 (egg), stage 2 (first larvae stage), stage 3 (second larvae stage) and stage 4 (third larvae stage). The only information that can be collected is the number of organisms in each stage. This is gathered from destructive samples taken at different sampling times. That

is, an organism's stage cannot be observed without harvesting and destroying the samples and, thus, different samples are examined at each sampling time. The effect of mortality in each stage is also considered in the models. Such models were studied in [51], [39], [25], [17] and [34].

As discussed in Chapter 3, stage-dependent maturation parameters have often been estimated through Laplace transform (LT) methods ([51]; [25]). These estimations rely on assumptions that reduce their generality. A common limitation of these methods is that stage-dependent maturation parameters are estimated only in either the constant shape or the constant rate cases. This means that the stage-dependent maturation distribution is assumed, even though in general biological situations it is usually unknown. A second limitation relates to the way of handling the hazard rate at each stage. Although the estimated stage-specific mortalities were expanded to the situation where non-trivial death rates apply, hazard rates were estimated separately from the maturation parameters. Thus, many more samples were required than in the simpler models. A third limitation is that calculating variances of the estimates is complicated.

Many different approaches appear in the literature that address theoretical and statistical aspects of this problem (e.g., see [33],[18], [19], [32], [41] and [43]). A major difficulty is a lack of general computational methods to estimate the maturation parameters and stage specific mortality ([41]). Methods for statistical comparison of phenology between populations have been presented based on a t-test and simple linear regression. De Valpine et al ([19]) presented models for repeated censuses of cohort stage structure data without destruction. Bayesian approaches have been used in order to estimate stage duration parameters and mortality rates. Other approaches have been used to estimate separate distributions by combining stages instead of separating stage duration distributions ([32]; [43]). De Valpine and Knape ([18]) studied computational methods for smoothed maximum likelihood estimation applied to general multi-stage models from cohort data. The Markov chain Monte Carlo (MCMC) method was proposed as an extension to solve the statistical and computational issues mentioned above.

Knape and De valpine ([33]) recently proposed Monte Carlo estimation for models with no hazard rate and models with stage-wise hazard rate. The differences between Knape and De Valpine's approach and our approach are the probabilities ($p_j(t)$ and $d(t)$ (4.2.2, 4.3.2 and 4.3.3)) that an individual will be alive and will be found dead in stage $j, j = 1, 2, ..., I$

at time t in the likelihood function. In Knape and De Valpine's paper, the probability $p_j(t)$ ([33], (2), p. 996) was defined as a summation over all possible life schedules that are in stage j at time t weighted by the probability of being alive. The probability $d(t)$ ([33], (1), p. 996), that an individual is dead at time t, is one minus the probability of being alive in any stage at time t. In our approach, these probabilities $p_j(t)$ and $d_j(t)$ ((4.2.2), (4.3.2) and (4.3.3)) are defined as the convolutions of probability density functions of the time spent from stage 1 to stage j . The summary of differences between Knape and De Valpine's approach and our approach are as follows:

First, Knape and De Valpine proposed the time spent by individuals in each stage obtained by generating from a given parameter vector θ_0. The model assumption, that stage times and times of death of individuals are unknown, is not considered. In our model, the data are censored severely. The time of transition from one stage to the next stage and the times at which death occurs cannot be observed. Thus, the probabilities $p_j(t), j = 1, 2, ..., I$ do not depend on the time spent in each stage of individuals.

Second, the probability $p_j(t)$ depended on the time individuals spent in each stage and these times are assumed to be unknown in the model ([33], (2), p. 996). Therefore, the likelihood function was approximated computationally. The estimates of $\hat{p}_j(t)$ were not based on the transitions between stages of an individual. The estimates were based on the simulated values of the time spent in each stage by individuals which is generated from the distribution $h_j(t) = g_j(t)S_j(t), j = 1, 2, ..., I$. They used the particle MCMC method. In our approach, the calculation of the likelihood function is feasible in the no hazard rate case and the stage-wise hazard rate case (4.2.5 and 4.3.6). We apply the Metropolis-Hastings algorithm based on deterministic transformations to improve the mixing of Markov chains.

Third, in Knape and De Valpine's model, the stage at which death occurs was not considered. Hence, the probability, that an individual is dead at time t, did not include stages. In our model, at each sampling time, the number of dead organisms in each stage can be counted. The probability of an individual in stage j being found dead at sampling time t is defined by (4.3.3). This leads to the estimates of hazard rate parameters at each stage more precisely.

Parameter estimation within a Bayesian setting provides at least three advantages com-

pared to the existing methods in this context. First, it relaxes common assumptions such as known constant shapes or known constant rates or assuming the same overall death rate. This is achieved by allowing uncertainties in shape and rate parameters through their prior distributions. Second, a Bayesian approach allows the number of samples to be reduced because the estimates of maturation and hazard rate parameters are implemented simultaneously at each stage. Third, this approach provides information about uncertainties within each parameter.

In this chapter, we further explore a Bayesian approach. Parameters are estimated using some well-known MCMC algorithms such as the MH algorithm. The method does not need any assumptions about the specific form of the stage-dependent maturation parameters. When the probability density functions of the time spent in each stage has a gamma distribution, the rate and shape parameters can be estimated simultaneously. Thus, the rate or shape parameters do not need to be assumed as known constants. Furthermore, hazard rates in each stage can also be estimated simultaneously with the maturation parameters. This makes the Bayesian approach an effective estimation method for general multi-stage models from single cohort data. In addition, this approach allows the number of sampling times to be reduced compared to the studies in Chapter 3.

It should be noted that Markov chains from the MH algorithm have bad mixing tendencies, especially when the number of stages is large. Therefore, parameters estimated from the Laplace transform methods are embedded in the MH algorithm to increase the speed of the convergence of the chains. This method is described as the MH algorithm based on deterministic transformations ([20]; [4]).

This chapter is divided into four sections. Section 4.2 presents Bayesian analysis for estimating stage parameters in a model with no hazard rates. Section 4.3 deals with Bayesian analysis for estimating stage parameters in a model with stage-wise constant hazard rates. This is then followed by Section 4.4 which presents simulation studies in order to evaluate the methods in Sections 4.2 and 4.3, respectively. The application of the methodology in case studies is presented in Chapter 5.

4.2 Bayesian analysis for the model with no hazard rates

This section introduces the model with no hazard rates and the MH algorithm ([11]; [9]). The MH algorithm based on deterministic transformations is included in order to improve the convergence of the Markov process.

4.2.1 The likelihood function

Recall in Section 2.1 where stage-duration distribution models are considered as single cohort stage-frequency models. In these models, the life cycle of an individual is divided into $(I+1)$ stages, where stage $(I+1)$ is a final stage (for example, the death stage or the adult stage). We assume that at the first sampling time, all individuals start at stage 1. In the model with no hazard rates, an individual transforms through each stage without missing a stage and death does not occur in any stage. Moreover, in our model we examine destructive sampling, which assesses each individual's stage in the sample before destroying it. Thus, a different sample is taken at each sampling time.

In stage j, the stage duration density $g_j(t)$, $j = 1, ..., I$ has a gamma distribution which is parameterized by shape and rate parameters (a_j, λ_j). The number of organisms alive in stage j at time T_k is defined as $N_j(T_k) = N_{kj}$, $j = 1, ..., I$, $k = 1, ..., K$, where K is the number of sampling times. The number of sampled individuals at time t_k is defined as

$$N_k = \sum_{j=1}^{I+1} N_j(T_k) \, . \tag{4.2.1}$$

Recall (2.1.2), the probability of the life time of each independent organism is calculated by [47] as

$$\begin{aligned} p_j(t) &= \mathrm{P}(\text{organism } X \text{ is alive in stage } j \text{ at time } t | X \text{ starts stage 1 at time 0}) \\ &= h_1 * h_2 * ... * h_{j-1} * H_j(t) \, , \quad j = 1, ..., I \, , \end{aligned} \tag{4.2.2}$$

where $h_i(t) = g_i(t) = t^{a_j - 1} e^{-\lambda_j t} \lambda_j^{a_j} / \Gamma(a_j)$, $i = 1, ..., j-1$ is the density function of an organism being alive in stage i, the notation $*$ denotes convolution and

$$H_j(t) = \int_t^{\infty} g_j(x) dx \, . \tag{4.2.3}$$

Define $y_k = (N_1(T_k), N_2(T_k), ..., N_{I+1}(T_k))$ to be the observed information from k^{th} sampling time, and $y = \{y_k,\ k = 1, ..., K\}$ to be the sequence of these observations. Let $\theta = (a_1, \lambda_1, ..., a_I, \lambda_I) = (\theta_1, \theta_2, ..., \theta_{2I})\,,\ \ m = 1, ..., 2I \in \mathbb{R}^+$ represents the maturation parameters from I stages and $\theta_{(-m)} = (\theta_1, ..., , \theta_{m-1}, \theta_{m+1}, ..., \theta_{2I})$.

Following De Valpine and Knape 18, at each sampling time T_k, we assume that y_k has a multinomial distribution with parameters N_k as defined in (4.2.1) and $p_1(T_k), p_2(T_k), ..., p_I(T_k)$ as in (4.2.2), that is

$$(y_k|N_k, \theta) \sim \text{Multinomial}\left(N_k, p_1(T_k), p_2(T_k), ..., p_{I+1}(T_k)\right)\ , \tag{4.2.4}$$

where $p_{I+1}(T_k) = 1 - \sum_{j=1}^{I} p_j(T_k)$. Let $p\left(N_k, p_1(T_k), p_2(T_k), ..., p_{I+1}(T_k)\right)$ denote its probability mass function.

By (4.2.4), the joint likelihood function of observed information y from K sampling times is shown to be

$$f(y|\theta) = \prod_{k=1}^{K} p\left(N_k, p_1(T_k), p_2(T_k), ..., p_{I+1}(T_k)\right)\ . \tag{4.2.5}$$

4.2.2 The posterior distribution

The $2I$ maturation parameters are set as $\theta = (\theta_1, \theta_2, ..., \theta_{2I}) = (a_1, \lambda_1, ..., a_I, \lambda_I)$. Note that odd indexed θ_j's (where $j = 2k - 1$) coincide with the shape parameters a_k and the even indexed θ_j's (where $j = 2k$) coincide with the rate parameters λ_k. We set the prior distributions of each θ_j for $j = 1, ..., 2I$ to be uniform on the interval $\Omega_j = (0, x_j)$. The range of the uniform distributions is defined as

$$\Omega = \{(\theta_1, \theta_2, ..., \theta_{2I}) \in \Omega_1 \times \Omega_2 \times ... \times \Omega_{\theta_{2I}}\}\ . \tag{4.2.6}$$

The values of $x_j,\ j = 1, ..., 2I$, depend on the data but will not affect the algorithm (Section 4.2.4.2). If more information about the parameters is known, this will reduce the time to explore the entire sample space and keep a reasonable acceptance rate.

In our model, we assume that maturation parameters are independent for simplicity. This is a reasonable assumption for this model. We could also explore this issue by examining the relevant correlations in the MCMC output. When iterations are large enough, the sample variance matrix of the Markov chain can

be used to propose new value for parameter vector. Using (4.2.4) and (4.2.5), the target posterior distribution has the form of

$$\begin{aligned}
\pi(\theta|y) &\propto f(y|\theta)p(\theta) \\
&\propto f(y_1|\theta)...f(y_K|\theta)p(\theta_1)...p(\theta_{2I}) \\
&\propto \prod_{k=1}^{K} f(y_k|\theta) \prod_{m=1}^{2I} p(\theta_m) \\
&\propto \prod_{k=1}^{K} p\left(N_k, p_1(T_k), p_2(T_k), ..., p_{I+1}(T_k)\right) \prod_{m=1}^{2I} p(\theta_m) \\
&\propto \begin{cases} \prod_{i=1}^{I+1} \prod_{k=1}^{K} p_i^{N_{ki}}(T_k) & \theta \in \Omega \\ 0 & \theta \notin \Omega\,, \end{cases}
\end{aligned} \tag{4.2.7}$$

where $y = \{y_i,\ i = 1, ..., K\}$ and y_i is observed information from the i^{th} sampling time. Note that

$$\prod_{m=1}^{2I} p(\theta_m) = \prod_{m=1}^{2I} \frac{1}{x_m},\ \theta \in \Omega$$

is independent of data y and hence can be omitted in the above.

Using (4.2.7), the conditional posterior distributions of the rate and shape parameters in stage j for $j = 1, 2, ..., I$ are obtained as follows:

i) The conditional posterior distribution of the shape parameter a_j which coincides with θ_{2j-1} is

$$\pi\left(\theta_{2j-1}|y, \theta_{(-(2j-1))}\right) \propto \begin{cases} \prod_{j=1}^{I} \prod_{k=1}^{K} p_j^{N_{kj}}(T_k) & \theta_{2j-1} \in \Omega_{2j-1}\,, \\ 0 & \theta_{2j-1} \notin \Omega_{2j-1}\,. \end{cases} \tag{4.2.8}$$

ii) The conditional posterior distribution of the rate parameter λ_j which coincides with θ_{2j} is

$$\pi\left(\theta_{2j}|y, \theta_{(-2j)}\right) \propto \begin{cases} \prod_{j=1}^{I} \prod_{k=1}^{K} p_j^{N_{kj}}(T_k) & \theta_{2j} \in \Omega_{2j}\,, \\ 0 & \theta_{2j} \notin \Omega_{2j}\,. \end{cases} \tag{4.2.9}$$

4.2.3 The single MH algorithm for the no hazard rate model

We apply the single MH method ([11]; [9]) to the no hazard rate model. Although the algorithm (Algorithm 1) produces bad mixing tendencies (Chapter 5), the output of the

algorithm is used to estimate the initial sampling rates $s = (s_1, s_2, ..., s_I)$. The sampling rate s_j in stage j is selected as the mean of gamma density function $g_j(t)$ which is estimated from the single MH algorithm. These sampling rates will be updated after applying MH algorithm based on deterministic transformations. It is a reasonable choice of sampling rate s for the Laplace transform in using the iterative method of Schuh and Tweedie (Section 2.2).

We first introduce some notation. Let $\theta_m^{(t)}$, $m = 1, ..., 2I$ be a current value of the parameter θ_m and $\theta_m^{(*)}$, $m = 1, ..., 2I$ be a proposed value at the t^{th} iteration of the parameter θ_m through a random walk of MH algorithm.

We choose the prior distributions as uniform distribution with range Ω (4.2.6) and initialise value $\theta^{(0)} = \left(a_1^{(0)}, \lambda_1^{(0)}, ..., a_I^{(0)}, \lambda_I^{(0)}\right)$. The proposal distributions $\left(q\left(\theta_j^{(t)}|\theta_j^{(*)}\right)\right)$ are normal random walk $\left(N\left(\theta_j^{(t)}, \sigma_j\right)\right)$ in which the means are the current values. Tuning variances (σ_j) from the normal random walk distributions are optimized by using adaptive MH method (Section 2.3.5). The single MH algorithm is applied to our models as follows.

Algorithm 1 The single MH algorithm at stage j for the model with no hazard rate

1: Initialise $\left(a_j^{(0)}, \lambda_j^{(0)}, ..., a_I^{(0)}, \lambda_I^{(0)}\right)$
2: **for** $t = 1$ **to T do**
Update the shape parameter a_j which coincides with θ_{2j-1}:
3: Given the current state $\theta_{2j-1}^{(t)}$, propose $\theta_{2j-1}^{(*)} \sim q\left(\theta_{2j-1}^{(*)}|\theta_{2j-1}^{(t)}\right)$
4: Calculate the acceptance probability $\alpha_1 = \min\left(1, \frac{\pi\left(\theta_{2j-1}^{(*)}|y,\theta_{(-(2j-1))}\right)q\left(\theta_{2j-1}^{(t)}|\theta_{2j-1}^{(*)}\right)}{\pi\left(\theta_{2j-1}^{(t)}|y,\theta_{(-(2j-1))}\right)q\left(\theta_{2j-1}^{(*)}|\theta_{2j-1}^{(t)}\right)}\right)$
5: Set $\theta_{2j-1}^{(t+1)} = \theta_{2j-1}^{(*)}$ with probability α_1, otherwise set $\theta_{2j-1}^{(t+1)} = \theta_{2j-1}^{(t)}$
Update the rate parameter λ_j which coincides with θ_{2j}:
6: Given the current state $\theta_{2j}^{(t)}$, propose $\theta_{2j}^{(*)} \sim q\left(\theta_{2j}^{(*)}|\theta_{2j}^{(t)}\right)$
7: Calculate the acceptance probability $\alpha_2 = \min\left(1, \frac{\pi\left(\theta_{2j}^{(*)}|y,\theta_{(-2j)}\right)q\left(\theta_{2j}^{(t)}|\theta_{2j}^{(*)}\right)}{\pi\left(\theta_{2j}^{(t)}|y,\theta_{(-2j)}\right)q\left(\theta_{2j}^{(*)}|\theta_{2j}^{(t)}\right)}\right)$
8: Set $\theta_{2j}^{(t+1)} = \theta_{2j}^{(*)}$ with probability α_2, otherwise set $\theta_{2j}^{(t+1)} = \theta_{2j}^{(t)}$
9: **end for**

In order to calculate the acceptance probability in Algorithm 1, we need to calculate the probability $p_j(T_k)$, $j = 1, ..., I$, in stage j at each sampling time T_k. The Laplace transform of the probability $p_j(t)$ has an explicit form which was proved in Chapter 3. From (4.2.2),

the Laplace transform of the probability $p_j(t)$ is expressed as

$$\begin{aligned}
\mathcal{L}\left\{p_j(t)\right\}(s) &= \int_0^\infty p_j(t)\exp\left(-st\right)dt \\
&= \int_0^\infty h_1 * h_2 * ... * h_{j-1} * H_j(t)\exp\left(-st\right)dt \\
&= \mathcal{L}\left\{h_1\right\}(s)\mathcal{L}\left\{h_2\right\}(s)...\mathcal{L}\left\{h_{j-1}\right\}(s)\mathcal{L}\left\{H_j(t)\right\}(s) \\
&= \beta_1(s)\beta_2(s)...\beta_{j-1}(s)\frac{1-\beta_j(s)}{s}\,. \\
&= \left(\frac{\lambda_1}{\lambda_1+s}\right)^{a_1} .... \left(\frac{\lambda_{j-1}}{\lambda_{j-1}+s}\right)^{a_{j-1}} \frac{1-\beta_j(s)}{s}\,.
\end{aligned} \tag{4.2.10}$$

Taking the inverse Laplace transform of (4.2.10), we then obtain the value of the probability $p_j(t)$.

4.2.4 MH algorithm based on deterministic transformations for the no hazard rate model

In this section, in order to improve mixing of the MH method ([11]; [9]), we apply a more efficient algorithm, so-called MH based on deterministic transformations ([20]; [4]; [53]; [35]). The estimations of the maturation parameters in [51] and [25] are implemented in this algorithm. This implies that the rate parameter proposal in each stage is a deterministic proposal conditional on the proposed shape parameter in this stage and data.

4.2.4.1 Acceptance probability of shape and rate estimates in stage j

Recall that $\lambda_j^{(*)}$ is a deterministic proposal of λ_j. Given a constant $a_j^{(*)}$, $\mu_j = 0$ and the data y, $\lambda_j^{(*)}$ is estimated using (3.2.8) in the shape constant case as

$$\widehat{\lambda}_j^{(*)}(s) = \boldsymbol{f}(a_j^{(*)}) = \frac{s}{\left[\hat{\beta}_j(s)\right]^{-1/a_j^{(*)}} - 1}\,, \tag{4.2.11}$$

where $\beta_j(s) = \mathcal{L}(h_j) = \int_0^\infty g_j(t)\exp(-st)dt$ and the estimation of $\beta_j(s)$ was given in (3.2.7).

The proposed $a_j^{(*)}$ is a continuous random variable and the deterministic proposal $\hat{\lambda}_j^{(*)}$ is an invertible function of $a_j^{(*)}$. Using the change of variable technique, the probability density function of $\hat{\lambda}_j^{(*)}$ conditional on $a_j^{(*)}$ is

$$p\left(\hat{\lambda}_j^{(*)}|a_j^{(*)}\right) = p\left(a_j^{(*)}\right)\left|\left(\boldsymbol{f}^{-1}\right)'_{\hat{\lambda}_j^{(*)}|a_j^{(*)}}\right|, \tag{4.2.12}$$

where $\left(f^{-1}\right)'$ denotes the derivative of f^{-1}, and

$$a_j^{(*)}(s) = \boldsymbol{f}^{-1}\left(\hat{\lambda}^{(*)}\right) = \frac{\log\widehat{\beta}_j(s)}{\log\left[\hat{\lambda}_j^{(*)}/(\hat{\lambda}_j^{(*)}+s)\right]}. \tag{4.2.13}$$

Using (4.2.12), the joint probability density of the proposal parameters in each stage is calculated as

$$p\left(a_j^{(*)},\hat{\lambda}_j^{(*)}\right) = p(a_j^{(*)})p\left(\hat{\lambda}_j^{(*)}|a_j^{(*)}\right) = p\left(a_j^{(*)}\right)^2\left|\left(\boldsymbol{f}^{-1}\right)'_{\hat{\lambda}_j^{(*)}|a_j^{(*)}}\right|, j = 1,...,I\,. \tag{4.2.14}$$

Note that f is an isomorphism between a_j and λ_j, hence the entire sample space of λ_j is explored. Because the sampling distribution of a_j has fatter tails than the target, the sampling distribution of λ_j also covers the target. Using (4.2.5), (4.2.8) and (4.2.14) the acceptance probability for the shape and rate estimates is calculated as

$$\begin{aligned}
\alpha &= \min\left(1, \frac{\pi\left(a_j^{(*)},\hat{\lambda}_j^{(*)}|y\right)q\left(a_j^{(t)}|a_j^{(*)}\right)}{\pi\left(a_j^{(t)},\hat{\lambda}_j^{(t)}|y\right)q\left(a_j^{(*)}|a_j^{(t)}\right)}\right)\\
&= \min\left(1, \frac{f\left(y|a_j^{(*)},\hat{\lambda}_j^{(*)}\right)p\left(a_j^{(*)},\hat{\lambda}_j^{(*)}\right)q\left(a_j^{(\mathrm{t})}|a_j^{(*)}\right)}{f\left(y|a_j^{(t)},\hat{\lambda}_j^{(t)}\right)p\left(a_j^{(t)},\hat{\lambda}_j^{(t)}\right)q\left(a_j^{(*)}|a_j^{(t)}\right)}\right)\\
&= \min\left(1, \frac{f\left(y|a_j^{(*)},\hat{\lambda}_j^{(*)}\right)p(a_j^{(*)})^2\left|\left(\boldsymbol{f}^{-1}\right)'_{\hat{\lambda}_j^{(*)}|a_j^{(*)}}\right|q\left(a_j^{(\mathrm{t})}|a_j^{(*)}\right)}{f\left(y|a_j^{(t)},\hat{\lambda}_j^{(t)}\right)p(a_j^{(t)})^2\left|\left(\boldsymbol{f}^{-1}\right)'_{\hat{\lambda}_j^{(t)}|a_j^{(t)}}\right|q\left(a_j^{(*)}|a_j^{(t)}\right)}\right),
\end{aligned} \tag{4.2.15}$$

where the proposal distribution $q\left(a_j^{(\mathrm{t})}|a_j^{(*)}\right)$ is taken from the random walk MH.

4.2.4.2 The algorithm

In order to improve mixing from a single MH algorithm (Algorithm 1), we introduce the MH algorithm based on deterministic transformations (Algorithm 2, below), adapted

to suit our application. As maturation parameters in this model are independent and an individual must pass from one stage to the next stage without missing a stage, parameters at each stage are estimated separately. We start from stage 1 and move to the following stages in sequential order. The estimated parameters in previous stages are used as accepted values in later estimates.

In order to estimate the rate parameters in (4.2.11), we need to estimate the initial sampling rates ($s = (s_1, s_2, ..., s_I)$) across all stages. The initial sampling rate s_j, for $j = 1, 2, .., I$ is selected as the mean of the gamma density distribution $g_j(t)$ which is estimated using Algorithm 1.

The main loop of a MH algorithm based on deterministic transformations consists of four steps. The first step is generating a proposal sample $a_j^{(*)}$ from a proposal distribution $q\left(\theta_j^{(t)}|\theta_j^{(*)}\right)$. The proposal distribution is a symmetric Gaussian distribution $N\left(\theta_j^{(t)}, \sigma_{\theta_j}\right)$, where the tuning parameter σ_{θ_j} is generated independently of the parameter $\theta_j^{(t)}$. In the second step, given $a_j^{(*)}$ and s_j, the rate $\lambda_j^{(*)}$ is estimated using (4.2.11), recalling that the parameters $\hat{\beta}_j(s)$, $j = 1, 2, .., I$, were already estimated using (3.2.7). In the next step, we calculate the acceptance probability α using (4.2.15). Finally, we accept the proposal sample $(a_j^{(*)}, \lambda_j^{(*)})$ with the probability α. This extended algorithm is fully presented in Algorithm 2.

Algorithm 2 MH algorithm based on deterministic transformations at stage j for the model with no hazard rate

1: Estimate the initial sampling rates s using Algorithm 1
2: Initialise $\left(a_j^{(0)}, \lambda_j^{(0)}, ..., a_I^{(0)}, \lambda_I^{(0)}\right)$
3: **for** $t = 1$ **to T do**
Update the parameters a_j and λ_j which coincide with θ_{2j-1} and θ_{2j}, respectively:
4: Given the current state $\theta_{2j-1}^{(t)}$, propose $\theta_{2j-1}^{(*)} \sim q\left(\theta_{2j-1}^{(*)}|\theta_{2j-1}^{(t)}\right)$
5: Given $\theta_{2j-1}^{(*)}$ and $\theta_{2j-1}^{(t)}$, estimate $\theta_{2j}^{(*)}$ and $\theta_{2j}^{(t)}$ using (4.2.11)
6: Calculate the acceptance probability α using (4.2.15)
7: Set $\left(\theta_{2j-1}^{(t+1)}, \theta_{2j}^{(t+1)}\right) = \left(\theta_{2j-1}^{(*)}, \theta_{2j}^{(*)}\right)$ with the probability α,
otherwise set $\left(\theta_{2j-1}^{(t+1)}, \theta_{2j}^{(t+1)}\right) = \left(\theta_{2j-1}^{(t)}, \theta_{2j}^{(t)}\right)$
8: **end for**

4.2.4.3 Reversibility and stationarity of the Markov chain in stage i

The MH algorithm based on deterministic transformations creates an irreducible Markov chain. The Markov chain with transition matrix $P = \left[p\left(\theta^{(i)}, \theta^{(j)}\right)\right]_{ij=1}^{m}$ satisfies the balance property (Section 2.3.6)

$$\pi\left(\theta^{(i)}\right) p\left(\theta^{(i)}, \theta^{(j)}\right) = \pi\left(\theta^{(j)}\right) p\left(\theta^{(j)}, \theta^{(i)}\right)$$

for all pairs of states $\left(\theta^{(i)}, \theta^{(j)}\right)$. This is shown as follows:

$$\begin{aligned}
\pi\left(\theta^{(i)}\right) p\left(\theta^{(i)}, \theta^{(j)}\right) &= \pi\left(\theta^{(i)}\right) q\left(\theta^{(j)}|\theta^{(i)}\right) \min\{1, \frac{\pi\left(\theta^{(j)}\right) q\left(\theta^{(i)}|\theta^{(j)}\right)}{\pi\left(\theta^{(i)}\right) q\left(\theta^{(j)}|\theta^{(i)}\right)}\} \\
&= f\left(y|a_i, \lambda_i\right) p(a_i)^2 \left|\left(\boldsymbol{f}^{-1}\right)'_{\lambda_i|a_i}\right| q(a_j, a_i) \times \\
&\quad \min\left\{1, \frac{f\left(y|a_j, \lambda_j\right) p(a_j)^2 \left|\left(\boldsymbol{f}^{-1}\right)'_{\lambda_j|a_j}\right| q(a_i, a_j)}{f\left(y|a_i, \lambda_i\right) p(a_i)^2 \left|\left(\boldsymbol{f}^{-1}\right)'_{\lambda_i|a_i}\right| q(a_j, a_i)}\right\} \\
&= \min\left\{f\left(y|a_j, \lambda_j\right) p(a_j)^2 \left|\left(\boldsymbol{f}^{-1}\right)'_{\lambda_j|a_j}\right| q(a_i, a_j), \right. \\
&\quad \left. f\left(y|a_i, \lambda_i\right) p(a_i)^2 \left|\left(\boldsymbol{f}^{-1}\right)'_{\lambda_i|a_i}\right| q(a_j, a_i)\right\} \\
&= \pi\left(\theta^{(j)}\right) p\left(\theta^{(j)}, \theta^{(i)}\right) \quad \text{(symnetric in } i \text{and } j\text{)}.
\end{aligned} \tag{4.2.16}$$

Thus, the Markov chain is time-reversible and the solution π is the unique stationary distribution.

4.3 Bayesian analysis for the model with stage-wise constant hazard rates

4.3.1 The likelihood function

For the stage-wise constant hazard rates model, the lifetime in each stage j is influenced by the hazard rate μ_j and the time in stage j which has the density function $g_j(t)$ depending on the parameter θ_j. Recall the notation $N_j(T_k), j = 1, ..., I$ to be the number of organisms alive in stage j at time T_k and let $D_j(T_k),\ j = 1, ..., I,\ k = 1, ..., K$, be the number of

organisms that are observed to be dead in stage j at time T_k. Recall N_k, $k = 1, ..., K$, is the number of sampled individuals at time T_k, which in the case of stage-wise constant hazard rates is given by

$$N_k = \sum_{j=1}^{I+1} N_j(T_k) + \sum_{j=1}^{I+1} D_j(T_k) \,. \tag{4.3.1}$$

By (2.1.2), the probability of the life time of each particular independent organism is calculated as

$$\begin{aligned} p_j(t) &= P(\text{organism X is alive in stage } j \text{ at time } t | \text{X starts stage 1 at time 0}) \\ &= h_1 * h_2 * ... * h_{j-1} * H_j(t) \,, \end{aligned} \tag{4.3.2}$$

and

$$\begin{aligned} d_j(t) &= P(\text{organism X is found dead in stage } j \text{ at time } t | \text{X starts stage 1at time 0}) \\ &= h_1 * h_2 * ... * h_{j-1} * H_j^d(t) \,, \end{aligned} \tag{4.3.3}$$

where $h_i(t) = g_i(t)S_i(t) = g_i(t) \exp\left(-\int_0^t \mu_i dx\right), i = 1, ..., j-1$, is the density function of an organism being alive in stage i and the notation $*$ denotes convolution. Furthermore, $S_i(t)$ is the probability of an organism being alive in stage i longer than time t and μ_i is the hazard function of the survival time in stage i. The $H_j(t)$ and $H_j^d(t)$ in (4.3.2) and (4.3.3) respectively are expressed as

$$\begin{aligned} H_j(t) &= S_j(t) \int_t^\infty g_j(x)dx = \exp\left(-\int_0^t \mu_j dx\right) \int_t^\infty g_j(x)dx, \\ H_j^d(t) &\approx \mu_j \int_0^t g_j(x)dx \,. \end{aligned} \tag{4.3.4}$$

If the average of survival time ($E(T) = 1/\mu_j$) is smaller than the mean of maturation time in each stage, most of the individuals die before they move to the next stage. In order to avoid this, we consider the cases in which the hazard rate in each stage is smaller than one ($0 < \mu_j < 1$, $j = 1, ..., I$).

In the stage-wise constant hazard rate models, the observed information from the k^{th} sampling time becomes

$$y_k = (N_1(T_k), N_2(T_k), ..., N_I(T_k), D_1(T_k), D_2(T_k), ..., D_I(T_k), N_{I+1}(T_k) + D_{I+1}(T_k))$$

and the parameters from I stages have the form

$$\theta = (a_1, \lambda_1, \mu_1, ..., a_I, \lambda_I, \mu_I) = (\theta_1, \theta_2, ..., \theta_{3I}) \ .$$

The life-time in stage j is influenced by the survival time competing with the maturation time in this stage. The survival time and the maturation time in stage j are independent. In order to estimate these parameters, we assume that the numbers of alive and dead individuals have multinomial distributions conditioned on the total number as follows.

At each sampling time, we assume that the $y_k,\ k = 1, ..., K$, have multinomial distributions with parameters N_k as defined in (4.3.1), as well as $p_1(T_k), ..., p_I(T_k), d_1(T_k), ..., d_I(T_k)$ defined in (4.3.2) and (4.3.3). This later, natural, assumption has been used in [25] and [18],

$$(y_k|N_k, \theta) \sim \text{Multinomial}\left(N_k, p_1(T_k), p_2(T_k), ..., p_I(T_k), d_1(T_k), d_2(T_k), ..., d_I(T_k), p_{I+1}(T_k)\right), \tag{4.3.5}$$

where $p_{I+1}(T_k) = 1 - \sum_{i=1}^{I} \left(p_i(T_k) + d_i(T_k)\right)$.

By (4.3.5), the joint likelihood function from K sampling times is

$$f(y|\theta) = \prod_{k=1}^{K} p\left(N_k, p_1(T_k), p_2(T_k), ..., p_I(T_k), d_1(T_k), d_2(T_k), ..., d_I(T_k), p_{I+1}\right) . \tag{4.3.6}$$

In order to reduce the estimation error from hazard parameters when estimating the maturation parameters, at each sampling time, we shall require the probability $p^d(T_k)$ that an individual is dead at time T_k. Clearly

$$p^d(T_k) = 1 - \left(p_1(T_k) + p_2(T_k) + ... + p_I(T_k)\right) . \tag{4.3.7}$$

The error incurred here is due to the fact that we do not know the exact death time of an individual. The dead stages of individuals are ignored.

The random variables

$$N_1(T_k),\ N_2(T_k), ...,\ N_I(T_k),\ D(T_k) := D_1(T_k) + D(T_k) + ... + D_I(T_k) + N_{I+1}(T_k) + D_{I+1}(T_k)$$

are multinomially distributed with parameters N_k and $p_1(T_k), p_2(T_k), ..., p_I(T_k), p^d(T_k)$ as follows

$$\left(N_1(T_k), N_2(T_k), ..., N_I(T_k), D(T_k) \,|\, N_k\right) \sim \text{Multinomial}\left(N_k, p_1(T_k), p_2(T_k), ..., p_I(T_k), p^d(T_k)\right). \tag{4.3.8}$$

The likelihood function from K sampling times is

$$f(y|\theta) = \prod_{k=1}^{K} p\left(N_k, p_1(T_k), p_2(T_k), ..., p_I(T_k), p^d(T_k)\right) . \tag{4.3.9}$$

4.3.2 The posterior distribution

Similarly to Section 4.2.2, set the prior distributions of

$$\theta = (\theta_1, \theta_2, ..., \theta_{3I}) = (a_1, \lambda_1, \mu_1, ..., a_I, \lambda_I, \mu_I)$$

to have a uniform distribution with range $\Omega = (\Omega_1, \Omega_2, ..., \Omega_{3I})$.

$$\Omega = \{\theta = (\theta_1, \theta_2, ..., \theta_{3I}) \in \Omega_1 \times \Omega_2 \times ... \times \Omega_{3I}\} ,$$

where $\Omega_m = (0, x_m)$, $x_m \in \mathbb{R}^+, m = 1, ..., 3I$. Note that, for $j = 1, 2, ..., I$, the index j identifies the triple (a_j, λ_j, μ_j) and the corresponding triple $(\theta_{3j-2}, \theta_{3j-1}, \theta_{3j})$ in the $3I$-tuple θ.

In order to estimate the hazard parameters in each stage, using (4.3.6) the target posterior distribution has the form

$$\begin{aligned}
\pi(\theta|y) &\propto f(y|\theta)p(\theta) \\
&\propto \prod_{k=1}^{K} f\left(y_k|\theta\right) \prod_{m=1}^{3I} p(\theta_m) \\
&\propto \prod_{k=1}^{K} p\left(N_k, p_1(T_k), p_2(T_k), ..., p_I(T_k), d_1(T_k), d_2(T_k), ..., d_I(T_k), p_{I+1}(T_k)\right) \prod_{m=1}^{3I} p(\theta_m) \\
&\propto \begin{cases} \prod_{j=1}^{I} \prod_{k=1}^{K} p_j^{N_{kj}}(T_k) d_j^{D_{kj}}(T_k) \left(p_{I+1}(T_k)\right)^{N_{k,I+1}} & \theta \in \Omega , \\ 0 & \theta \notin \Omega , \end{cases}
\end{aligned} \tag{4.3.10}$$

where $N_{k,I+1} = N_k - \sum_{m=1}^{I} N_m(T_k) - \sum_{n=1}^{I} D_n(T_k)$. Note that

$$\prod_{m=1}^{3I} p(\theta_m) = \prod_{m=1}^{3I} \frac{1}{x_m}, \theta \in \Omega$$

is independent of data y and hence can be omitted in the above.

In order to estimate the maturation parameters in each stage, using (4.3.9) and (4.3.7), the target posterior distribution has the form

$$\begin{aligned}\pi(\theta|y) &\propto f(y|\theta)p(\theta)\\ &\propto \prod_{k=1}^{K} f(y_j|\theta) \prod_{m=1}^{3I} p(\theta_i)\\ &\propto \prod_{k=1}^{K} p\left(N_k, p_1(T_k), p_2(T_k), ..., p_I(T_k), p^d(T_k)\right) \prod_{m=1}^{3I} p(\theta_m)\\ &\propto \begin{cases} \prod_{j=1}^{I}\prod_{k=1}^{K} p_j^{N_{kj}}(T_k)\left(p^d(T_k)\right)^{\left(N_k-\sum_{j=1}^{I} N_j(T_k)\right)} & \theta \in \Omega\,,\\ 0 & \theta \notin \Omega\,. \end{cases}\end{aligned} \tag{4.3.11}$$

Using (4.3.10) and (4.3.11), the conditional posterior distributions of the rate, shape and hazard rate parameters in stage j for $j = 1, 2, ..., I$ are expressed as follows:

i) The conditional posterior distribution of the shape parameter a_j, which coincides with θ_{3j-2}, is

$$\begin{aligned}\pi\left(\theta_{3j-2}|y, \theta_{(-(3j-2))}\right) &\propto \prod_{k=1}^{K} f(y_j|\theta_{3j-2}).p(\theta_{3j-2})\\ &\propto \begin{cases} \prod_{j=i}^{I}\prod_{k=1}^{K} p_j^{N_{kj}}(T_k)\left(p^d(T_k)\right)^{\left(N_k-\sum_{l=1}^{I} N_l(T_k)\right)} & \theta_{3j-2} \in \Omega_{3j-2}\,,\\ 0 & \theta_{3j-2} \notin \Omega_{3j-2}\,. \end{cases}\end{aligned} \tag{4.3.12}$$

ii) The conditional posterior distribution of the rate parameter λ_j, which coincides with θ_{3j-1}, is

$$\begin{aligned}\pi\left(\theta_{3j-1}|y, \theta_{(-(3j-1))}\right) &\propto \prod_{k=1}^{K} f(y_j|\theta_{3j-1}).p(\theta_{3j-1})\\ &\propto \begin{cases} \prod_{j=i}^{I}\prod_{k=1}^{K} p_j^{N_{kj}}(T_k)\left(p^d(T_k)\right)^{\left(N_k-\sum_{l=1}^{I} N_l(T_k)\right)} & \theta_{3j-1} \in \Omega_{3j-1}\,,\\ 0 & \theta_{3j-1} \notin \Omega_{3j-1}\,. \end{cases}\end{aligned} \tag{4.3.13}$$

iii) The conditional posterior distribution of the hazard rate parameter μ_j, which coincides with θ_{3j}, is

$$\pi\left(\theta_{3j}|y,\theta_{(-3j))}\right) \propto \prod_{k=1}^{K} f(y_j|\theta_{3j}).p(\theta_{3j}) \\ \propto \begin{cases} \prod_{j=i}^{I}\prod_{k=1}^{K} p_j^{N_{kj}}(T_k)d_j^{D_{kj}}(t_k)\left(p_{I+1}(T_k)\right)^{N_{k,I+1}} & \theta_{3j}\in\Omega_{3j}\,, \\ 0 & \theta_{3j}\notin\Omega_{3j}\,. \end{cases} \tag{4.3.14}$$

4.3.3 The single MH algorithm for the stage-wise constant hazard rate model

Similarly to Section 4.2.3, we shall now introduce a single MH algorithm designed for the model with stage-wise constant hazard rates. The only difference between this algorithm and Algorithm 1 is that we need to update hazard rate parameters simultaneously with maturation parameters in each stage.

Choosing the prior distributions to be uniform distributions with range $\Omega = (\Omega_1 \times \Omega_2 \times ... \times\Omega_{3I})$ (4.2.6), initialise value $\theta^{(0)} = \left(\theta_1^{(0)},\theta_2^{(0)},...,\theta_{3I}^{(0)}\right) = \left(a_1^{(0)},\lambda_1^{(0)},\mu_1^{(0)},...,a_I^{(0)},\lambda_I^{(0)},\mu_I^{(0)}\right)$. The proposal distributions are normal random walks in which means are current values. Tuning variances from the normal random walk distributions are optimized by using an adaptive MH method (Section 2.3.5). The new single MH algorithm in the stage-wise constant hazard model is presented in Algorithm 3 below.

In order to calculate the acceptance probabilities in Algorithm 3, we should calculate the probabilities $p_j(t)$ and $d_j(t)$ in each stage at each sampling time. Using (3.2.4), the Laplace transform of the probabilities $p_j(t)$ and $d_j(t)$ in (4.3.2) and (4.3.3), can be expressed as

$$\begin{aligned} \mathcal{L}\left\{p_j(t)\right\}(s) &= \int_0^\infty p_j(t)e^{-st}dt \\ &= \int_0^\infty h_1 * h_2 * ... * h_{j-1} * H_j(t)\exp\left(-st\right)dt \\ &= \mathcal{L}\left\{h_1\right\}(s)\mathcal{L}\left\{h_2\right\}(s)...\mathcal{L}\left\{h_{j-1}\right\}(s)\mathcal{L}\left\{H_j(t)\right\}(s) \\ &= \beta_1(s)\beta_2(s)...\beta_{j-1}(s)\frac{1-\beta_j(s)}{s+\mu_j} \\ &= \left(\frac{\lambda_1}{\lambda_1+\mu_1+s}\right)^{a_1} ... \left(\frac{\lambda_{j-1}}{\lambda_{j-1}+\mu_{j-1}+s}\right)^{a_{j-1}} \frac{1-\beta_j(s)}{s+\mu_j} \end{aligned} \tag{4.3.15}$$

and

$$\begin{aligned}\mathcal{L}\left\{d_j(t)\right\}(s) &= \mathcal{L}\left\{h_1 * h_2 * ... * h_{j-1} * H_j^d(t)\right\}(s) \\ &= \mathcal{L}\left\{h_1\right\}(s)\mathcal{L}\left\{h_2\right\}(s)...\mathcal{L}\left\{h_{j-1}\right\}(s)\mathcal{L}\left\{H_j^d(t)\right\}(s) \\ &\approx \beta_1(s)\beta_2(s)...\beta_{j-1}(s)\mathcal{L}\left\{\mu_j \int_0^t g_j(x)dx\right\}(s) \\ &= \mu_j \left(\frac{\lambda_1}{\lambda_1+\mu_1+s}\right)^{a_1} .... \left(\frac{\lambda_{j-1}}{\lambda_{j-1}+\mu_{j-1}+s}\right)^{a_{j-1}} \frac{\left(\frac{\lambda_j}{\lambda_j+s}\right)^{a_j}}{s} .\end{aligned} \tag{4.3.16}$$

By taking the inverse Laplace transform of the (4.3.15) and (4.3.16), we obtain the probabilities $p_j(t)$ and $d_j(t)$ respectively.

Algorithm 3 The single MH algorithm at stage j for the model with stage-wise constant hazard rates

1: Initialise $\left(a_j^{(0)}, \lambda_j^{(0)}, \mu_j^{(0)}, ..., a_I^{(0)}, \lambda_I^{(0)}, \mu_I^{(0)}\right)$
2: **for** $t = 1$ **to T do**

Update the shape parameter a_j which coincides with θ_{3j-2}:

3: Given the current state $\theta_{3j-2}^{(t)}$, propose $\theta_{3j-2}^{(*)} \sim q\left(\theta_{3j-2}^{(*)}|\theta_{3j-2}^{(t)}\right)$
4: Calculate the acceptance probability $\alpha_1 = \min\left(1, \frac{\pi\left(\theta_{3j-2}^{(*)}|y,\theta_{(-(3j-2))}\right)q\left(\theta_{3j-2}^{(t)}|\theta_{3j-2}^{(*)}\right)}{\pi\left(\theta_{3j-2}^{(t)}|y,\theta_{(-(3j-2))}\right)q\left(\theta_{3j-2}^{(*)}|\theta_{3j-2}^{(t)}\right)}\right)$
5: Set $\theta_{3j-2}^{(t+1)} = \theta_{3j-2}^{(*)}$ with probability α_1, otherwise set $\theta_{3j-2}^{(t+1)} = \theta_{3j-2}^{(t)}$

Update the rate parameter λ_j which coincides with θ_{3j-1}:

6: Given the current state $\theta_{3j-1}^{(t)}$, propose $\theta_{3j-1}^{(*)} \sim q\left(\theta_{3j-1}^{(*)}|\theta_{3j-1}^{(t)}\right)$
7: Calculate the acceptance probability $\alpha_2 = \min\left(1, \frac{\pi\left(\theta_{3j-1}^{(*)}|y,\theta_{(-(3j-1))}\right)q\left(\theta_{3j-1}^{(t)}|\theta_{3j-1}^{(*)}\right)}{\pi\left(\theta_{3j-1}^{(t)}|y,\theta_{(-(3j-1))}\right)q\left(\theta_{3j-1}^{(*)}|\theta_{3j-1}^{(t)}\right)}\right)$
8: Set $\theta_{3j-1}^{(t+1)} = \theta_{3j-1}^{(*)}$ with probability α_2, otherwise set $\theta_{3j-1}^{(t+1)} = \theta_{3j-1}^{(t)}$

Update the hazard rate parameter μ_j which coincides with θ_{3j}:

9: Given the current state $\theta_{3j}^{(t)}$, propose $\theta_{3j}^{(*)} \sim q\left(\theta_{3j}^{(*)}|\theta_{3j}^{(t)}\right)$
10: Calculate the acceptance probability $\alpha_3 = \min\left(1, \frac{\pi\left(\theta_{3j}^{(*)}|y,\theta_{(-3j)}\right)q\left(\theta_{3j}^{(t)}|\theta_{3j}^{(*)}\right)}{\pi\left(\theta_{3j}^{(t)}|y,\theta_{(-3j)}\right)q\left(\theta_{3j}^{(*)}|\theta_{3j}^{(t)}\right)}\right)$
11: Set $\theta_{3j}^{(t+1)} = \theta_{3j}^{(*)}$ with probability α_3, otherwise set $\theta_{3j}^{(t+1)} = \theta_{3j}^{(t)}$
12: **end for**

4.3.4 MH algorithm based on deterministic transformations for the stage-wise constant hazard rate model

In this section, in order to improve the mixing of the chains in the above MH method (Algorithm 3), we introduce an MH algorithm based on deterministic transformations as in Section 4.2.4. The estimations of the maturation parameters in Chapter 3 are implemented for the algorithm. This implies that the rate parameter proposal in each stage is a deterministic proposal conditioned on the proposed shape parameter in this stage, given the observed data.

4.3.4.1 The acceptance probability

Recall that the $\lambda_j^{(*)}$ and $a_j^{(*)}$ are deterministic proposals for λ_j and a_j, respectively. In the constant shape case, $\lambda_j^{(*)}$ is estimated as

$$\hat{\lambda}_j^{(*)}(s) = \boldsymbol{f}\left(a_j^{(*)}\right) = \frac{s+\mu_j}{\left[\hat{\beta}_j(s)\right]^{-1/a_j^{(*)}} - 1}\,, \tag{4.3.17}$$

where $\beta_j(s) = \mathcal{L}(h_j) = \int_0^\infty g_j(t)e^{-(s+\mu_j)t}dt$. The estimation of $\beta_j(s)$ was presented in Chapter 3.

Using the change of variable technique, $a_i^{(*)}$ is a continuous random variable, $\hat{\lambda}_i^{(*)}$ is an invertible function of $a_i^{(*)}$. The probability density function of $\hat{\lambda}_i^{(*)}$ conditional on $a_i^{(*)}$ has the form

$$p\left(\hat{\lambda}_i^{(*)}|a_i^{(*)}\right) = p\left(a_i^{(*)}\right).\left|\left(\boldsymbol{f}^{-1}\right)'_{\hat{\lambda}_i^{(*)}|a_i^{(*)}}\right|\,, \tag{4.3.18}$$

where

$$a_j^{(*)}(s) = \boldsymbol{f}^{-1}\left(\hat{\lambda}_j^{(*)}\right) = \frac{log\widehat{\beta}_j(s)}{log\left[\hat{\lambda}_j^{(*)}/(\hat{\lambda}_j^{(*)}+\mu_j+s)\right]}\,. \tag{4.3.19}$$

Using (4.3.18) and (4.3.19), the joint probability density of parameters in each stage is calculated as

$$p\left(a_j^{(*)},\hat{\lambda}_j^{(*)}\right) = p\left(a_j^{(*)}\right)p\left(\hat{\lambda}_j^{(*)}|a_j^{(*)}\right) = p\left(a_j^{(*)}\right)^2\left|\left(\boldsymbol{f}^{-1}\right)'_{\hat{\lambda}_j^{(*)}|a_j^{(*)}}\right|\,. \tag{4.3.20}$$

Using (4.3.11),(4.3.12) and (4.3.20), the acceptance probability of the shape and rate estimates is calculated as

$$\alpha = \min\left(1, \frac{f\left(y|a_i^{(*)}, \hat{\lambda}_i^{(*)}\right) p(a_i^{(*)})^2 \left|\left(\boldsymbol{f}^{-1}\right)'_{\hat{\lambda}_i^{(*)}|a_i^{(*)}}\right| q\left(a_i^{(t)}|a_i^{(*)}\right)}{f\left(y|a_i^{(t)}, \hat{\lambda}_i^{(t)}\right) p(a_i^{(t)})^2 \left|\left(\boldsymbol{f}^{-1}\right)'_{\hat{\lambda}_i^{(t)}|a_i^{(t)}}\right| q\left(a_i^{(*)}|a_i^{(t)}\right)}\right). \tag{4.3.21}$$

4.3.4.2 The algorithm

As in Section 4.2.4.2, we now introduce a MH algorithm based on deterministic transformations for the stage-wise constant hazard model in order to improve mixing from a single MH algorithm (Algorithm 3). The latter is presented as, below, Algorithm 4.

We first estimate the initial sampling rate s_j in each stage j, using Algorithm 3. These sampling rates are updated using the iterative method of Schuh and Tweedie (Section 2.2). The main loop of the MH algorithm based on deterministic transformations in the stage-wise constant hazard model is the same as in the no hazard rate model, with the exception of updating the hazard rate parameters $\mu_j, j = 1, ..., I$, from Step 4 to Step 6.

Algorithm 4 The MH algorithm based on deterministic transformations at stage j for the model with stage-wise hazard rates

1: Estimate the initial sampling rates s using Algorithm 3
2: Initialise $\left(a_j^{(0)}, \lambda_j^{(0)}, \mu_j^{(0)}, ..., a_I^{(0)}, \lambda_I^{(0)}, \mu_I^{(0)}\right)$
3: **for** $t = 1$ **to T do**

Update the hazard rate parameter μ_j which coincides with θ_{3j}:

4: Given the current state $\theta_{3j}^{(t)}$, propose $\theta_{3j}^{(*)} \sim q\left(\theta_{3j}^{(*)}|\theta_{3j}^{(t)}\right)$
5: Calculate the acceptance probability $\alpha_1 = \min\left(1, \frac{\pi\left(\theta_{3j}^{(*)}|y,\theta_{(-3j)}\right) q\left(\theta_{3j}^{(t)}|\theta_{3j}^{(*)}\right)}{\pi\left(\theta_{3j}^{(t)}|y,\theta_{(-3j)}\right) q\left(\theta_{3j}^{(*)}|\theta_{3j}^{(t)}\right)}\right)$
6: Set $\theta_{3j}^{(t+1)} = \theta_{3j}^{(*)}$ with probability α_1, otherwise set $\theta_{3j}^{(t+1)} = \theta_{3j}^{(t)}$

Update the shape a_j and the rate λ_j parameters which coincides with θ_{3j-2} and θ_{3j-1}, respectively:

7: Given the current state $\theta_{3j-2}^{(t)}$, propose $\theta_{3j-2}^{(*)} \sim q\left(\theta_{3j-2}^{(*)}|\theta_{3j-2}^{(t)}\right)$
8: Given $\theta_{3j-2}^{(*)}$ and $\theta_{3j-2}^{(t)}$, estimate $\theta_{3j-1}^{(*)}$ and $\theta_{3j-1}^{(t)}$ using (4.3.17)
9: Calculate the acceptance probability α_2 using (4.3.21)
10: Set $\left(\theta_{3j-2}^{(t+1)}, \theta_{3j-1}^{(t+1)}\right) = \left(\theta_{3j-2}^{(*)}, \theta_{3j-1}^{(*)}\right)$ with probability α_2, otherwise set $\left(\theta_{3j-2}^{(t+1)}, \theta_{3j-1}^{(t+1)}\right) = \left(\theta_{3j-2}^{(t)}, \theta_{3j-1}^{(t)}\right)$
11: **end for**

4.4 Bayesian analysis for the model with linear time-dependent hazard rates

4.4.1 The likelihood function

The multi-stage model for this section is similar to that of Section 4.3.1 except that the hazard rate in each stage is linear time-dependent and has the form $\mu_j(t) = \gamma_j t, \gamma_j \geq 0, j = 1, ..., I$. The intercept in the hazards is set to zero because we assume that at time zero, all individuals start at stage 1. Set $\mathcal{T}_j$ be the survival time of an individual at stage j having density $f(\mathcal{T}_j)$.

Recall from (2.1.2), that the probability of the life time of each particular independent organism is calculated as

$$\begin{aligned} p_j(t) &= P(\text{an organism is alive in stage } j \text{ at time t} \mid \text{the organism starts stage 1 at time 0}) \\ &= h_1 * h_2 * ... * h_{j-1} * H_j(t) \,, \end{aligned} \tag{4.4.1}$$

and the probabilities of an individual in stage j being found dead at sampling time t are calculated as

$$\begin{aligned} d_j(t) &= P(\text{an organism is dead in stage } j \text{ at time } t \mid \text{the organism starts stage 1at time 0}) \\ &\approx h_1 * h_2 * ... * h_{j-1} * H_j^d(t) \,, \end{aligned} \tag{4.4.2}$$

where $h_i(t) = g_i(t)S_i(t) = g_i(t)\exp\left(-\int_0^t \mu_i(x)dx\right) = g_i(t)\exp\left(-\gamma_{i1}t^2/2\right), i = 1, ..., j-1$ is the density function of an organism being alive in stage i. Furthermore, $S_i(t)$ is probability of an organism being alive in stage i longer than time t and $\mu_i(t)$ is the hazard function of the survival time in stage i at time t, and

$$\begin{aligned} H_j(t) &= S_j(t)\int_t^\infty g_j(x)dx = \exp\left(-\int_0^t \mu_j(x)dx\right)\int_t^\infty g_j(x)dx \\ &= \exp\left(-\gamma_j\frac{t^2}{2}\right)\int_t^\infty g_j(x)dx \end{aligned} \tag{4.4.3}$$

$$H_j^d(t) \approx f\left(E(\mathcal{T}_j)\right)\int_0^t g_j(x)dx \,.$$

4.4.2 The posterior distribution

Set the $3I$ maturation parameters $\theta = (\theta_1, \theta_2, ..., \theta_{3I}) = (a_1, \lambda_1, \gamma_1, ..., a_I, \lambda_I, \gamma_I)$. The prior distributions of θ, as before, are chosen as uniformly distributed with range $\Omega = (\Omega_1 \times \Omega_2 \times ... \times \Omega_{3I})$.

In order to estimate the hazard parameters in each stage, the target posterior distribution has the form

$$\begin{aligned}
\pi(\theta|y) &\propto f(y|\theta)\, p(\theta) \\
&\propto \prod_{k=1}^{K} f(y_j|\theta) \prod_{m=1}^{3I} p(\theta_i) \\
&\propto \prod_{k=1}^{K} p\left(N_k, p_1(T_k), p_2(T_k), ..., p_I(T_k), d_1(T_k), d_2(T_k), ..., d_I(T_k), p_{I+1}(T_k)\right) \prod_{m=1}^{3I} p(\theta_i) \\
&\propto \begin{cases} \prod_{i=1}^{I} \prod_{k=1}^{K} p_i^{N_{ki}}(T_k) d_i^{D_{ki}}(T_k) \left(p_{I+1}(T_k)\right)^{\left(N_k - \sum_{j=1}^{I+1} N_j(T_k) - \sum_{j=1}^{I+1} D_j(T_k)\right)} & \theta \in \Omega \\ 0 & \theta \notin \Omega. \end{cases}
\end{aligned} \tag{4.4.4}$$

Similar to the stage-wise constant hazard rate case, in order to estimate the maturation parameters, the target posterior distribution in the linear time-dependent hazard rates model has the form

$$\begin{aligned}
\pi(\theta|y) &\propto f(y|\theta)\, p(\theta) \\
&\propto \prod_{k=1}^{K} f(y_k|\theta) \prod_{m=1}^{3I} p(\theta_i) \\
&\propto \prod_{k=1}^{K} p\left(N_k, p_1(T_k), p_2(T_k), ..., p_I(T_k), d_1(T_k), d_2(T_k), ..., d_I(T_k), p_{I+1}(T_k)\right) \prod_{m=1}^{3I} p(\theta_i) \\
&\propto \begin{cases} \prod_{j=1}^{I} \prod_{k=1}^{K} p_j^{N_{kj}}(T_k) d_j^{D_{kj}}(T_k) \left(p_{I+1}(T_k)\right)^{N_{k,I+1}} & \theta \in \Omega\,, \\ 0 & \theta \notin \Omega\,, \end{cases}
\end{aligned} \tag{4.4.5}$$

where $N_{k,I+1} = N_k - \sum_{m=1}^{I} N_m(T_k) - \sum_{n=1}^{I} D_n(T_k)$.

Using (4.4.5), the conditional probability density functions of the rate, shape and slope parameters in stage j for $j = 1, 2, ..., I$, are expressed as follows

i) The conditional probability density function of the shape parameter a_j, which coincides

with θ_{3j-2}, is

$$\pi\left(\theta_{3j-2}|y,\theta_{(-(3j-2))}\right) \propto \prod_{k=1}^{K} f\left(y_j|\theta_{3j-2}\right) p(\theta_{3j-2}) \propto \begin{cases} \prod_{j=i}^{I}\prod_{k=1}^{K} p_j^{N_{kj}}(T_k)\left(p^d(T_k)\right)^{\left(N_k-\sum_{l=1}^{I}N_l(T_k)\right)} & \theta_{3j-2}\in\Omega_{3j-2}\,, \\ 0 & \theta_{3j-2}\notin\Omega_{3j-2}\,. \end{cases} \tag{4.4.6}$$

ii) The conditional probability density function of the rate parameter λ_j, which coincides with θ_{3j-1}, is

$$\pi\left(\theta_{3j-1}|y,\theta_{(-(3j-1))}\right) \propto \prod_{k=1}^{K} f\left(y_j|\theta_{3j-1}\right) p(\theta_{3j-1}) \propto \begin{cases} \prod_{j=i}^{I}\prod_{k=1}^{K} p_j^{N_{kj}}(T_k)\left(p^d(T_k)\right)^{\left(N_k-\sum_{l=1}^{I}N_l(T_k)\right)} & \theta_{3j-1}\in\Omega_{3j-1}\,, \\ 0 & \theta_{3j-1}\notin\Omega_{3j-1}\,. \end{cases} \tag{4.4.7}$$

iii) The conditional probability density function of the slope parameter γ_j in the hazard rate stage j, which coincides with θ_{3j}, is

$$\pi\left(\theta_{3j}|y,\theta_{(-3j)}\right) \propto \prod_{k=1}^{K} f\left(y_j|\theta_{3j}\right) p(\theta_{3j}) \propto \begin{cases} \prod_{j=1}^{I}\prod_{k=1}^{K} p_j^{N_{kj}}(T_k) d_j^{D_{kj}}(T_k)\left(p_{I+1}(T_k)\right)^{N_{k,I+1}} & \theta_{3j}\in\Omega_{3j}\,, \\ 0 & \theta_{3j}\notin\Omega_{3j}\,, \end{cases} \tag{4.4.8}$$

where $p^d(T_k) = 1-(p_1(T_k)+p_2(T_k)+...+p_I(T_k))$ and $N_{k,I+1} = N_k - \sum_{m=1}^{I} N_m(T_k) - \sum_{n=1}^{I} D_n(T_k)$.

4.4.3 The single MH algorithm for the linear time-dependent hazard rates model

Similar to Section 4.4.3, the single MH algorithm is applied to our model with additional updates for slope parameters. The single MH algorithm for the linear time-dependent hazard rates model is presented as Algorithm 5, below.

Algorithm 5 The single MH algorithm at stage j for the model with time-dependent hazard rates

1: Initialise $\left(a_j^{(0)}, \lambda_j^{(0)}, \mu_j^{(0)}, ..., a_I^{(0)}, \lambda_I^{(0)}, \mu_I^{(0)}\right)$
2: **for** $t = 1$ **to T do**
Update the shape parameter a_j which coincides with θ_{3j-2}:
3: Given the current state $\theta_{3j-2}^{(t)}$, propose $\theta_{3j-2}^{(*)} \sim q\left(\theta_{3j-2}^{(*)}|\theta_{3j-2}^{(t)}\right)$
4: Calculate the acceptance probability $\alpha_1 = \min\left(1, \frac{\pi\left(\theta_{3j-2}^{(*)}|y,\theta_{(-(3j-2))}\right)q\left(\theta_{3j-2}^{(t)}|\theta_{3j-2}^{(*)}\right)}{\pi\left(\theta_{3j-2}^{(t)}|y,\theta_{(-(3j-2))}\right)q\left(\theta_{3j-2}^{(*)}|\theta_{3j-2}^{(t)}\right)}\right)$
5: Set $\theta_{3j-2}^{(t+1)} = \theta_{3j-2}^{(*)}$ with probability α_1, otherwise set $\theta_{3j-2}^{(t+1)} = \theta_{3j-2}^{(t)}$
Update the rate parameter λ_j which coincides with θ_{3j-1}:
6: Given the current state $\theta_{3j-1}^{(t)}$, propose $\theta_{3j-1}^{(*)} \sim q\left(\theta_{3j-1}^{(*)}|\theta_{3j-1}^{(t)}\right)$
7: Calculate the acceptance probability $\alpha_2 = \min\left(1, \frac{\pi\left(\theta_{3j-1}^{(*)}|y,\theta_{(-(3j-1))}\right)q\left(\theta_{3j-1}^{(t)}|\theta_{3j-1}^{(*)}\right)}{\pi\left(\theta_{3j-1}^{(t)}|y,\theta_{(-(3j-1))}\right)q\left(\theta_{3j-1}^{(*)}|\theta_{3j-1}^{(t)}\right)}\right)$
8: Set $\theta_{3j-1}^{(t+1)} = \theta_{3j-1}^{(*)}$ with probability α_2, otherwise set $\theta_{3j-1}^{(t+1)} = \theta_{3j-1}^{(t)}$
Update the slope parameter γ_j which coincides with θ_{3j}:
9: Given the current state $\theta_{3j}^{(t)}$, propose $\theta_{3j}^{(*)} \sim q\left(\theta_{3j}^{(*)}|\theta_{3j}^{(t)}\right)$
10: Calculate the acceptance probability $\alpha_3 = \min\left(1, \frac{\pi\left(\theta_{3j}^{(*)}|y,\theta_{(-3j)}\right)q\left(\theta_{3j}^{(t)}|\theta_{3j}^{(*)}\right)}{\pi\left(\theta_{3j}^{(t)}|y,\theta_{(-3j)}\right)q\left(\theta_{3j}^{(*)}|\theta_{3j}^{(t)}\right)}\right)$
11: Set $\theta_{3j}^{(t+1)} = \theta_{3j}^{(*)}$ with probability α_3, otherwise set $\theta_{3j}^{(t+1)} = \theta_{3j}^{(t)}$
12: **end for**

In order to calculate the acceptance probabilities in Algorithm 5, we should calculate the probability $p_j(t)$ and $d_j(t)$ in each stage, at each sampling time. The Laplace transform of the probabilities $p_j(t)$ and $d_j(t)$ in (4.4.1) and (4.4.2) can be expressed as

$$\begin{aligned}
\mathcal{L}\left\{p_j(t)\right\}(s) &= \int_0^\infty p_j(t)\exp\left(-st\right)dt \\
&= \int_0^\infty h_1 * h_2 * ... * h_{j-1} * H_j(t)\exp\left(-st\right)dt \\
&= \mathcal{L}\left\{h_1\right\}(s)\mathcal{L}\left\{h_2\right\}(s)...\mathcal{L}\left\{h_{j-1}\right\}(s)\mathcal{L}\left\{H_j\right\}(s) \\
&= \beta_1(s)\beta_2(s)...\beta_{j-1}(s)\mathcal{L}\left\{H_j(t)\right\}(s)
\end{aligned} \tag{4.4.9}$$

and

$$\begin{aligned}
\mathcal{L}\left\{d_j(t)\right\}(s) &= \mathcal{L}(h_1 * h_2 * ... * h_{j-1} * H_j^d(t)) \\
&= \mathcal{L}\left\{h_1\right\}(s)\mathcal{L}\left\{h_2\right\}(s)...\mathcal{L}\left\{h_{j-1}\right\}(s)\mathcal{L}\left\{H_j^d\right\}(s)
\end{aligned} \tag{4.4.10}$$

$$\approx \beta_1(s)\beta_2(s)...\beta_{j-1}(s)\mathcal{L}\left\{f(E(\mathcal{T}_j))\int_0^t g_j(x)dx\right\}(s)\,.$$

Recall from (3.2.41) in Chapter 3, we have

$$\begin{aligned}\beta_j(s) \approx \exp\left(\frac{(\lambda_j+s)^2}{2\gamma_{j1}}\right)\frac{\lambda_j^{a_j}}{\Gamma(a_j)}\left[\frac{1}{\gamma_{j1}}\left(\frac{2}{\gamma_{j1}}\right)^{(a_j-2)/2}\Gamma\left(\frac{\gamma_{j1}}{2}\left[\frac{(\lambda_j+s)}{\gamma_{j1}}\right]^2,\frac{a_j}{2}\right)\right.\\ \left.-\frac{1}{\gamma_{j1}}\left(\frac{2}{\gamma_{j1}}\right)^{(a_j-3)/2}(a_j-1)\frac{(\lambda_j+s)}{\gamma_{j1}}\Gamma\left(\frac{\gamma_{j1}}{2}\left[\frac{(\lambda_j+s)}{\gamma_{j1}}\right]^2,\frac{a_j-1}{2}\right)\right].\end{aligned} \tag{4.4.11}$$

Substituting $u_0 = (\lambda_j+s)^2/(2\gamma_{j1})$ in (4.4.11), we obtain

$$\begin{aligned}\beta_j(s) \approx \frac{\lambda_j^{a_j}\exp(u_0)}{\Gamma(a_j)}\left[\frac{1}{\gamma_{j1}}\left(\frac{2}{\gamma_{j1}}\right)^{(a_j-2)/2}\Gamma\left(u_0,\frac{a_j}{2}\right)\right.\\ \left.-\frac{1}{\gamma_{j1}}\left(\frac{2}{\gamma_{j1}}\right)^{(a_j-3)/2}(a_j-1)\frac{(\lambda_j+s)}{\gamma_{j1}}\Gamma\left(u_0,\frac{a_j-1}{2}\right)\right].\end{aligned} \tag{4.4.12}$$

In order to simplify the result in (4.4.12), we apply the following approximation of the incomplete gamma function ([2], p. 2)

$$\Gamma(u,a) \approx u^{a-1}e^{-u}\left(1+\frac{a-1}{u}\right). \tag{4.4.13}$$

Now, combinig (4.4.13) with (4.4.12) leads to:

$$\begin{aligned}\beta_j(s) &\approx \frac{\lambda_j^{a_j}\exp(u_0)}{\Gamma(a_j)}\left[\frac{1}{\gamma_{j1}}\left(\frac{2}{\gamma_{j1}}\right)^{(a_j-2)/2}\exp(-u_0)(u_0)^{(a_j-2)/2}\left(1+\frac{a_j-2}{2u_0}\right)\right.\\ &\left.-\frac{1}{\gamma_{j1}}\left(\frac{2}{\gamma_{j1}}\right)^{(a_j-3)/2}(a_j-1)\frac{(\lambda_j+s)}{\gamma_{j1}}\exp(-u_0)(u_0)^{(a_j-3)/2}\left(1+\frac{a_j-3}{2u_0}\right)\right]\\ &= \frac{\lambda_j^{a_j}}{\Gamma(a_j)}\left[\frac{1}{\gamma_{j1}}\left(\frac{2}{\gamma_{j1}}\right)^{(a_j-2)/2}(u_0)^{(a_j-2)/2}\left(1+\frac{a_j-2}{2u_0}\right)\right.\\ &\left.-\frac{1}{\gamma_{j1}}\left(\frac{2}{\gamma_{j1}}\right)^{(a_j-3)/2}(a_j-1)(u_0)^{(a_j-3)/2}\frac{(\lambda_j+s)}{\gamma_{j1}}\left(1+\frac{a_j-3}{2u_0}\right)\right]\\ &= \frac{\lambda_j^{a_j}}{\Gamma(a_j)}\frac{(\lambda_j+s)^{a_j-2}}{\gamma_{j1}^{a_j-1}}\left(2-a_j+\frac{-a_j^2+4a_j-5}{2u_0}\right).\end{aligned} \tag{4.4.14}$$

By (3.2.23), (4.4.3) and (4.4.14), we now have

$$\begin{aligned}\mathcal{L}\left\{H_j(t)\right\}(s) &= \mathcal{L}\left\{\exp\left(-\gamma_{j1}\frac{t^2}{2}\right)\int_t^\infty g_j(x)dx\right\}(s)\\ &\approx \left\{\exp\left(\frac{s^2}{2\gamma_{j1}}\right)\sqrt{\frac{\pi}{2\gamma_{j1}}}\left[1-\operatorname{erf}\left(\sqrt{\frac{\gamma_{j1}}{2}}\frac{s}{\gamma_{j1}}\right)\right]-0.6\sqrt{\frac{\pi}{2\gamma_{j1}}}\beta_j(s)\right\}\\ &\approx 0.6\sqrt{\frac{\pi}{2\gamma_{j1}}}\left(1-\beta_j(s)\right)\ .\end{aligned} \tag{4.4.15}$$

By (3.2.4), in which $\mu_i = 0$, and (4.4.3), we obtain

$$\begin{aligned}\mathcal{L}\left\{H_j^d(t)\right\}(s) &\approx \mathcal{L}\left\{f\left(E(\mathcal{T}_j)\right)\int_0^t g_j(x)dx\right\}(s)\\ &\approx f\left(E(\mathcal{T}_j)\right)\mathcal{L}\left\{\int_0^t g_j(x)dx\right\}(s)\\ &\approx \frac{f\left(E(\mathcal{T}_j)\right)}{s}\mathcal{L}\left\{g_j(t)\right\}(s)\\ &\approx \frac{f\left(E(\mathcal{T}_j)\right)}{s}\left(\frac{\lambda_j}{s+\lambda_j}\right)^{a_j}\ .\end{aligned} \tag{4.4.16}$$

Taking the inverse Laplace transform of (4.4.9) and (4.4.10), we obtain the value of the probabilities $p_j(t)$ and $d_j(t)$.

Because of the complex form of (4.4.14), the inverse functions $\left(f^{-1}\right)'_{\lambda_j^{(*)}|a_j^{(*)}}$, $j = 1, 2, ..., I$ in the acceptance probabilities are difficult to calculate. We cannot apply the MH algorithm based on deterministic transformations in the linear time-dependent hazard rate case as in Section 4.3.4. Moreover, the approximations in (4.4.15) and (4.4.16) produce biases as the number of stages having linear time-dependent hazard rates increases. We also note that estimates will have bias when the average survival times are larger than the average maturation time at each stage. Thus, in order to reduce biases, number of stages having time-dependent hazard rate should be small. A combination of no hazard rate, stage-wise hazard rate and time-dependent hazard rate on simulated data is implemented in Section 5.3.

Chapter 5

Simulation Studies

We applied the MH algorithm based on deterministic transformations as explained in Sections 4.2, 4.3 and 4.4 in order to estimate the θ parameters. We updated each stage separately, starting from stage 1 and moving consecutively to the following stages. The initial values for the parameters were selected randomly. Tuning variances from the normal random walk distributions were optimized using an adaptive MH method (Section 2.3.5).

When we applied MH Algorithms 1 and 3, we ran $100,000$ iterations. The chains from MCMC output showed slow mixing and auto-correlation. By appling the MH Algorithms 2, 4 and 5 based on deterministic transformations, we ran $10,000$ iterations. The chains from MCMC output converged well demonstrated in Gelman and Rubin and Geweke diagnostics. The sampling rates s_j were chosen as the estimated mean of the time spent in each stage j. The estimated mean duration of a stage was taken from the summary output from a single MH algorithm with $T = 100,000$ iterations in which the first $95,000$ iterations were discarded as burn-in. Next, the MH algorithms based on the deterministic transformations were applied. In order to assess the convergence of the MCMC chains, we used the Gelman and Rubin multiple sequence and the Geweke diagnostic tests, respectively ([45]). We ran five MCMC chains at different starting points, of length $T = 10,000$ with the first $5,000$ iterations discarded as burn-in.

5.1 Simulation data in the no hazard rate model

We applied the MCMC methods as explained in Section 4.2 to simulated data. Data were simulated using methods analogous to those used in [25]. At each sampling time, ten individuals were generated from Erlangian distributions with shapes $a_j = 2,\ j = 1, 2, 3$ and rates $\lambda_j = 1.5,\ j = 1, 2, 3$ for stage 1, stage 2 and stage 3, respectively. Fifteen

sampling time points were taken between 0.1 and 6. In this simulation, the cut-points between the 15 fixed sampling times were

$$S = \{c_0 = 0, c_1 = \frac{T_1 + T_2}{2}, ..., c_{14} = \frac{T_{14} + T_{15}}{2}, c_{15} = \infty\}\,.$$

5.1.1 The probabilities at each stage

This section presents the acceptance probability calculations at each stage as described in Section 4.2.4.1. From (4.2.2), the probabilities of an individual being alive in each stage at each sampling time were calculated as follows for given maturation parameters over three stages $(a_1, \lambda_1, \mu_1, a_2, \lambda_2, \mu_2, a_3, \lambda_3, \mu_3)$.

- The probability $p_1(T_k)$ of being alive in stage 1 at sampling time T_k was obtained as

$$\begin{aligned} p_1(T_k) &= H_1(T_k) \\ &= \int_{t_k}^{\infty} g_1(x)dx = \int_{T_k}^{\infty} \frac{x^{a_1-1}\exp(-\lambda_1 x)\,\lambda_1^{a_1}}{(a_1-1)!}dx \\ &= \frac{\lambda_1^{a_1}}{(a_1-1)!}\int_{T_k}^{\infty} x^{a_1-1}\exp(-\lambda_1 x)\,dx\,, \end{aligned} \tag{5.1.1}$$

and the Laplace transform $\mathcal{L}\{p_1(t)\}(s)$ was obtained as

$$\begin{aligned} \mathcal{L}\{p_1(t)\}(s) &= \mathcal{L}\{H_1(t)\}(s) \\ &= \frac{1-\mathcal{L}\{h_1\}(s)}{s} = \frac{1}{s}\left(1-\left(\frac{\lambda_1}{\lambda_1+s}\right)^{a_1}\right)\,. \end{aligned} \tag{5.1.2}$$

- The probability $p_2(T_k)$ of being alive in stage 2 at sampling time T_k and the Laplace transform $\mathcal{L}(p_2(t))$ were obtained as

$$\begin{aligned} p_2(T_k) &= h_1 * H_2(T_k)\,, \\ \mathcal{L}\{p_2(t)\}(s) &= \mathcal{L}\{h_1\}(s)\mathcal{L}\{H_2(t)\}(s) \\ &= \frac{1}{s}\left(\frac{\lambda_1}{\lambda_1+s}\right)^{a_1}\left(1-\left(\frac{\lambda_1}{\lambda_2+s}\right)^{a_2}\right)\,. \end{aligned} \tag{5.1.3}$$

- The probability $p_3(T_k)$ of being alive in stage 3 at sampling time T_k and the Laplace

transform $\mathcal{L}(p_3(t))$ were obtained as

$$\begin{aligned} p_3(T_k) &= h_1 * h_2 * H_3(T_k)\,, \\ \mathcal{L}\left\{p_3(t)\right\}(s) &= \mathcal{L}\left\{h_1\right\}(s)\mathcal{L}\left\{h_2\right\}(s)\mathcal{L}\left\{H_3(t)\right\}(s) \\ &= \frac{1}{s}\left(\frac{\lambda_1}{\lambda_1+s}\right)^{a_1}\left(\frac{\lambda_2}{\lambda_2+s}\right)^{a_2}\left(1-\left(\frac{\lambda_3}{\lambda_3+s}\right)^{a_3}\right)\,. \end{aligned} \tag{5.1.4}$$

Taking the inverse Laplace transform of (5.1.2),(5.1.3) and (5.1.4) numerically, we obtain the values of the probabilities $p_1(t_k)$, $p_2(t_k)$ and $p_3(t_k)$, respectively. These probabilities were substituted in (4.2.7) in order to calculate the posterior distributions needed to evaluate acceptance probabilities α_1 and α_2 in Algorithm 1.

5.1.2 The single MH algorithm

We applied Algorithm 1 in Section 4.2.3 to estimate the maturation parameters and hazard rate in each stage. For each of three stages, the starting values were set at 0.1 and the initial tuning parameters for proposal values of maturation parameters were set at 0.1. The optimal tuning parameters for proposal values were determined with the adaptive MH method (Section 2.3). A summary of the maturation parameters from 100,000 MH iterations with 95,000 samples discarded as burn-in is shown in Table 5.1. The means of the maturation parameters in the three stages converged slowly to the true values of these parameters. The autocorrelation plots for our MH samples in Figure 5.1 show very slow mixing as indicated by slowly decaying dependence as lag increases. Figures 5.2 and 5.3 show the trace plots confirming the mixing and distributions of the maturation parameter samples in three stages. In practice, our MH Algorithm 1 was run several times from different starting points. The results from these runs were compared and all agreed with each other.

The results indicate that the chains have failed to converge. Therefore, we applied Algorithm 2 based on deterministic transformations, as described in Section 4.2.4, in the next section in order to improve the convergence rates of the chains. The output from Table 5.1 is used to calculate the sampling rate s in the next section.

Table 5.1: Summary results of parameter estimation resulting from applying Algorithm 1 based on the last 5,000 iterations for the no hazard rate model.

Parameter	True value	Acceptance rate %	Mean	SD	2.5%	50%	97.5%
a_1	2	53.48	2.043	0.208	1.637	2.043	2.418
a_2	2	59.39	2.539	0.504	1.604	2.552	3.632
a_3	2	66.42	1.658	0.381	1.008	1.627	2.490
λ_1	1.5	44.67	1.551	0.152	1.268	1.549	1.869
λ_2	1.5	49.97	1.979	0.391	1.264	1.974	2.821
λ_3	1.5	59.56	1.229	0.304	0.713	1.197	1.902

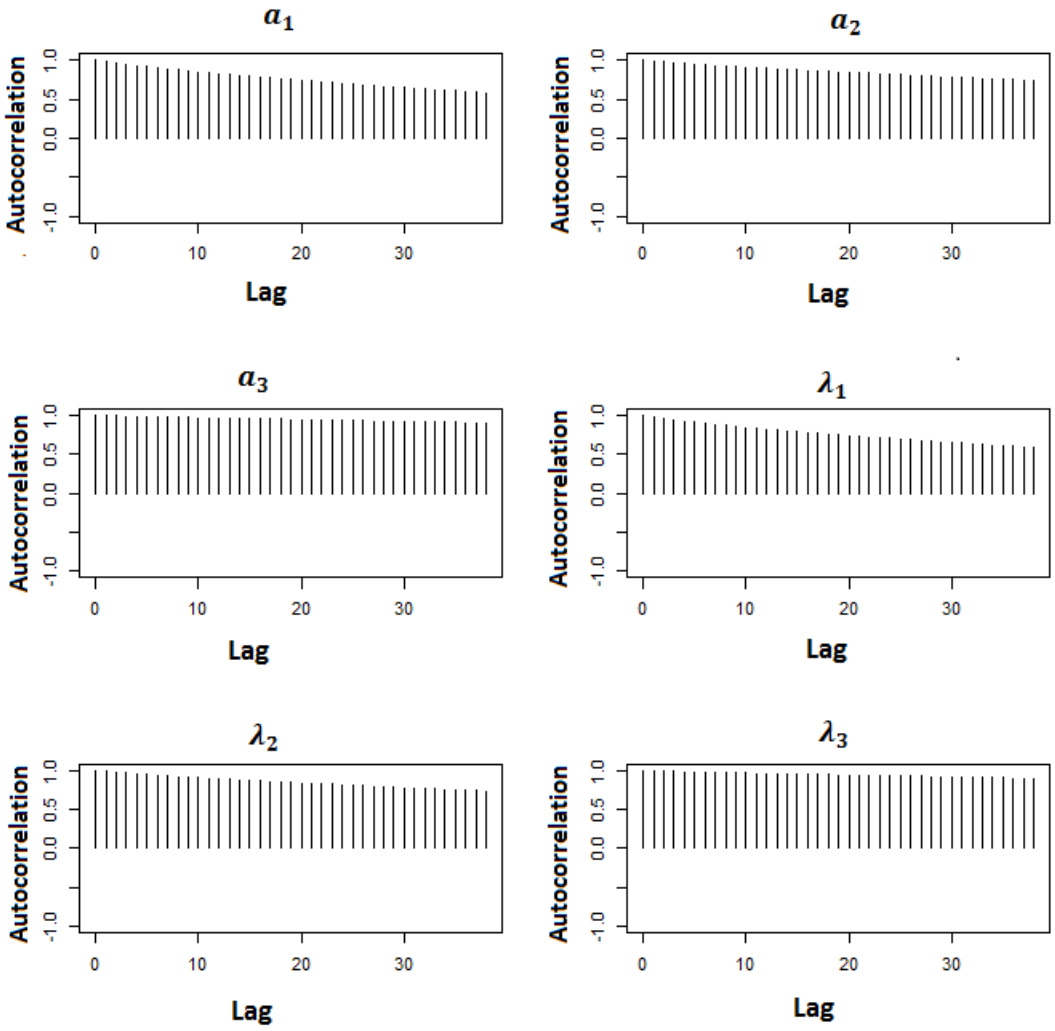

Figure 5.1: Autocorrelation plots of the maturation parameters $(a_j, \lambda_j,\ j = 1, 2, 3)$ estimates for the no hazard rates model.

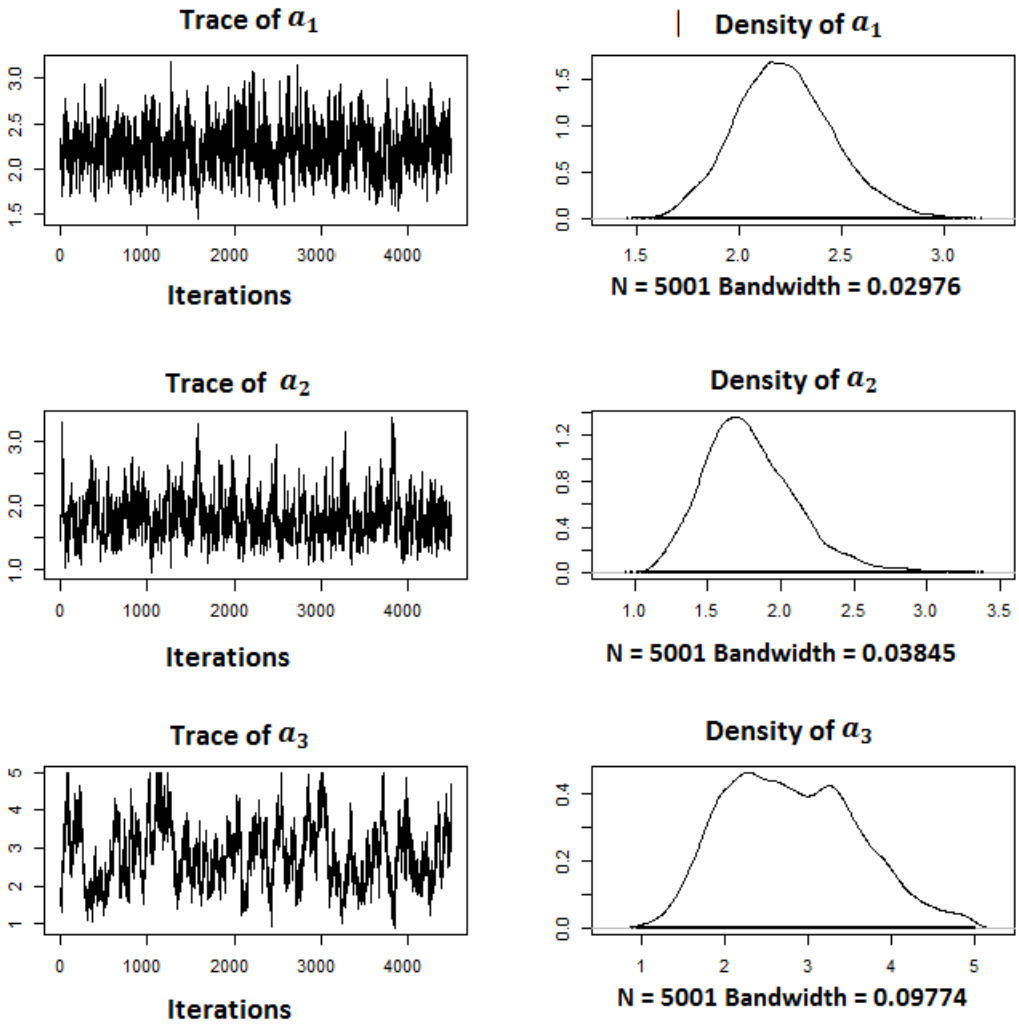

Figure 5.2: MCMC traces and density plots of the shape parameter (a_j, $j = 1, 2, 3$) estimates for the no hazard rates model.

5.1.3 The MH algorithm based on deterministic transformation

We applied the Metropolis-Hastings algorithm based on deterministic transformations (Section 4.2.4.2) to estimate the maturation parameters in order to increase the mixing from Section 5.1.2. The initial values $\left(a_1^{(0)}, \lambda_1^{(0)}, a_2^{(0)}, \lambda_2^{(0)}, a_3^{(0)}, \lambda_3^{(0)}\right)$ were chosen either randomly or from the single MH algorithm output (Table 5.1).

The sampling rates $s = (s_1, s_2, s_3)$ were estimated as $(0.75, 0.77, 0.70)$ for stage 1, stage 2 and stage 3, respectively. The optimal tuning parameters (σ_{a_j}, $j = 1, 2, 3$) for proposal

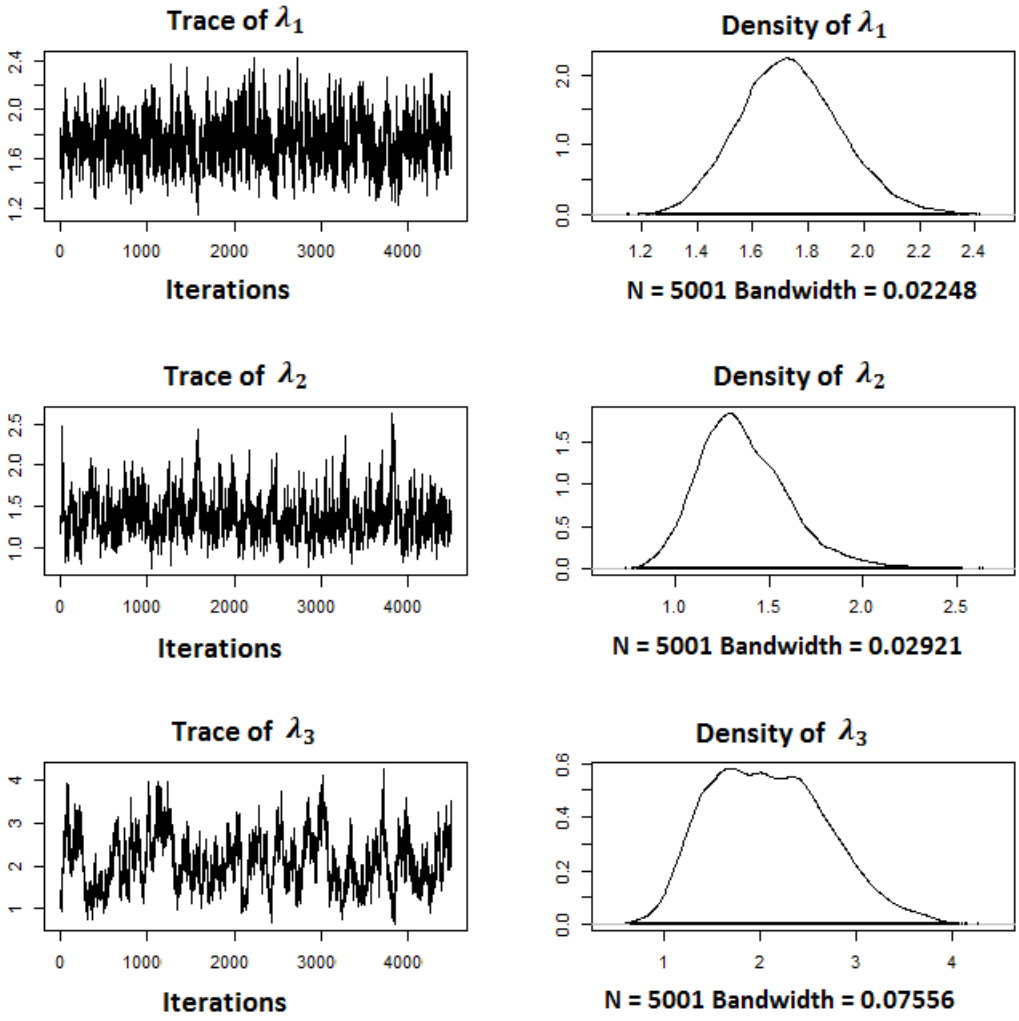

Figure 5.3: MCMC traces and density plots of the rate parameter (λ_j, $j = 1, 2, 3$) estimates for the no hazard rates model.

values of shape parameters (a_j, $j = 1, 2, 3$) were set at 0.238, 0.476 and 0.932 for the three stages (Section 2.3.5). The trace, density and autocorrelation plots of the shape and rate estimates are shown in Figures 5.4, 5.5 and 5.6. The trajectories of the chains are visually consistent over time and the marginal distributions of the parameters are generally unimodal. These figures indicate that the chains converge to the target stationary distribution of interest. Moreover, the autocorrelation plots of the shape and the rate samples show that the chains are mixing well, with less dependence as lag increases. The acceptance rates of the shape variables fell in a reasonable range at 33.48%, 32.84% and 42.38% in the three stages.

A summary of the Gelman and Rubin and the Geweke diagnostic tests is presented in Table 5.2 and the potential scale reduction factors changing through the iterations are shown in Figures 5.7 and 5.8. Because the potential scale reduction factors approach 1, the MCMC chains are diagnosed as converging to the stationary distribution of the parameters (Cowles and Carlin 14). In the Geweke diagnostic tests, the standard Z-scores all have absolute values less than two ($|Z| \leq 2$), which also indicates convergence of the MCMC chains.

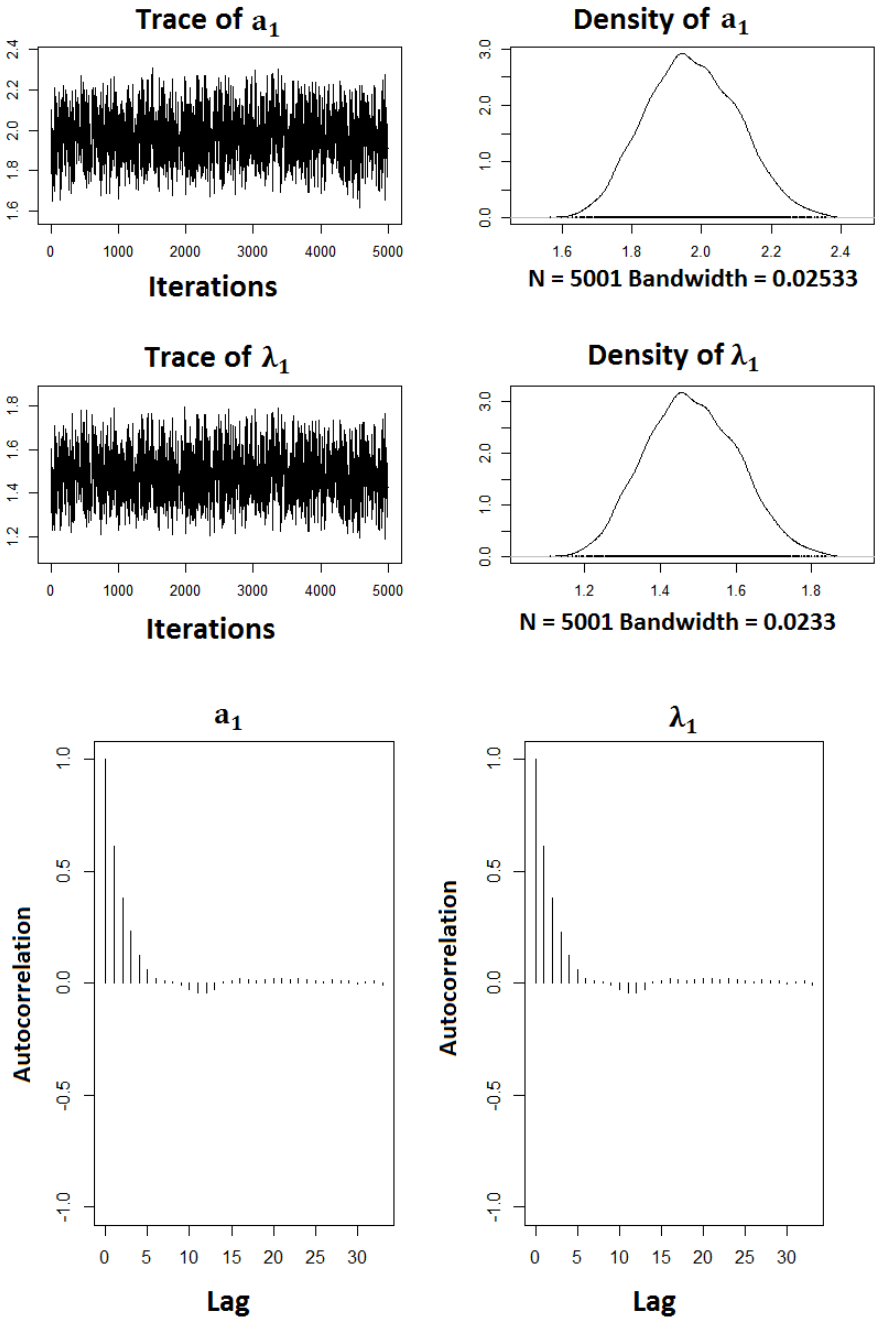

Figure 5.4: MCMC trace, density and autocorrelation plots of shape (a_1) and rate (λ_1) estimates at stage 1 for the no hazard rate model.

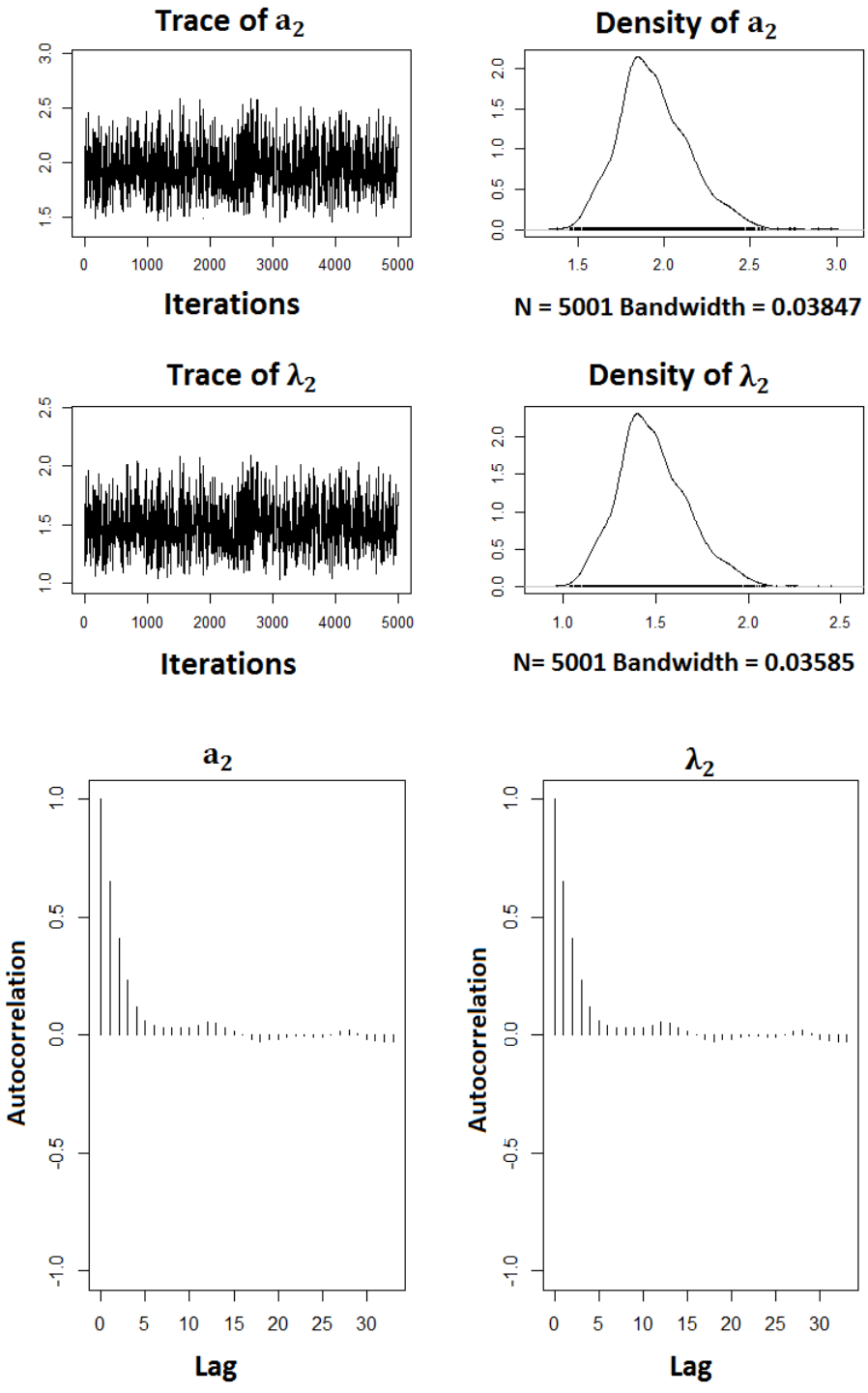

Figure 5.5: MCMC trace, density and autocorrelation plots of shape (a_2) and rate (λ_2) estimates at stage 2 for the no hazard rate model.

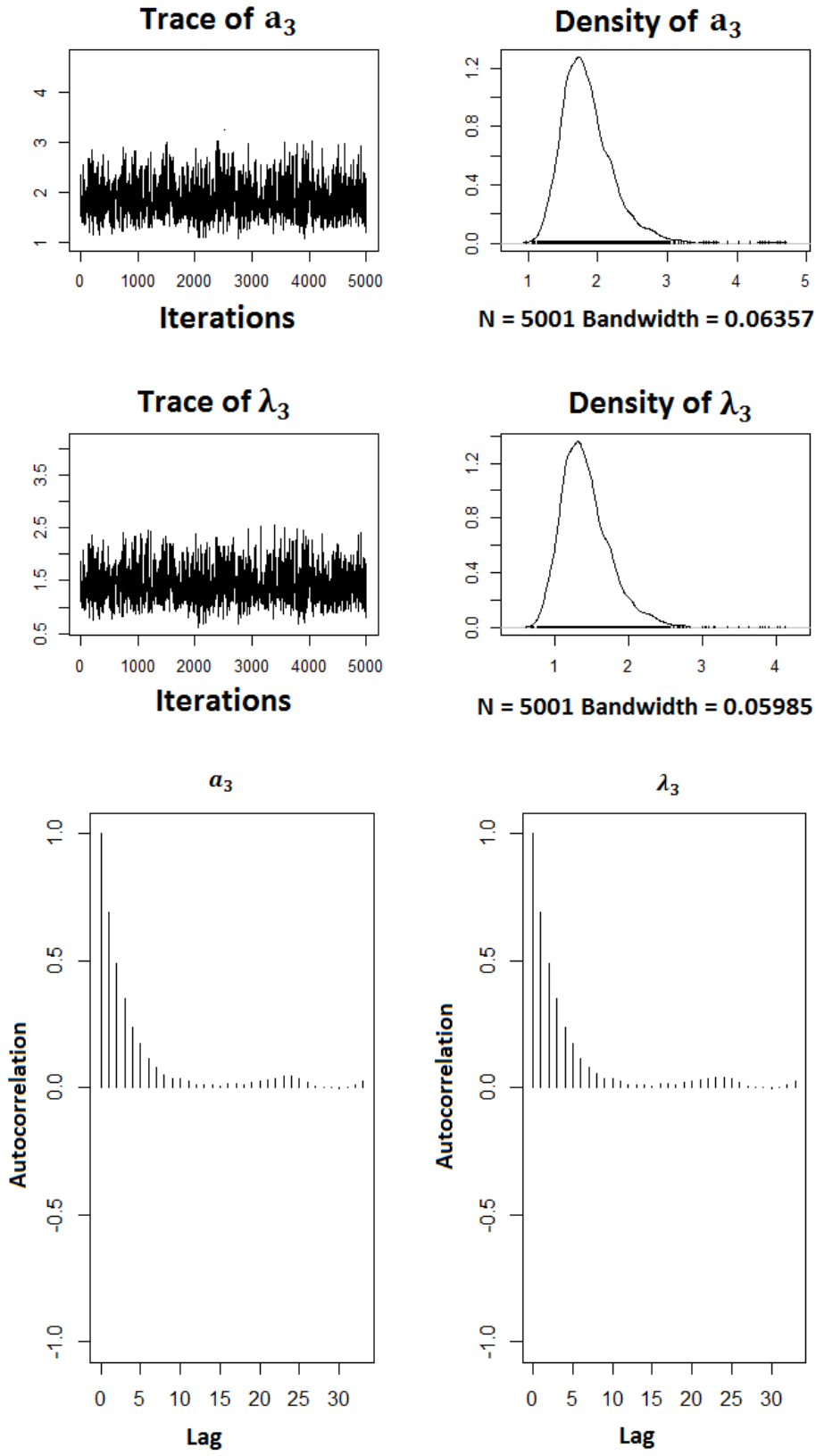

Figure 5.6: MCMC trace, density and autocorrelation plots of shape (a_3) and rate (λ_3) estimates at stage 3 for the no hazard rate model.

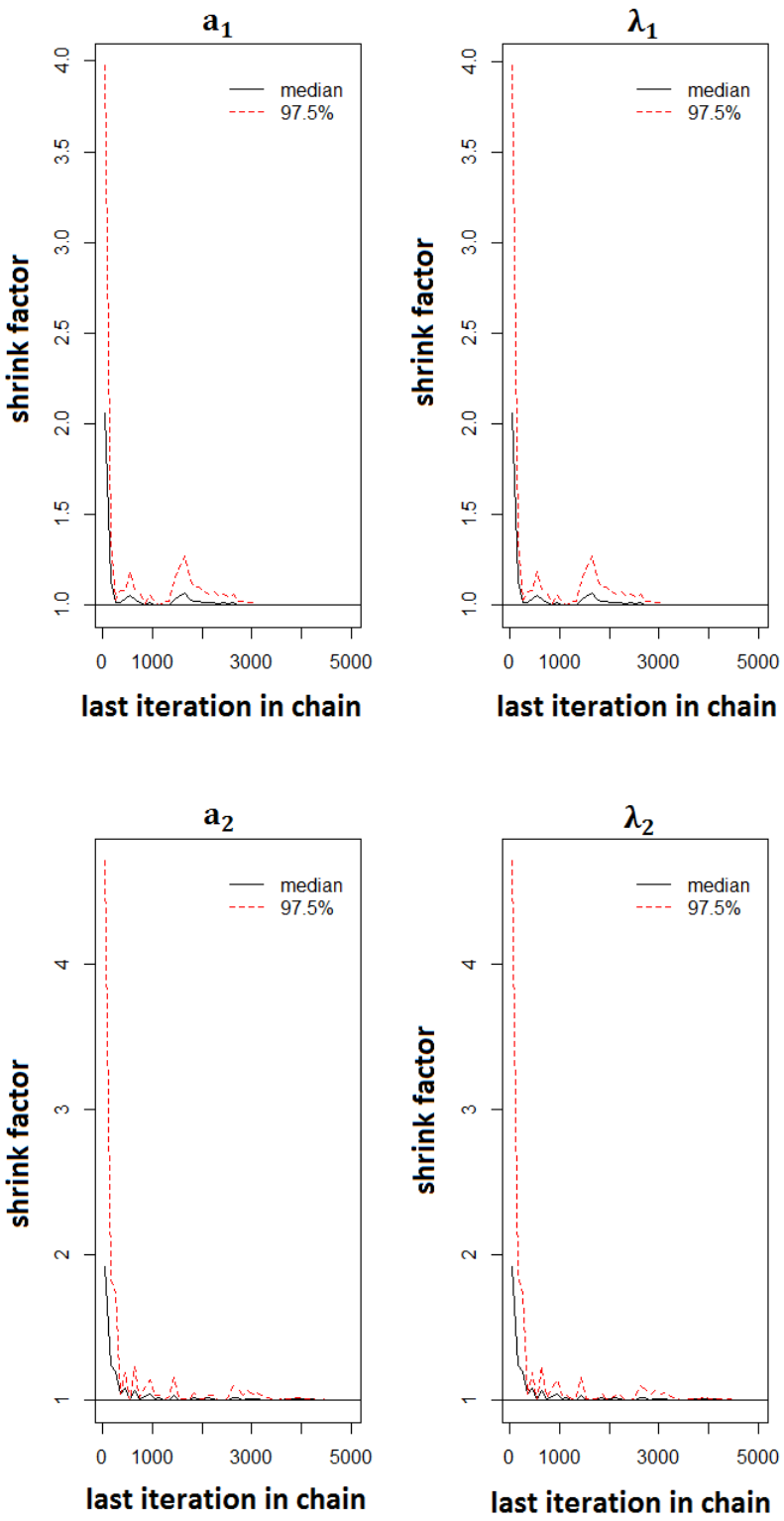

Figure 5.7: The potential rate reduction factor plots from the Gelman and Rubin diagnostic test of shape (a_j, $j = 1, 2$) and rate (λ_j, $j = 1, 2$) estimates at stage 1 and stage 2 for the no hazard rate model for five Markov chains of length 10,000 iterations.

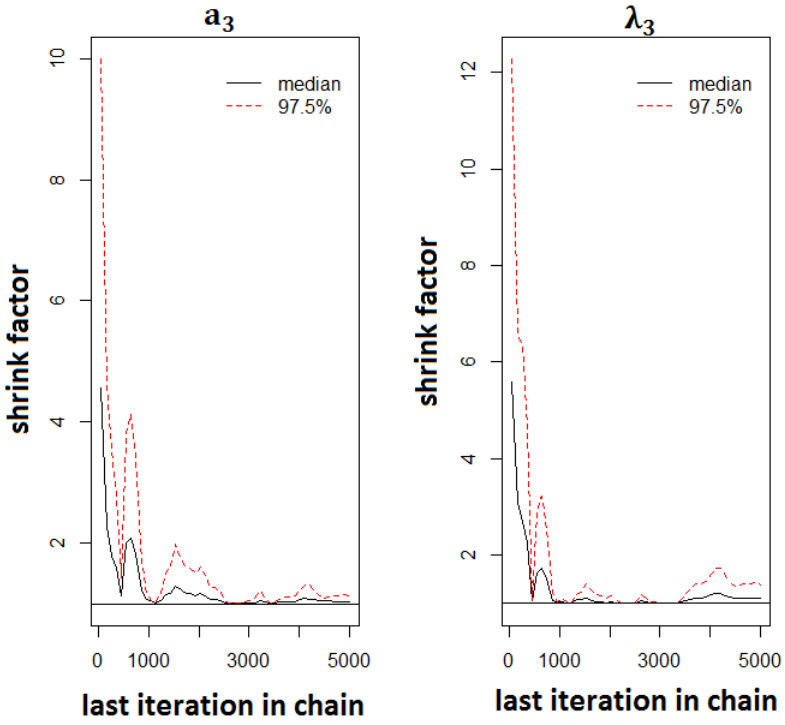

Figure 5.8: The potential rate reduction factor plots from the Gelman and Rubin diagnostic test of shape (a_3) and rate (λ_3) estimates at stage 3 for the no hazard rate model for five Markov chains of length 10,000 iterations.

For the sake of brevity, similar supporting figures in Sections 5.2 and 5.3 have been placed in Appendices C and D.

In order to evaluate the accuracy of the estimates, we generated 50 simulated datasets. The means of the estimates are close to the true values and the true values are also within their 95% credible intervals (Table 5.3). Interval estimations were described by credible intervals (CrI) estimating the probability that true values lie in the interval. The credible performance is measured by the percentage of estimated values that lie in the CrI.

The fitted curves for the three stages are shown in Figure 5.9. The fitted solid curve represents estimated proportions of live individuals obtained by first estimating the parameters from the simulated data and then plotting the probabilities that depend on these estimates. The dotted curve represents the observed proportions of live individuals calculated directly from the simulated data. Finally, the two dashed line curves represent the 95% confidence bands around the estimated proportions. It can be seen that the

Table 5.2: Summary of MCMC convergence diagnostic tests of shape a_j and rate $\lambda_j, j = 1, 2, 3$, estimates for the no hazard rate model.

Gelman and Rubin diagnostic			**Geweke diagnostic**	
Potential scale reduction factors			Fraction in 1st window	0.1
	Point est.	Upper C.I.	Fraction in 2nd window	0.5
a_1	1.00	1.00	a_1	-1.179
λ_1	1.00	1.00	λ_1	-1.178
Multivariable psrf	1.00			
a_2	1.00	1.00	a_2	0.957
λ_2	1.00	1.00	λ_2	0.958
Multivariable psrf	1.00			
a_3	1.01	1.03	a_3	1.259
λ_3	1.01	1.03	λ_3	1.250
Multivariable psrf	1.01			

estimated proportions are close to the observed proportions calculated directly. However, the estimations in the third stage show some biases, especially at the peak of the distribution. A poor performance of the credible intervals in stage 3 also displays these biases, in comparison to the previous stages. Estimation errors from the later stages are bigger due to the accumulation of estimation errors from the previous stages.

Table 5.3: Summary of results for parameter estimations from 50 simulated data for the no hazard rate model.

Parameter	**True value**	**Mean**	**95% credible interval**	**CrI performance**
a_1	2	2.109	[1.800;2.340]	98%
λ_1	1.5	1.547	[1.314;1.796]	100%
a_2	2	1.963	[1.545;2.443]	96%
λ_2	1.5	1.474	[1.127;1.915]	100%
a_3	2	2.232	[1.455;3.220]	98%
λ_3	1.5	1.675	[1.027;2.707]	94%

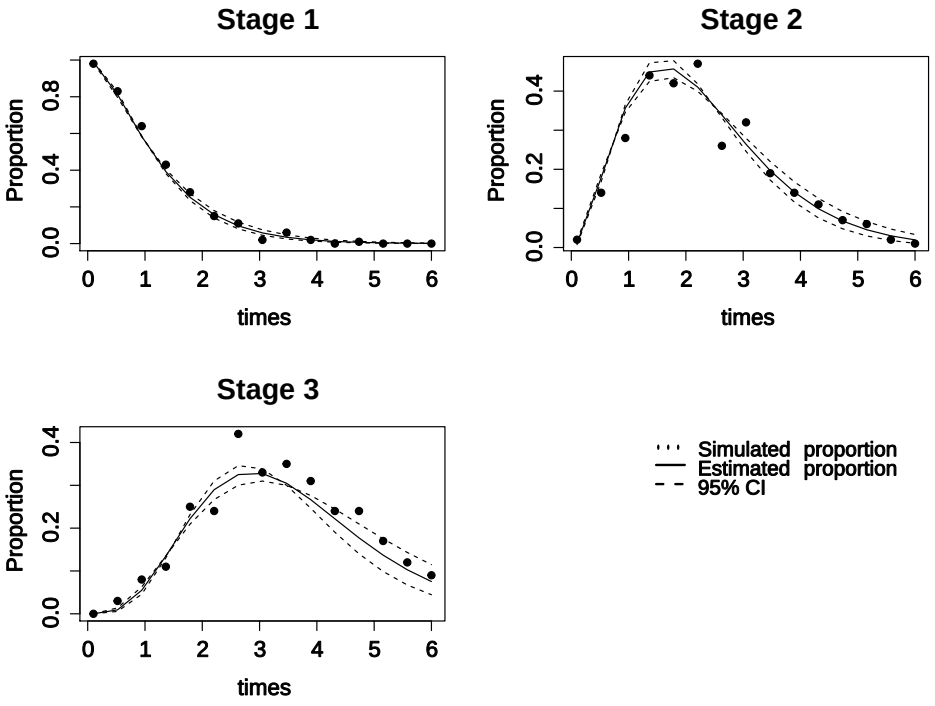

Figure 5.9: Proportions of alive individuals and estimated proportions of alive individuals and the 95% CrI of the estimations in three stages for the no hazard rate model.

5.2 Simulation data in stage-wise constant hazard rate model

The data in this case were simulated following Section 3.4.1. Unlike the study in Chapter 3, which used 50 sampling time points, this simulation only used 15 sampling time points taken between 0.1 and 6. One thousand individuals at each sampling time were generated from Erlangian distributions with constant shape $a = 2$ and with rates $\lambda_1 = 1.5$, $\lambda_2 = 1.5$ and $\lambda_3 = 1.5$ for stage 1, stage 2 and stage 3, respectively. The number of dead individuals was generated from an exponential distribution with the rates of $\mu_1 = 0.3$, $\mu_2 = 0.5$ and $\mu_3 = 0.7$ in stage 1, stage 2 and stage 3 respectively.

5.2.1 The probabilities at each stage

This section presents the calculation of probability at each stage in order to calculate the acceptance probabilities (Section 4.3.3). For given maturation parameters

$$\theta = (\theta_1, \theta_2, ..., \theta_9) = (a_1, \lambda_1, \mu_1, a_2, \lambda_2, \mu_2, a_3, \lambda_3, \mu_3)$$

in each stage, using (4.3.2) and (4.3.3) the probability of the number of alive and dead individuals in each stage at each sampling time were calculated as shown below.

The probability of an individual being alive in stage 1 at sampling time T_k is

$$\begin{aligned} p_1(T_k) &= H_1(T_k) \\ &= S_1(t_k) \int_{t_k}^{\infty} g_1(x)dx = \exp\left(-\int_0^{T_k} \mu_1 dx\right) \int_{T_k}^{\infty} \frac{x^{a_1-1} \exp(-\lambda_1 x) \lambda_1^{a_1}}{(a_1-1)!} dx \\ &= \frac{\exp(-\mu_1 T_k) \lambda_1^{a_1}}{(a_1-1)!} \int_{T_k}^{\infty} x^{a_1-1} \exp(-\lambda_1 x)\, dx\,, \end{aligned} \tag{5.2.1}$$

and the Laplace transform is

$$\begin{aligned} \mathcal{L}\{p_1(t)\}(s) &= \mathcal{L}\{H_1(t)\}(s) \\ &= \frac{1-\mathcal{L}(h_1)}{s+\mu_1} = \frac{1}{s+\mu_1}\left(1-\left(\frac{\lambda_1}{\lambda_1+\mu_1+s}\right)^{a_1}\right). \end{aligned} \tag{5.2.2}$$

Similarly, the probability of an individual being alive in stage 2 at sampling time T_k and its Laplace transform are given by

$$\begin{aligned} p_2(T_k) &= h_1 * H_2(T_k) \\ \Rightarrow \mathcal{L}\{p_2(t)\}(s) &= \mathcal{L}\{h_1\}(s)\mathcal{L}\{H_2(t)\}(s) \\ &= \frac{1}{s+\mu_2}\left(\frac{\lambda_1}{\lambda_1+\mu_1+s}\right)^{a_1}\left(1-\left(\frac{\lambda_2}{\lambda_2+\mu_2+s}\right)^{a_2}\right). \end{aligned} \tag{5.2.3}$$

The probability of an individual being alive in stage 3 at sampling time T_k and its Laplace transform are given by

$$
\begin{aligned}
p_3(T_k) &= h_1 * h_2 * H_3(T_k) \\
\Rightarrow \mathcal{L}\left\{p_3(t)\right\}(s) &= \mathcal{L}\left\{h_1\right\}(s)\mathcal{L}\left\{h_2\right\}(s)\mathcal{L}\left\{H_3(t)\right\}(s) \\
&= \frac{1}{s+\mu_3}\left(\frac{\lambda_1}{\lambda_1+\mu_1+s}\right)^{a_1}\left(\frac{\lambda_2}{\lambda_2+\mu_2+s}\right)^{a_2} \times \\
&\quad \left(1-\left(\frac{\lambda_3}{\lambda_3+\mu_3+s}\right)^{a_3}\right).
\end{aligned} \tag{5.2.4}
$$

The Laplace transform of the probability of an individual being found dead in stage 1 at sampling time t is

$$
\begin{aligned}
\mathcal{L}\left\{d_1(t)\right\}(s) &= \mathcal{L}\left\{H_1^d(t)\right\}(s) \approx \mathcal{L}\left\{\mu_1 \int_0^t g_1(x)dx\right\}(s) \\
&= \frac{\mu_1}{s}\left(\frac{\lambda_1}{\lambda_1+s}\right)^{a_1}.
\end{aligned} \tag{5.2.5}
$$

The Laplace transform of the probability of an individual being found dead in stage 2 at sampling time t is

$$
\begin{aligned}
\mathcal{L}\left\{d_2(t)\right\}(s) &= \mathcal{L}\left\{h_1 * H_2^d(t)\right\}(s) \\
&\approx \frac{\mu_2}{s}\left(\frac{\lambda_1}{\lambda_1+\mu_1+s}\right)^{a_1}\left(\frac{\lambda_2}{\lambda_2+s}\right)^{a_2}.
\end{aligned} \tag{5.2.6}
$$

The Laplace transform of the probability of an individual being found dead in stage 3 at sampling time t is

$$
\begin{aligned}
\mathcal{L}\left\{d_3(t)\right\}(s) &= \mathcal{L}\left\{h_1 * h_2 * H_3^d(t)\right\}(s) \\
&\approx \frac{\mu_3}{s}\left(\frac{\lambda_1}{\lambda_1+\mu_1+s}\right)^{a_1}\left(\frac{\lambda_2}{\lambda_2+\mu_2+s}\right)^{a_2}\left(\frac{\lambda_3}{\lambda_3+s}\right)^{a_3}.
\end{aligned} \tag{5.2.7}
$$

Practically, we need to adjust sampling times if there are no deaths occurring at the beginning of stage 3 for some of the sampling times. We assume that dead individuals in stage 3 are still alive at these sampling times. The adjustment is presented in Lemma 4 as follows.

Lemma 4. *In stage 3 at the initial sampling time τ, where no deaths occur, we need to subtract τ from our estimation, namely*

$$
\mathcal{L}\left\{H_3^d(t-\tau)u_\tau(t)\right\}(s) = \exp\left(-\tau(s+\mu_3)\right)\mathcal{L}\left\{H_3^d(t)\right\}(s), \tag{5.2.8}
$$

where the step function $u_\tau(t)$ is defined by

$$u_\tau(t) = \begin{cases} 0 & t < \tau\,, \\ e^{-\tau\mu_3} & t > \tau\,. \end{cases} \tag{5.2.9}$$

Proof. We have

$$\begin{aligned} \exp\left(-\tau(s+\mu_3)\right)\mathcal{L}\left\{H_3^d(t)\right\}(s) &= \exp\left(-\tau(s+\mu_3)\right)\int_0^\infty \exp(-st)H_3^d(t)dt \\ &= \int_0^\infty \exp\left(-s(t+\tau)\right)\exp(-\tau\mu_3)H_3^d(t)dt\,. \end{aligned} \tag{5.2.10}$$

Setting $x = t + \tau$ in (5.2.10) and observing that $x \geq \tau$, we obtain

$$\begin{aligned} \exp\left(-\tau(s+\mu_3)\right)\mathcal{L}\left\{H_3^d(t)\right\}(s) &= \int_\tau^\infty \exp(-sx)\exp(-\tau\mu_3)H_3^d(x-\tau)dx \\ &= \int_0^\infty \exp(-sx)H_3^d(x-\tau)u_\tau(x)dx \\ &= \mathcal{L}\left\{H_3^d(t-\tau)u_\tau(t)\right\}(s)\ . \end{aligned} \tag{5.2.11}$$

□

Taking the inverse Laplace transforms of (5.2.2), (5.2.3), (5.2.4), (5.2.5), (5.2.6) and (5.2.7), we obtain the probabilities $p_1(T_k)$, $p_2(T_k)$, $p_3(T_k)$, $d_1(T_k)$, $d_2(T_k)$ and $d_3(T_k)$, respectively. These probabilities were used in (4.3.10) in order to calculate the posterior distributions needed to evaluate the acceptance probabilities in Algorithm 3.

5.2.2 The single MH algorithm

Similar to Section 5.1.2, the MH algorithm in Section 4.3.3 was applied in order to estimate the maturation parameters and hazard rates in each stage. For each of three stages, the starting values and the initial tuning parameters for proposal values of maturation parameters and hazard rate parameters were set at different values. The results from the Markov chains were compared and all agreed with each other. The optimal tuning parameter for proposal values were chosen by the adaptive MH method (Section 2.3.5). A

summary of the maturation parameters from 100,000 MH iterations with 95,000 samples discarded as burn-in is shown in Table 5.4. The means of the maturation parameters in the three stages converge slowly to true values of these parameters. The acceptance rates were too high for some parameters. The autocorrelation plots for our MH samples in Figure B1.3 (Appendix B) show very slow mixing and strong dependence as lags increase. The trajectories of the chains (Figures B1.1 and B1.2, Appendix B) are visually inconsistent over time and indicate that some of the marginal distributions of the parameters are bimodal. The figures show that the chains do not converge to the target stationary distribution. Therefore, the MH Algorithm 4 based on deterministic transformations in Section 4.3.4 was applied in order to improve mixing of the chains.

Table 5.4: Summary of the maturation parameters and hazard rates estimates resulting from appling single Metropolis-Hastings algorithm from the last 5,000 iterations for stage-wise constant hazard rates model.

Parameter	**True value**	**Acceptance rate %**	**Mean**	**SD**	**2.5%**	**50%**	**97.5%**
a_1	2	30.56	1.924	0.079	1.773	1.926	2.088
a_2	2	26.16	2.296	0.249	1.826	2.278	2.764
a_3	2	49.94	3.040	1.025	1.570	2.767	4.864
λ_1	1.5	50.92	1.473	0.061	1.362	1.475	1.596
λ_2	1.5	45.58	1.741	0.214	1.324	1.729	2.153
λ_3	1.5	68.68	2.287	0.856	1.040	2.069	3.856
μ_1	0.3	76.24	0.279	0.010	0.261	0.279	0.300
μ_2	0.5	73.40	0.510	0.020	0.471	0.510	0.548
μ_3	0.7	82.39	0.762	0.037	0.687	0.765	0.836

5.2.3 The MH algorithm based on deterministic transformations

We applied the Metropolis-Hastings Algorithm 4 based on deterministic transformations (Section 4.3.4) to estimate the maturation parameters. The initial values $\left(a_1^{(0)}, \lambda_1^{(0)}, a_2^{(0)}, \lambda_2^{(0)}, a_3^{(0)}, \lambda_3^{(0)}\right)$ were taken either randomly or from the single MCMC output (Table 5.4).

The sampling rate $s = (s_1, s_2, s_3)$ was estimated as $(0.77, 0.70, 0.85)$ for stage 1, stage 2 and stage 3, respectively. As in the previous section, $T = 10,000$ iterations were generated with the first 5,000 iterations discarded as burn-in. The trace, density and autocorrelation plots of the shape and rate estimates are shown in Figures D1.1, D1.2

and D1.3 (Appendix D). The trace plots show that the Markov chains appear to have reached their stationary distributions. Moreover, the density plots show smooth and unimodal posterior marginal distributions for each parameter. The autocorrelation plots of the shape, rate and hazard rate samples indicate that the chains are mixing well and are independent as lag increases. These figures confirm that the MCMC chains converge to the target stationary distributions. Acceptance rates of 48.80%, 34.74% and 51.55% for the shape variables ($a_j, j = 1, 2, 3$) and 45.37%, 47.85% and 38.08% for the hazard rate variables ($\lambda_j, j = 1, 2, 3$) were obtained at the three stages.

A summary of the Gelman and Rubin and the Geweke diagnostic test results is provided in Table 5.5 and the potential scale reduction factors changing through the iterations are shown in Figures C1.1, C1.2 and C1.3 (Appendix C). Because the potential scale reduction factors all approach one, the MCMC chains are diagnosed as converging to the stationary distribution of the parameters (Cowles and Carlin 14). In the Geweke diagnostic tests, the absolute values of the standard Z-scores were less than two ($|Z| \leq 2$), also indicating that the MCMC chains converged.

We applied the method to 50 simulated datasets. As presented in Table 5.6, the means of estimates are sufficiently close to the true values, with reasonably small standard deviations. Furthermore, the true values are within their 95% credible intervals. The credible interval performances are more accurate at earlier stages than later stages, because the biases in the later stages tend to have accumulated from the previous ones.

The fitted curves for the three stages are shown in Figure 5.10. The estimated proportions of alive individuals seem to be acceptably close to the simulated probabilities of an alive individual across the three stages. However, the estimated proportions of dead individuals (Figure 5.10) seem to be biased in the simulated probabilities, especially at the highest observed proportions. The explanation for this bias is that the exact death times of individuals were unknown.

Table 5.5: Summary of MCMC convergence diagnostic tests of shape, rate and hazard rate (a_j, λ_j and $\mu_j, j = 1, 2, 3$) estimates for the stage-wise constant hazard rate model.

Gelman and Rubin diagnostic			**Geweke diagnostic**	
Potential scale reduction factors			Fraction in 1st window	0.1
	Point est.	Upper C.I.	Fraction in 2nd window	0.5
a_1	1.00	1.00	a_1	-1.534
λ_1	1.00	1.00	λ_1	-1.328
μ_1	1.00	1.00	μ_1	-1.653
Multivariable psrf	1.00			
a_2	1.00	1.00	a_2	-0.217
λ_2	1.00	1.00	λ_2	-0.055
μ_2	1.00	1.00	μ_2	-1.201
Multivariable psrf	1.00			
a_3	1.02	1.02	a_3	1.451
λ_3	1.02	1.02	λ_3	1.306
μ_3	1.00	1.01	μ_3	2.025
Multivariable psrf	1.00			

Table 5.6: Summary of results for parameter estimations from 50 simulated datasets for the stage-wise constant hazard rate model.

Parameter	**True value**	**Mean**	95% **credible interval**	**CrI performance**
a_1	2	1.945	[1.939;2.321]	100%
λ_1	1.5	1.470	[1.353;1.621]	100%
μ_1	0.3	0.297	[0.290;0.319]	100%
a_2	2	2.178	[1.935;2.494]	96%
λ_2	1.5	1.612	[1.284;1.932]	100%
μ_2	0.5	0.503	[0.469;0.545]	98%
a_3	2	2.373	[1.609;2.847]	92%
λ_3	1.5	1.722	[1.177;2.635]	98%
μ_3	0.7	0.755	[0.685;0.809]	94%

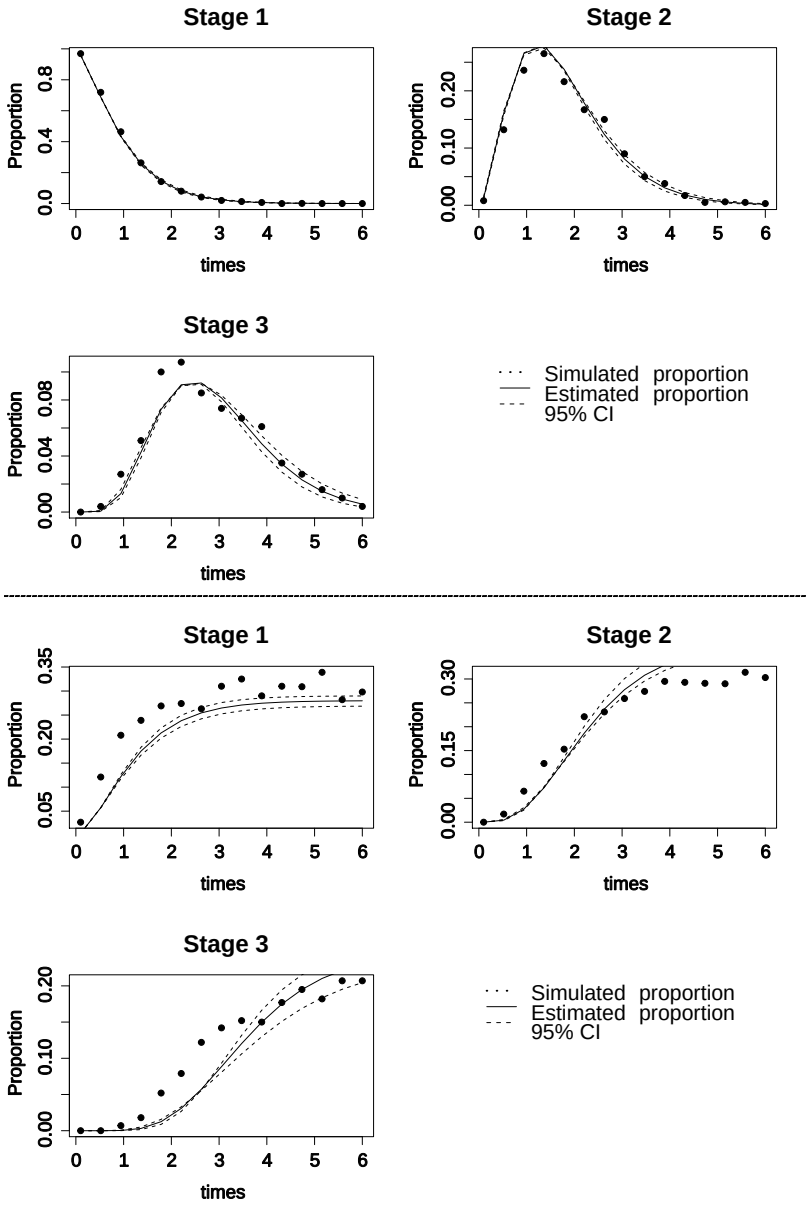

Figure 5.10: The figures in the top half are the observed proportions of alive individuals and the estimated proportion of alive individuals and the 95% CrI of the estimations for the three stages. The figures in the bottom half are the observed proportions of dead individuals and the estimated proportion of dead individuals and the 95% CrI of the estimations for three stages.

5.3 Simulation data in the linear time-dependent hazard rates model

5.3.1 The probabilities at each stage

We applied the single MH Algorithm 5 (Section 4.4) to simulated data that were cohort stage frequency data with 3 stages having Erlangian distributions. The hazard rates were set at 0, 0.3 and $2t$ for stage 1, stage 2 and stage 3, respectively. One thousand stage times at each sampling time were generated from Erlangian distributions with constant shape $a_j = 2$ and rates $\lambda_j = 1.5$, $j = 1, 2, 3$ for stage 1, stage 2 and stage 3, respectively. Fifteen sampling time points were taken between 0.1 and 6.

Because the hazard rate in stage 1 was zero, from (5.1.1) and (5.1.2), the Laplace transform of the probability of an individual being alive in stage 1 at sampling time T_k is presented as follows:

The probability of an individual being alive in stage 1 at sampling time T_k is

$$\begin{aligned} p_1(T_k) &= H_1(T_k) \\ &= \int_{T_k}^{\infty} g_1(x)dx = \int_{T_k}^{\infty} \frac{x^{a_1-1}\exp\left(-\lambda_1 x\right)\lambda_1^{a_1}}{(a_1-1)!}dx \\ &= \frac{\lambda_1^{a_1}}{(a_1-1)!}\int_{T_k}^{\infty} x^{a_1-1}\exp\left(-\lambda_1 x\right)dx\,, \end{aligned} \tag{5.3.1}$$

and the Laplace transform is

$$\begin{aligned} \mathcal{L}\left\{p_1(t)\right\}(s) &= \mathcal{L}\left\{H_1(t)\right\}(s) \\ &= \frac{1-\mathcal{L}\left\{h_1\right\}(s)}{s} = \frac{1}{s}\left(1-\left(\frac{\lambda_1}{\lambda_1+s}\right)^{a_1}\right)\,. \end{aligned} \tag{5.3.2}$$

The probability of an individual being alive in stage 2 at sampling time T_k and its Laplace transform are given by

$$\begin{aligned} p_2(T_k) &= h_1 * H_2(T_k) \\ \Rightarrow \mathcal{L}\left\{p_2(t)\right\}(s) &= \mathcal{L}\left\{h_1\right\}(s)\mathcal{L}\left\{H_2(t)\right\}(s) \\ &= \frac{1}{s+\mu_2}\left(\frac{\lambda_1}{\lambda_1+s}\right)^{a_1}\left(1-\left(\frac{\lambda_1}{\lambda_2+\mu_2+s}\right)^{a_2}\right)\,. \end{aligned} \tag{5.3.3}$$

From (4.4.14), $\beta_3(s) = \mathcal{L}\{h_3(t)\}(s) = \mathcal{L}\{g_j(t)\exp(-\gamma_{j1}t^2/2)\}(s)$ is calculated as

$$\beta_3(s) \approx \frac{\lambda_j^{a_j}}{\Gamma(a_j)} \frac{(\lambda_j + s)^{a_j-2}}{\gamma_{j1}^{a_j-1}} \left(2 - a_j + \frac{-a_j^2 + 4a_j - 5}{2u_0}\right) \quad , \tag{5.3.4}$$

where $u_0 = (\lambda_3 + s)^2/(2\gamma_3)$.

From (4.4.15), the probability of an individual being alive in stage 3 at sampling time T_k and its Laplace transform are given by

$$\begin{aligned}
p_3(T_k) &= h_1 * h_2 * H_3(T_k) \\
\Rightarrow \mathcal{L}\{p_3(t)\}(s) &= \mathcal{L}\{h_1\}(s)\mathcal{L}\{h_2\}(s)\mathcal{L}\{H_3(t)\}(s) \\
&\approx \left(\frac{\lambda_1}{\lambda_1 + s}\right)^{a_1} \left(\frac{\lambda_2}{\lambda_2 + \mu_2 + s}\right)^{a_2} \\
&\quad \left\{\exp\left(\frac{s^2}{2\gamma_3}\right)\sqrt{\frac{\pi}{2\gamma_3}}\left[1 - \operatorname{erf}\left(\sqrt{\frac{\gamma_3}{2}}\frac{s}{\gamma_3}\right)\right] - 0.6\sqrt{\frac{\pi}{2\gamma_3}}\beta_3(s)\right\} \\
&\approx \left(\frac{\lambda_1}{\lambda_1 + s}\right)^{a_1} \left(\frac{\lambda_2}{\lambda_2 + \mu_2 + s}\right)^{a_2} \\
&\quad \left\{\exp\left(\frac{s^2}{2\gamma_3}\right)\sqrt{\frac{\pi}{2\gamma_3}}\operatorname{erfc}\left(\sqrt{\frac{\gamma_3}{2}}\frac{s}{\gamma_3}\right) - 0.6\sqrt{\frac{\pi}{2\gamma_3}}\beta_3(s)\right\} \\
&\approx \left(\frac{\lambda_1}{\lambda_1 + s}\right)^{a_1} \left(\frac{\lambda_2}{\lambda_2 + \mu_2 + s}\right)^{a_2} \\
&\quad \left\{\exp\left(\frac{s^2}{2\gamma_3}\right)\sqrt{\frac{\pi}{2\gamma_3}}0.6\exp\left(-\frac{s^2}{2\gamma_3}\right) - 0.6\sqrt{\frac{\pi}{2\gamma_3}}\beta_3(s)\right\} \\
&\approx \left(\frac{\lambda_1}{\lambda_1 + s}\right)^{a_1} \left(\frac{\lambda_2}{\lambda_2 + \mu_2 + s}\right)^{a_2} 0.6\sqrt{\frac{\pi}{2\gamma_3}}\left(1 - \beta_3(s)\right) \ .
\end{aligned} \tag{5.3.5}$$

The Laplace transform of the probability of an individual being found dead in stage 2 at sampling time t is

$$\begin{aligned}
\mathcal{L}\{d_2(t)\}(s) &= \mathcal{L}\left\{h_1 * H_2^d(t)\right\}(s) = \mathcal{L}\{h_1\}(s)\,\mathcal{L}\left\{H_2^d(t)\right\}(s) \\
&\approx \frac{\mu_2}{s}\left(\frac{\lambda_1}{\lambda_1 + s}\right)^{a_1}\left(1 - \left(\frac{\lambda_1}{\lambda_2 + s}\right)^{a_2}\right) \ .
\end{aligned} \tag{5.3.6}$$

From (4.4.16), the Laplace transform of the probability of an individual being found dead in stage 3 at sampling time t is

$$\begin{aligned}
\mathcal{L}\{d_3(t)\}(s) &= \mathcal{L}\left\{h_1 * h_2 * H_3^d(t)\right\}(s) = \mathcal{L}\{h_1\}(s)\,\mathcal{L}\{h_2\}(s)\,\mathcal{L}\left\{H_3^d(t)\right\}(s) \\
&\approx \left(\frac{\lambda_1}{\lambda_1 + s}\right)^{a_1}\left(\frac{\lambda_2}{\lambda_2 + \mu_2 + s}\right)^{a_2}\frac{f(E(\mathcal{T}_3))}{s}\left(\frac{\lambda_3}{s + \lambda_3}\right)^{a_3} \ .
\end{aligned} \tag{5.3.7}$$

Similar to Section 5.2.1, we need to adjust the sampling times in the case that no deaths occur in stage 3 at the beginning of some sampling times. The adjustment is presented in Lemma 5 as follows.

Lemma 5. *In stage 3 at initial sampling time τ, where no deaths occur, we need to subtract this time τ from our estimation.*

$$\mathcal{L}\left\{H_3^d(t-\tau)u_\tau(t)\right\}(s) = \exp\left(-\frac{\gamma_3}{2}\tau^2 - s\tau\right)\mathcal{L}\left\{H_3^d(t)\right\}(s)\ , \tag{5.3.8}$$

where step function $u_\tau(t)$ is defined by

$$u_\tau(t) = \begin{cases} 0 & t < \tau\ , \\ \exp\left(-\frac{\gamma_3}{2}\tau^2\right) & t > \tau\ . \end{cases} \tag{5.3.9}$$

Proof. By an argument analogous to that used to prove Lemma 4. □

By taking the inverse Laplace transform of (5.3.2),(5.3.3), 5.3.5, 5.3.6 and (5.3.7), we obtain the probabilities $p_1(T_k)$, $p_2(T_k)$, $p_3(T_k), d_2(T_k)$ and $d_3(T_k)$, respectively. These probabilities are used in (4.4.5) in order to calculate the acceptance probabilities in Algorithm 5.

5.3.2 The single MH algorithm

Similar to the previous model, we applied the MH Algorithm 5 (Section 4.4.3) to estimate the maturation parameters and the hazard rate in each stage. The purpose of doing this was to investigate prior distributions and choose appropriate initial values. For each of the three stages, the starting values were set at 0.1 and the initial tuning parameters for proposal values of maturation parameters and hazard rate parameters were set at 0.1 and 0.01 respectively. The optimal tuning parameter for proposal values were chosen by applying the adaptive MH method (Section 2.3). A summary of the maturation parameters from 100,000 MH iterations with 95,000 samples discarded as burn-in is shown in Table 5.7. The means of the maturation parameters in the three stages converge slowly to true values of these parameters. The autocorrelation plots of samples in Figure B2.3 (Appendix B) show the chains are not mixing well and exhibit too much dependence as lag increases. The trace plots (Figures B2.1 and B2.2) show that the Markov chains appear not to have approached their stationary distributions sufficiently closely.

Table 5.7: Summary of the matuation parameters and hazard rates estimates result from appling single MH algorithm from the last 5,000 iterations for linear time-dependent hazard rates model.

Parameter	True value	Acceptance rate %	Mean	SD	2.5%	50%	97.5%
a_1	2	22.55	1.905	0.064	1.788	1.900	2.027
a_2	2	17.25	2.861	0.118	4.580	4.893	4.997
a_3	2	45.84	2.899	0.115	2.575	2.944	2.998
λ_1	1.5	37.98	1.442	0.044	1.363	1.440	1.536
λ_2	1.5	78.71	2.018	0.090	1.849	2.011	2.219
λ_3	1.5	43.21	1.703	0.169	1.377	1.700	2.014
μ_2	0.3	23.52	0.415	0.069	0.282	0.418	0.539
γ_3	2	47.69	1.398	0.334	0.861	1.357	2.081

5.3.3 The MH algorithm based on deterministic transformations

We applied the MH Algorithms 2 and 4 based on deterministic transformations (Sections 4.2.4.2 and 4.3.4.2) to the problems in stage 1 and stage 2, respectively. After we obtained parameter estimates from stage 1 and stage 2, we applied the MH Algorithm 5 in Section 4.4.3 for stage 3. The initial values $\left(a_1^{(0)}, \lambda_1^{(0)}, a_2^{(0)}, \lambda_2^{(0)}, a_3^{(0)}, \lambda_3^{(0)}\right)$ were taken either randomly or from the single MCMC output (Table 5.7).

The sampling rates (s_1, s_2) were estimated as 0.77 and 0.7 for stage 1 and stage 2 respectively. From 10,000 iterations with 5,000 samples discarded as burn-in, the traces of the shape and rate estimates in the three stages are shown in Figures D2.1, D2.2 and D2.3 (Appendix D). We see that the trajectories of the Markov chains are consistent over time and that the distributions of the parameters look appropriately normal. The figure shows that the Markov chains converged to the target stationary distribution of interest. Moreover, the autocorrelation plots of the shape and the rate estimates in stage 1 show that the chains are mixing well and that there is not too much dependence as lag increases (Figure D2.1).

Means and standard deviations of the Markov chains are close to the true values of the parameters (Table 5.9). The acceptance rates for the MH algorithm are at 44.13% and 37.04% for the shape parameters (a_1 and a_2) in stage 1 and stage 2, respectively and 59.59%, 41.75%,

42.10%, 34.50% for μ_2, a_3, λ_3 and γ_3 parameters respectively. In order to evaluate the accuracy of the estimates, we generated 50 data sets. Means and standard deviations of the parameters were computed from the 50 datasets. The means of estimates are close to the true values, with a reasonably small standard deviations (Table 5.9). Furthermore, the true values are within their 95% credible intervals. The credible interval performances are better in early stages than later stages, because the bias in later stages tend to accumulate from the previous ones.

However, there is bias in the estimate of the slope parameter γ_3. The credible interval of slope parameter γ_3 is the worst in term of CrI performance compared to the other parameters (Table 5.9). This is due to bias in the approximations in (5.3.4), (5.3.5) and (5.3.7). Note that in the simulated data, only Stage 3 has a time-dependent hazard rate. The method is reliable in a model which has only one time-dependent hazard rate in Stage 1 or Stage 2. If the number of stages having a time-dependent hazard rate increases or the time-dependent hazard rate occurs in a later stage, biases from the estimates will increase. In such cases, we should consider an approach that does not depend on the probabilities $p_j(t)$ and $d_j(t), j = 1, ..., I$. This can reduce the bias in the approximations in (5.3.4), (5.3.5) and (5.3.7).

Table 5.8: Summary of MCMC convergence diagnostic tests of shape, rate and hazard rate estimates for the linear time-dependent hazard rate model.

Gelman and Rubin diagnostic			**Geweke diagnostic**	
Potential scale reduction factors			Fraction in 1st window	0.1
	Point est.	Upper C.I.	Fraction in 2nd window	0.5
a_1	1.00	1.01	a_1	-0.260
λ_1	1.00	1.01	λ_1	-0.260
Multivariable psrf	1.00			
a_2	1.03	1.10	a_2	0.473
λ_2	1.08	1.32	λ_2	0.472
μ_2	1.10	1.37	μ_2	1.453
Multivariable psrf	1.08			
a_3	1.02	1.02	a_3	-0.217
λ_3	1.26	1.90	λ_3	-0.055
γ_3	1.01	1.03	μ_3	-1.201
Multivariable psrf	1.65			

Table 5.9: Summary of results for parameter estimations from 50 simulated datasets for the linear time-dependent hazard rate model.

Parameter	**True value**	**Mean**	95% **credible interval**	**CrI performance**
a_1	2	1.997	[1.904;2.060]	100%
λ_1	1.5	1.497	[1.409;1.552]	100%
a_2	2	2.014	[1.839;2.309]	96%
λ_2	1.5	1.515	[1.432;1.903]	98%
μ_2	0.3	0.390	[0.281;0.409]	96%
a_3	2	2.012	[1.877;2.326]	96%
λ_3	1.5	1.614	[1.424;1.734]	92%
γ_3	1	1.1945	[0.946;1.303]	90%

In order to assess convergence of the Markov chain, we used Gelman and Rubin multiple sequence diagnostic and Geweke diagnostic. We ran 5 MCMC chains of length 10,000 with 5,000 samples discarded as burn-in. A summary of Gelman and Rubin diagnostics and Geweke diagnostics are presented in Table 5.8 and the potential scale reduction factors changing through the iterations are shown in Figures C2.1, C2.2 and C2.3 (Appendix C). Because the potential scale reduction factors are close to 1, the MCMC chains are diagnosed as converging to the stationary distribution of the parameters. The standard Z-scores in Geweke diagnostic tests, all have $|Z| \leq 2$, and so indicate that the MCMC chains converged.

The fitted curves for the three stages are shown in Figures 5.11. The estimated proportions of alive individuals seem acceptably close to the simulated probabilities. Moreover, the estimated values follow the simulated data set very closely at each sampling time. However, in the case of dead individuals in stages 2 and 3, Figure 5.11 indicates lack of agreement between the estimated and the simulated proportions, except in the range where these proportions and sampling times have low values. The explanation for this bias probably stems from the fact that the exact death times of individuals are unknown. This leads to some bias in the estimated dataset.

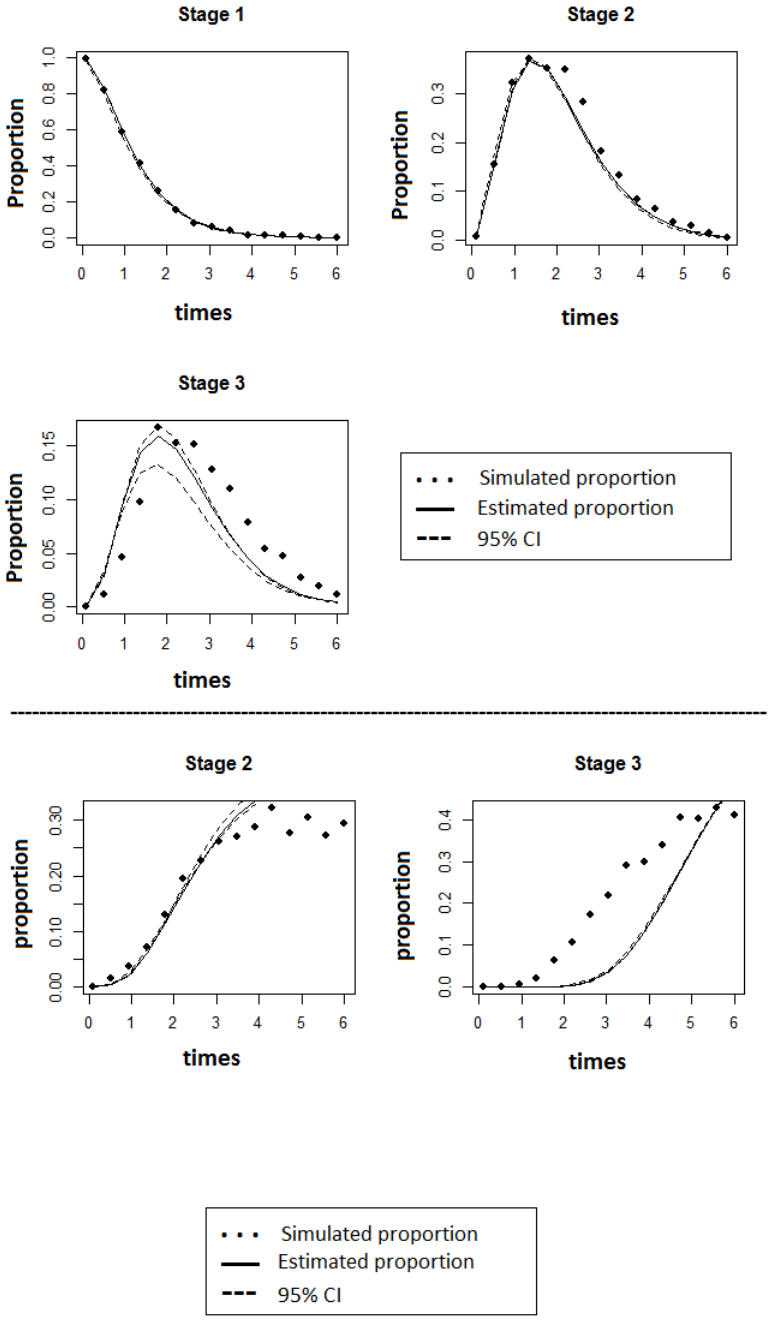

Figure 5.11: The figures in the top half are proportions of living individuals and estimated proportion of alive individuals and the 95% CrI of the estimations for three stages. The figures in bottom half are proportion of dead individuals and estimated proportion of dead individuals and the 95% CrI of the estimations for Stage 2 and Stage 3 in the linear time-dependent hazard rates model.

Chapter 6

Case Studies

In this chapter, we apply the techniques in Chapter 4 to cattle parasitic data ([58]) and breast development of New Zealander schoolgirls data ([39], p. 98). The no hazard rate model was applied to these data. We used the Bayesian approach of Section 4.2 to estimate parameters for the model. The results show that the proposed methods are able to estimate the parameters well, as compared to using the Laplace transform methods ([25] and [39]).

6.1 Parasitic nematode Data

The data for cattle parasite was taken from [58]. Data are presented in Table 2.1 (Section 2.1). The data consist of four stages of the parasite life cycle including stage 1 (eggs), stage 2 (first stage larvae), stage 3 (second stage larvae) and stage 4 (third stage larvae). The authors of Hoeting et al. [25] introduced the Laplace transform methods in order to estimate the parameters for the first three stages of the parasite life cycle. In this section, the techniques from Section 4.2 will be used in order to estimate maturation parameters from the first three stages of the parasitic nematode life cycle. This technique improves on the technique used in [25] by estimating the unknown shape parameters simultaneously with the rate parameters. In [25], the values of shape parameters are assumed to be the same and equal two. However, in most real situation, the values of the shape parameters are not known. Assumed values may not be correct and lead to inaccurate estimates of rate parameters. Although there are not many differences between the fitted curves for the three stages shown in Figure 6.1, the methods presented in this work will have more pronounced impact if the number of stages is large.

6.1.1 The single MH algorithm

Similar to Section 5.1, we applied the MH Algorithm 1, from Section 4.2.3, to estimate the maturation parameters in each stage. For each of three stages, the starting values were set at 0.1 and the initial tuning parameters for proposal values of shape parameters were set at 0.2 and 0.01 for rate parameters. A summary of the maturation parameters from 100,000 MH iterations with 95,000 samples discarded as burn-in is shown in Table 6.1. The means of the maturation parameters in three stages converge slowly to true values of these parameters. The autocorrelation plots for our MH samples in Figure B3.3 (Appendix B) show bad mixing and substantial dependence as lag increases. The trace plots and the distributions of the chains are presented in Figures B3.1 and B3.2 (Appendix B). The plots indicate that the chains failed to converge. Therefore, we apply the joint MH Algorithm 2 based on deterministic transformations in the next section in order to improve the convergence rates of the chains.

Table 6.1: Summary results of parameter estimation applied single MH Algorithm 1 based on the last 5,000 iterations for parasitic nematode data.

Parameter	**Acceptance rate %**	**Mean**	**SD**	**2.5%**	**50%**	**97.5%**
a_1	37.30	1.700	0.223	1.301	1.704	2.186
a_2	64.75	2.482	0.537	1.539	2.457	3.610
a_3	28.96	1.077	0.171	0.807	1.063	1.455
λ_1	19.64	0.040	0.005	0.032	0.040	0.050
λ_2	58.78	0.104	0.024	0.065	0.103	0.157
λ_3	9.89	0.015	0.003	0.010	0.014	0.020

6.1.2 The MH algorithm based on deterministic transformations

As in Section 4.2, the initial values for parameters were set randomly. The sampling rate s was chosen as the estimated mean duration, which was taken from the output of a single MH algorithm (Algorithm 1). The sampling rate $s = (s_1, s_2, s_3)$ was estimated as $(0.02, 0.04, 0.01)$ for stage 1, stage 2 and stage 3, respectively. Tuning variances from the normal random walk distributions were optimized using an adaptive MH method (Section

2.3.5).

The trace, density and autocorrelation plots of the shape and rate samples are shown in Figures D3.1, D3.2 and D3.3 (Appendix D). The autocorrelation plots of the shape samples and the rate samples show that the chains are mixing well and that they are reasonably independent as the lag increases. The trace plots show that the means of the Markov chains are constant and stabilized. The density plots show the desired stabilizations and indicate the convergence of the chains.

To further assess the MCMC convergence, we used the Gelman and Rubin multiple sequence diagnostic and the Geweke diagnostic tests. We ran five MCMC chains of length $T = 10,000$ iterations with the first 5,000 iterations discarded as burn-in. A summary of the Gelman and Rubin and the Geweke diagnostic tests is expressed in Table 6.2 and the potential scale reduction factor plots from Gelman and Rubin diagnostic are shown in Figures C3.1 and C3.2 (Appendix C). These tests indicate that the MCMC chains have converged. Because the potential scale reduction factors approach 1, the MCMC chains are diagnosed as converging to the stationary distribution of the parameters. The standard Z-scores, which all have absolute values less than two ($|Z| \leq 2$) in the Geweke diagnostic tests, also indicate that the MCMC chains converged. Means of the estimates presented in Table 6.3 have reasonably small standard deviations. The acceptance rate for the MH algorithm is at 32.25%, 54.36% and 63.86% for the three stages.
The fitted curves for the three stages are shown in Figure 6.1. The empirical proportions and the Laplace transform estimated proportions ([25]) are very close to the estimated probabilities of an individual being alive across each stage. Although the proportions are not very different between the Laplace transform method and the MH method based on deterministic transformations, note that in a Bayesian approach, parameters in each stage are estimated without any information regarding shape and rate parameters, in contrast to the Laplace transform methods. The Bayesian approach produces good estimates for multi-stage models.

6.2 Breast development of New Zealander schoolgirls

We chose the dataset of breast development of New Zealander schoolgirls ([39], p. 98) as an example of a stage-duration model with no hazard rate. A survey of New Zealan-

Table 6.2: Summary of MCMC convergence diagnostic tests of shape and rate estimates for the parasitic nematode data.

Gelman and Rubin diagnostic			**Geweke diagnostic**	
Potential scale reduction factors			Fraction in 1st window	0.1
	Point est.	Upper C.I.	Fraction in 2nd window	0.5
a_1	1.00	1.00	a_1	-0.341
λ_1	1.00	1.00	λ_1	-0.340
Multivariable psrf	1.00			
a_2	1.00	1.01	a_2	-0.760
λ_2	1.00	1.01	λ_2	-0.762
Multivariable psrf	1.00			
a_3	1.00	1.01	a_3	0.782
λ_3	1.00	1.01	λ_3	0.778
Multivariable psrf	1.01			

Table 6.3: Summary for shape and rate estimates applied Algorithm 2 based on MCMC runs of length 10,000 of three stages of parasitic nematode data.

Parameter	**Mean**	**SD**	**2.5%**	**50%**	**97.5%**
a_1	1.922	0.149	1.635	1.918	2.224
λ_1	0.043	0.004	0.035	0.043	0.052
a_2	1.745	0.369	1.241	1.683	2.533
λ_2	0.077	0.020	0.050	0.074	0.120
a_3	1.277	0.159	0.986	1.272	1.608
λ_3	0.016	0.003	0.011	0.016	0.021

der schoolgirls at different ages was conducted to assess the stages of breast development. The breast development was divided into five stages. At different ages, a different random number of New Zealander schoolgirls were taken to assess their breast development. Further analysis of each sample was not possible. Thus, this dataset is considered to consist of destructive samples. We could not access actual numbers of girls at each stage since the data available only recorded percentages at each stage, perhaps, after some processing or cleansing. The data are given in Table 6.4, including the 5 stages of breast development. Percentages of New Zealander schoolgirls were recorded from 20 sampling times from 6.5 years old to 25.5 years old. In this section, the techniques from Section 4.3 are used in

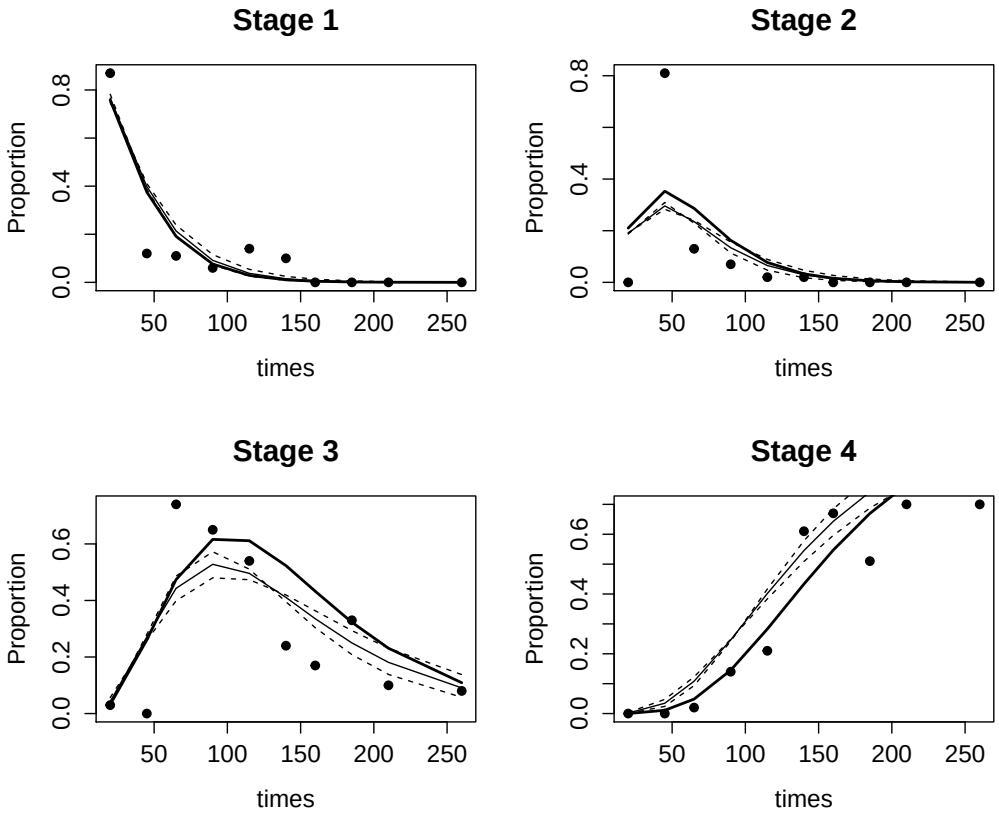

Figure 6.1: Plots of the proportions from sampling data (dotted line), the estimated proportion from MCMC method with the 95% CrI (solid line and dash lines) and the estimated proportion from Hoeting et al. [25] (bold solid line) from the 4 stages of the parasite life cycle.

order to estimate maturation parameters from the five stages of breast development. The improvements of these techniques in estimations of maturation durations at each stage are compared to methods in [39] that just estimated the percentages with different stages of the breast development and the mean ages of entry in stages 1-5. Our methods estimate not only the percentages with different stages of the breast development and the mean ages of entry in stages 1-5, but also the distributions of breast development in each stage. Therefore, our results give more information about breast development of New Zealander schoolgirls.

6.2.1 The single MH algorithm

Similar to the previous section, we applied the MH Algorithm 1 (Section 4.2.3) to estimate the maturation parameters in each stage. The dataset, comprising percentages of New Zealander schoolgirls in different stages at different ages, is considered to have a multinomial distribution. This natural assumption was used in [25] and [18].

For each of the four stages, the starting values were set at 0.1 and the initial tuning parameters for proposal values of parameters were set at 0.1. Table 6.5 shows the summary of the maturation parameters from 100,000 MH iterations with 95,000 samples discarded as burn-in. The autocorrelation plots for our MH samples in Figure B4.3 (Appendix B) shows bad mixing and substantial dependence as lag increase. The trace plots and the distributions of the chains are presented in Figures B4.1 and B4.2 (Appendix B). The plots show that the means of the Markov chains are not stable. These results indicated that the chains failed to converge. Therefore, we will apply the joint MH Algorithm 2 based on deterministic transformations in the next section in order to improve the convergence rates of the chains.

6.2.2 The MH algorithm based on deterministic transformations

We applied the MH Algorithm 2 based on deterministic transformations, described in Section 4.2.4, to estimate the maturation parameters in order to increase the mixing from Section 5.1.2. We updated each stage separately, starting from stage 1 and moving to the next stages sequentially. The initial values for parameters were randomly taken from 95% CrI of the above MCMC chain (Table 6.5). The sampling rate s was chosen as the estimated mean duration which is the output from 10,000 MCMC iterations in Table 6.5. Tuning variances from the normal random walk distributions were optimized using an adaptive MH method (Section 2.3.4).

The trace, density and autocorrelation plots of the shape and rate estimates are shown in Figures D4.1, D4.2, D4.3 and D4.4 (Appendix D). The autocorrelation plots of the shape and the rate estimates showed that the chains mixed well and that there was not too

Table 6.4: Percentages of New Zealander schoolgirls over five stages of development from 20 sampling times from 6.5 to 25.5 years old.

Time	**Stage 1**	**Stage 2**	**Stage 3**	**Stage 4**	**Stage 5**
6.5	100	0	0	0	0
7.5	100	0	0	0	0
8.5	100	0	0	0	0
9.5	95.1	4.8	0.1	0	0
10.5	66.4	27.6	5.3	0.7	0.1
11.5	32.6	37.7	22.8	6.2	0.7
12.5	13.3	27.5	35.8	19.4	3.9
13.5	5	15.2	34.4	33.3	12
14.5	1.9	7.5	25.7	40.2	24.8
15.5	0.7	3.5	16.8	39.1	40
16.5	0.2	1.6	10.2	33.2	54.8
17.5	0.1	0.7	5.9	26	67.3
18.5	0	0.3	3.4	19.2	77.1
19.5	0	0.1	1.9	13.7	84.3
20.5	0	0.1	1.1	9.5	89.4
21.5	0	0	0.6	6.5	92.9
22.5	0	0	0.3	4.4	95.3
23.5	0	0	0.2	3	96.9
24.5	0	0	0.1	2	97.9
25.5	0	0	0.1	1.3	98.6

Table 6.5: Summary results of parameter estimation applied the single MH Algorithm 1 based on the last 5,000 iterations for breast development of New Zealander schoolgirls data.

Parameter	**Acceptance rate %**	**Mean**	**SD**	**2.5%**	**50%**	**97.5%**
a_1	53.95	1.507	0.163	1.208	1.499	1.856
a_2	74.32	2.444	0.645	1.211	2.417	3.740
a_3	61.62	1.407	0.328	0.912	1.359	2.206
a_4	38.49	1.596	0.270	1.134	1.562	2.185
λ_1	69.22	0.928	0.092	0.766	0.925	1.131
λ_2	43.71	1.901	0.515	0.956	1.861	2.983
λ_3	31.57	0.843	0.188	0.564	0.817	1.300
λ_4	45.56	0.617	0.101	0.451	0.604	0.843

much dependence as lag increases. The trace plots show that the means of the Markov chains are stable. The density plots also indicate stabilization and convergence.

Means of the estimates presented in Table 6.3 have reasonably small standard deviations. The acceptance rates for the MH algorithm were 20.35%, 43.51%, 41.90% and 54.17% respectively for the four stages.

Table 6.6: Summary of MCMC convergence diagnostic tests of shape and rate estimates for the breast development data of New Zealander schoolgirls.

Gelman and Rubin diagnostic			**Geweke diagnostic**	
Potential scale reduction factors			Fraction in 1st window	0.1
	Point est.	Upper C.I.	Fraction in 2nd window	0.5
a_1	1.01	1.02	a_1	1.072
λ_1	1.01	1.08	λ_1	1.067
Multivariable psrf	1.30			
a_2	1.01	1.01	a_2	0.837
λ_2	1.01	1.01	λ_2	0.864
Multivariable psrf	1.01			
a_3	1.00	1.01	a_3	0.606
λ_3	1.00	1.01	λ_3	0.607
Multivariable psrf	1.00			
a_4	1.00	1.00	a_3	-0.159
λ_4	1.00	1.00	λ_3	-0.061
Multivariable psrf	1.00			

In order to assess convergence of the Markov chains, we used Gelman and Rubin multiple sequence diagnostic and Geweke diagnostic. We ran five MCMC chains of length 10,000 with 5,000 iterations discarded as burn-in. A summary of the Gelman and Rubin and Geweke diagnostic tests are presented in Table 6.6 and the potential scale reduction factor plots from the Gelman and Rubin diagnostic are shown in Figures C4.1, C4.2, C4.3 and C4.4 (Appendix C). The diagnostic tests all indicated that the MCMC chains have converged. Because the potential scale reduction factors are 1, the MCMC chains are diagnosed as converging to the stationary distribution of the parameters.

The fitted curves for the five stages are shown in Figure 6.2. The empirical proportions and the Laplace transform estimated proportions ([25]) are acceptably close to the esti-

Table 6.7: A summary results of parameter estimation using the MCMC Algorithm 2 based on deterministic transformations based on the last 5,000 iterations for the breast development of New Zealander schoolgirls.

Parameter	**Mean**	**SD**	**2.5%**	**50%**	**97.5%**
a_1	1.151	0.071	1.021	1.149	1.295
a_2	0.605	0.050	0.514	0.603	0.706
a_3	0.634	0.048	0.548	0.631	0.736
a_4	0.514	0.060	0.408	0.510	0.642
λ_1	1.076	0.097	0.917	1.066	1.295
λ_2	0.584	0.072	0.465	0.576	0.748
λ_3	2.042	0.301	1.586	1.999	2.744
λ_4	0.861	0.163	0.616	0.838	1.241

mated probabilities of an individual being alive across time in each stage (Figure 6.2). Note that in the Bayesian approach, parameters in each stage are estimated without any information about the shape and rate parameters in each stage compared to Laplace transform methods in Chapter 3. The Bayesian approach produces good estimates for multi-stage models.

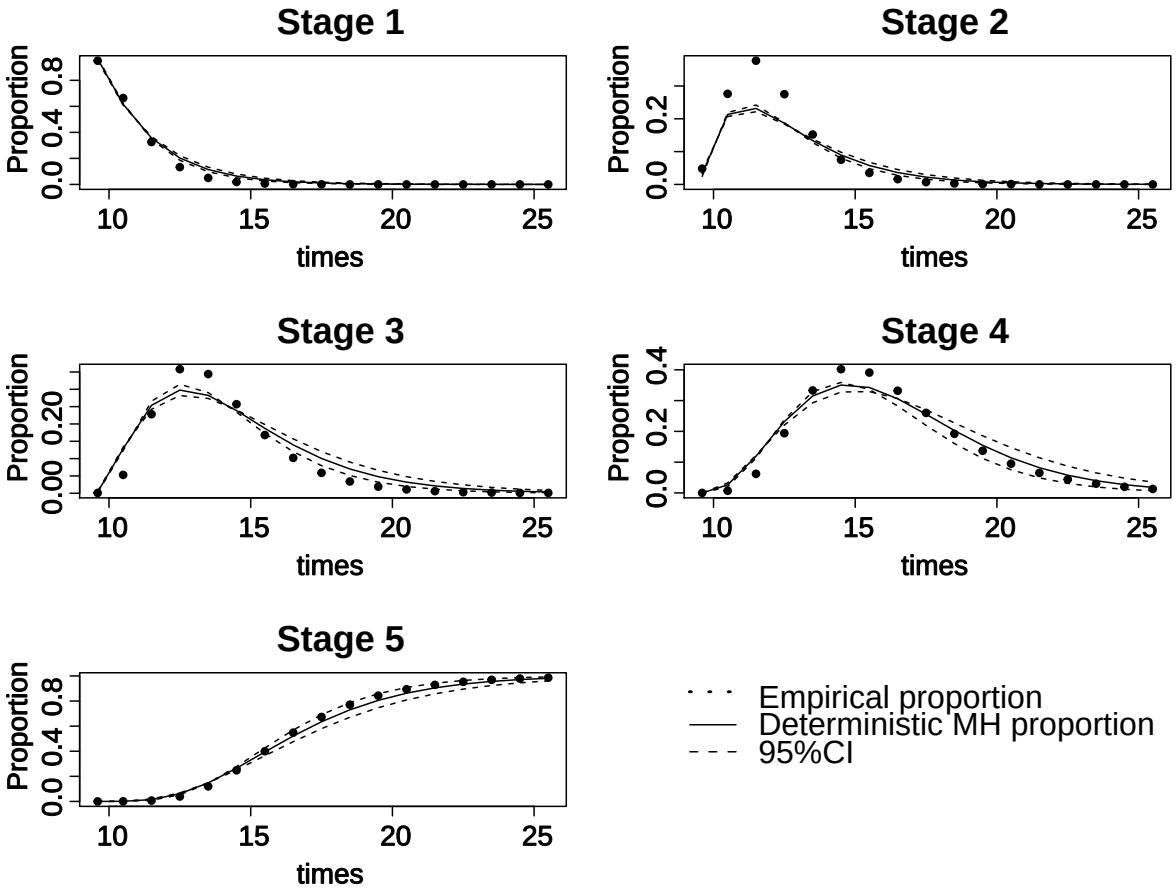

Figure 6.2: Plots of the proportions from sampling data, the estimated proportion from MCMC method with the 95% CrI from 5 stages of the breast development of New Zealander schoolgirls data.

Chapter 7

Summary, Conclusions and Discussion

Summary

Estimating how much time an average individual spends in each stage and the probability that an individual transits to a specific stage at a given time are central issues of stage-duration models. The models proposed in this thesis are distinguished by multi-stage models in which the stage of sampled individuals is assessed and the individuals are removed from further consideration. Stage duration and hazard rate in each stage provide a basic understanding of population biology and ecology. Exploring the development in each stage can yield basic understanding of progress of an individual. Treatments can be compared by estimating growth rates of focused stages. By affecting the covariates, new treatments could be developed through the improvement of the growth rates in stages of interest.

The contributions of the thesis consist of novel methods to estimate maturation parameters and hazard rate parameters for stage-duration distribution models. These methods are evaluated on both simulated data and experimental data. These methods contribute to the current literature of parameter estimation for stage-duration models. More specifically, the contributions include (but are not limited to) the following.

First, we used Laplace transform methods in order to estimate parameters for stage frequency data in stage-wise constant and linear time-dependent hazard rate cases (Chapter 3). The parameters are estimated with assumptions that constant shape parameters or constant rate parameters are known. Moreover, we explored relationships of maturation parameters in each stage. These fundamental relationships are subsequently embedded in MCMC methods based on deterministic transformations in Chapter 4.

Second, we applied MCMC method based on deterministic transformations in order to relax the assumptions above (Chapter 4). New methods were developed in order to estimate

maturation and hazard rate parameters for stage frequency data. The first contribution of these methods is that we have relaxed the assumptions of the known constant maturation parameters by introducing priors for these parameters. The parameters in each stage are estimated through a Bayesian analysis approach without the need for the initial information for shape or rate parameters required by earlier approaches ([51]; [25]). The other improvement is that the number of sampling times is reduced compared to the Laplace transform methods (Chapter 3). The third achievement is that hazard rate parameters in each stage are estimated simultaneously with the maturation parameters using a Bayesian approach. Thus the approach provides more information about uncertainties of the parameters for stage frequency data.

Third, the above methods were applied to the parasitic nematode data as well as data from the breast development of New Zealander schoolgirls (Chapter 6). The results show that the proposed methods are able to estimate the parameters well compared to previous studies ([39]; [25]). The contribution of this study is that the assumptions about shape and rate parameters are reduced compared to the Laplace transform methods. Parameters in each stage are estimated without any information regarding the shape and rate parameters.

Conclusions and discussion

Naturally, many problems still remain and are worth considering in the future. These include (but are not limited to) the following, interesting, areas for investigation.

- From Chapter 3: We explored the relationships between maturation parameters and estimated the maturation parameters. These relationships were used for the MH algorithm based on deterministic transformations in the following chapter. However, we did not focus on estimating hazard rate at each stage. Thus, the technique for estimating linear time-dependent hazard rate was not evaluated. Simulated data based on complex distributions of survival times with covariates in each stage will need to be investigated in the future.

- From Chapter 4: Maturation parameters and hazard rate parameters were estimated using the MH algorithm based on deterministic transformations. This Bayesian approach provides great advantages compared to other methods in the literature. The assumption about maturation parameters was relaxed and hazard rate parameters were estimated simultaneously with maturation parameters at each stage.

However, biases in later stages as well as the restriction to the hazard rates are limitations of the presented methodology. The biases in the later stages tend to accumulate from the previous stages. This affects the accuracy of the estimates when the number of stages increases. Moreover, maturation parameters and hazard rate parameters were only estimated when the mean of maturation time was larger than the mean of survival time in each stage. Furthermore, the proposed methods could not be applied in the situations where there are very high hazard rates. In particular, hazard rates must be less than one. An approximate Bayesian computation (ABC) approach will be considered in order to reduce biases when number of stages increases. This approach also allows determination of the largest hazards rates that can fit the model.

Large bias in the linear time-dependent hazard rate case needs to be overcome by applying other advanced MCMC methods. Future computational extensions of this work may include using advanced MCMC methods in order to reduce biases and speed up the convergence of Markov chains. In particular, in a model with linear time-dependent hazard rates, computing the likelihood functions is computationally intractable. The approximation in estimates creates biases. An ABC method will be used in this context.

- From Chapter 6: The methods in the previous chapters were implemented to evaluate parasitic nematode data and the breast development of New Zealander schoolgirl data. The no hazard rate model was applied to these data. We concluded that the proposed methods yielded results that fit well with the data. However, the data collected long time ago. The impact of the chapter's results on understanding the life cycle of the cattle parasite and the development of breasts of New Zealand girls might be not important. Further field work is necessary to confirm potential insight of our methods. We also did not have empirical data for models with stage-wise constant hazard rates and linear time-dependent hazard rates. In the future, we intend to identify and analyze data sets with these characteristics.

- The methods developed are of a mathematical nature. As they stand, they apply only in the context described in the thesis. Further extensions are likely to be possible but will require a level of development at least equivalent to that in the current thesis. Any discussion on the shape or extent of such extensions would be speculative.

Appendices

A. Typical simulated data for the models (Section 3.4).

A.1 Simulated data for the no hazard rate model

Table A.1: The distribution of 10 sampled individuals over four stages at 15 sampling time points, in which stage-specific mortality does not occur. The sample of 10 individuals were not the same at different time points and were referred to as stage times in previous literature. Stage 4 is the final stage (for example, the adult stage).

t	stage1	stage2	stage3	stage4
0.1	10	0	0	0
0.5	8	2	0	0
0.9	5	3	2	0
1.4	3	4	3	0
1.8	2	4	3	1
2.2	3	4	3	0
2.6	1	5	3	1
3.0	1	0	2	7
3.5	1	0	4	5
3.9	0	2	2	6
4.3	0	0	7	3
4.7	0	1	2	7
5.2	0	0	2	8
5.6	0	0	1	9
6	0	0	1	9

A.2 Simulated data for the stage-wise hazard rate model

Table A.2: The distribution of 1,000 sampled individuals over the four stages at 50 sampling time points in the stage-wise constant hazard rates case. The sample of 1,000 individuals were not the same at different time points. Stage 4 is the final stage. The table is continued on the next page.

t	stage1	stage2	stage3	stage4	death stage1	death stage2	death stage3
0.1	961	13	0	0	25	1	0
0.2	879	57	0	0	63	1	0
0.3	841	73	0	0	84	2	0
0.5	736	118	9	0	125	12	0
0.6	648	179	12	0	138	23	0
0.7	564	231	12	0	159	31	3
0.8	486	227	34	1	193	53	6
0.9	445	247	36	2	209	53	8
1.1	374	272	45	3	225	70	11
1.2	322	264	49	3	258	94	10
1.3	271	288	60	10	238	117	16
1.4	223	269	75	16	280	112	25
1.5	190	266	88	22	270	141	23
1.7	177	245	85	23	283	150	37
1.8	135	237	102	27	272	173	54
1.9	101	231	96	31	294	193	54
2.0	91	192	126	46	282	203	60
2.1	97	178	109	35	311	198	72
2.3	62	178	125	56	310	194	75
2.4	74	165	86	65	290	226	94
2.5	46	135	111	73	303	242	90
2.6	37	120	103	75	321	250	94
2.7	29	121	95	100	292	252	111
2.9	31	94	93	87	312	244	139
3.0	27	91	85	118	295	260	124
3.1	21	84	73	91	305	292	134
3.2	10	73	79	112	314	259	153
3.4	15	61	73	120	305	271	155
3.5	13	48	67	127	315	274	156

Table A.2 is continued from the previous page.

t	stage1	stage2	stage3	stage4	death stage1	death stage2	death stage3
3.6	13	45	67	118	291	277	189
3.7	5	47	59	134	327	268	160
3.8	10	35	59	123	295	288	190
4.0	7	29	37	141	309	301	176
4.1	2	33	29	156	300	293	187
4.2	4	27	40	143	296	301	189
4.3	1	16	33	150	313	287	200
4.4	1	18	21	155	320	304	181
4.6	3	16	19	157	307	310	188
4.7	1	14	21	152	304	319	189
4.8	1	13	20	155	307	294	210
4.9	0	12	13	166	300	304	205
5.0	1	9	16	186	297	305	186
5.2	0	9	15	162	321	293	200
5.3	2	7	7	160	272	333	219
5.4	1	3	8	199	302	298	189
5.5	1	4	8	171	308	308	200
5.6	0	4	5	160	303	312	216
5.8	0	2	7	168	299	320	204
5.9	0	2	10	158	288	308	234
6.0	0	0	3	155	323	306	213

A.3 Simulated data for the linear time-dependent hazard rate model

Table A.3: The distribution of 100 sampled individuals over the four stages at 15 sampling time points in linear time-dependent hazard rates case. The sample of 100 individuals are not the same at different time points. Stage 4 is the final stage.

t	stage1	stage2	stage3	stage4	death stage2	death stage3
0.1	99	1	0	0	0	0
0.5	79	18	1	0	2	0
0.9	59	29	5	4	3	0
1.4	48	36	8	5	3	0
1.8	32	38	8	6	15	1
2.2	13	39	13	14	18	3
2.6	8	28	14	19	28	3
3.0	5	19	9	34	24	9
3.5	4	9	6	44	30	7
3.9	5	10	3	46	27	9
4.3	3	9	5	47	26	10
4.7	1	4	1	55	28	11
5.2	0	2	1	49	35	13
5.6	1	1	1	52	30	15
6	0	1	2	55	28	14

B. Traces, density and autocorrelation plots for parameter estimates with single MH algorithm.

B.1 Model with stage-wise constant hazard rates (Section 5.2).

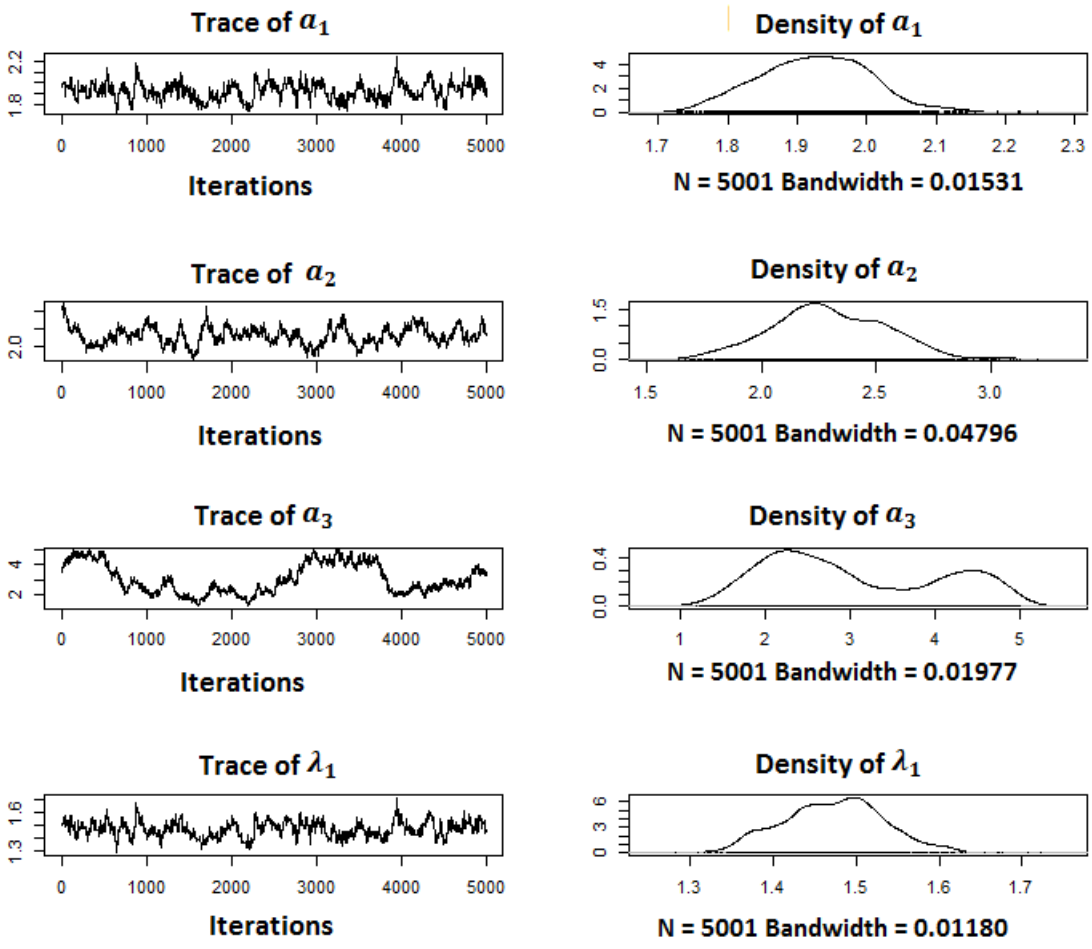

Figure B1.1: MCMC traces and density plots of the parameter estimates for stage-wise hazard rate model.

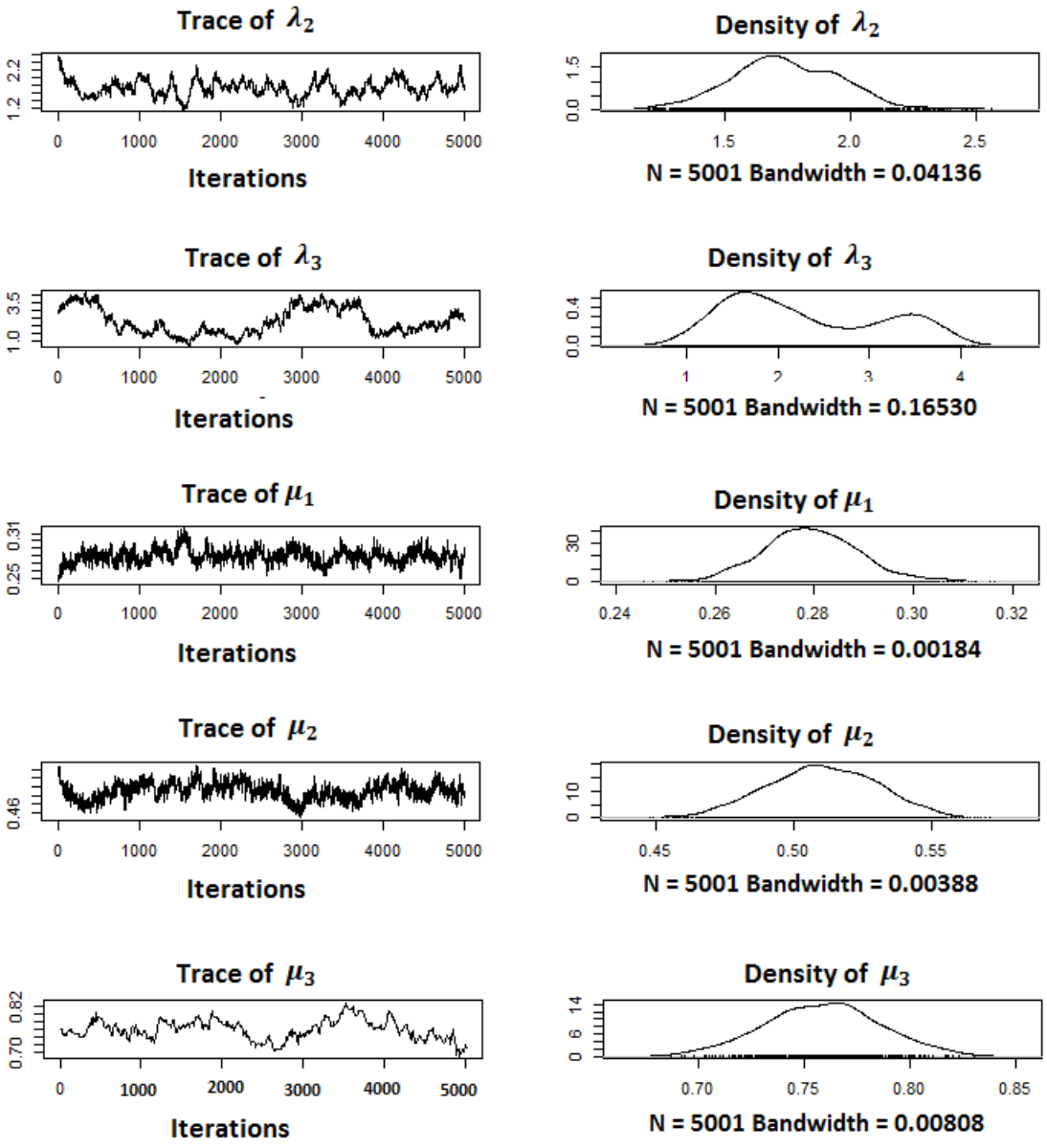

Figure B1.2: MCMC trace and density plots of the parameter estimates for stage-wise hazard rate model.

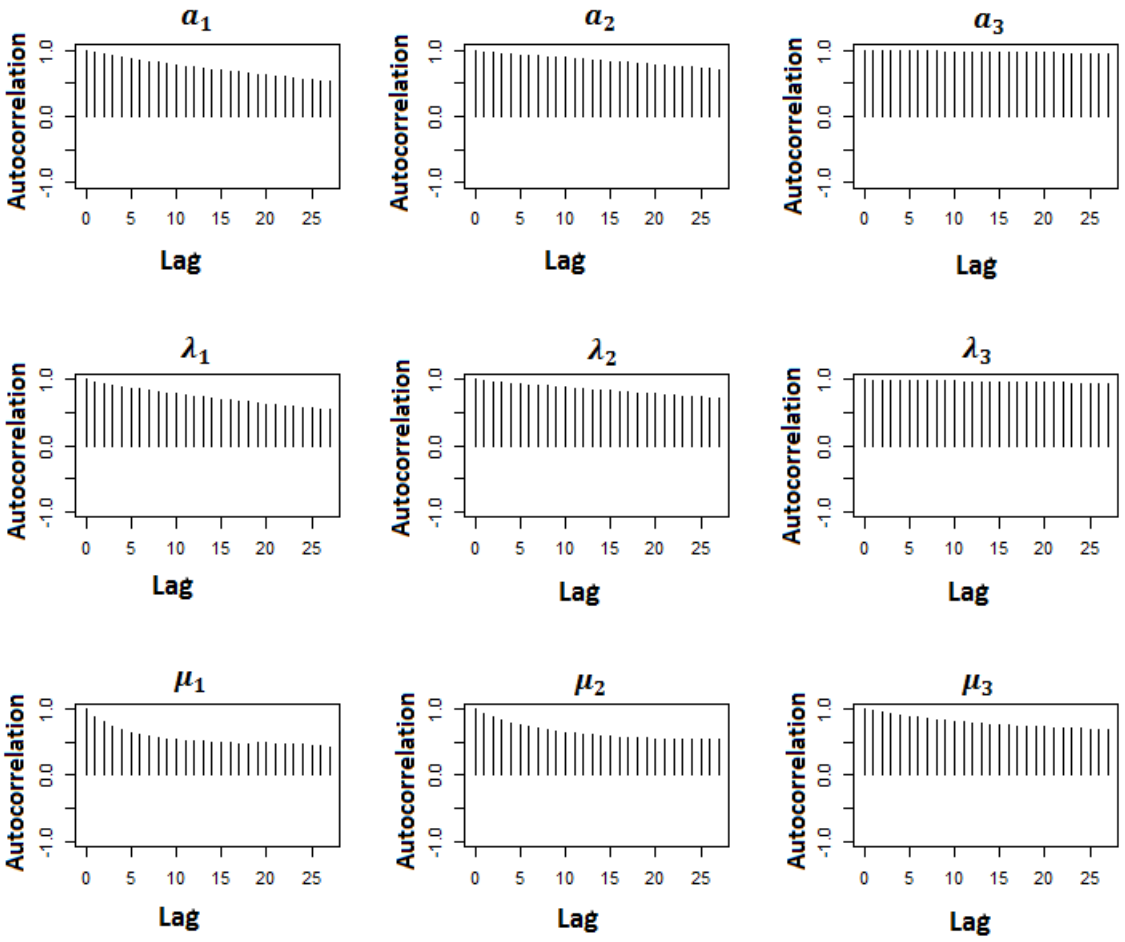

Figure B1.3: Autocorrelation plots of the parameter $(a_j, \lambda_j, \mu_j \, j = 1, 2, 3)$ estimates for stage-wise hazard rate model.

B.2 Model with time-dependent hazard rates (Section 5.3).

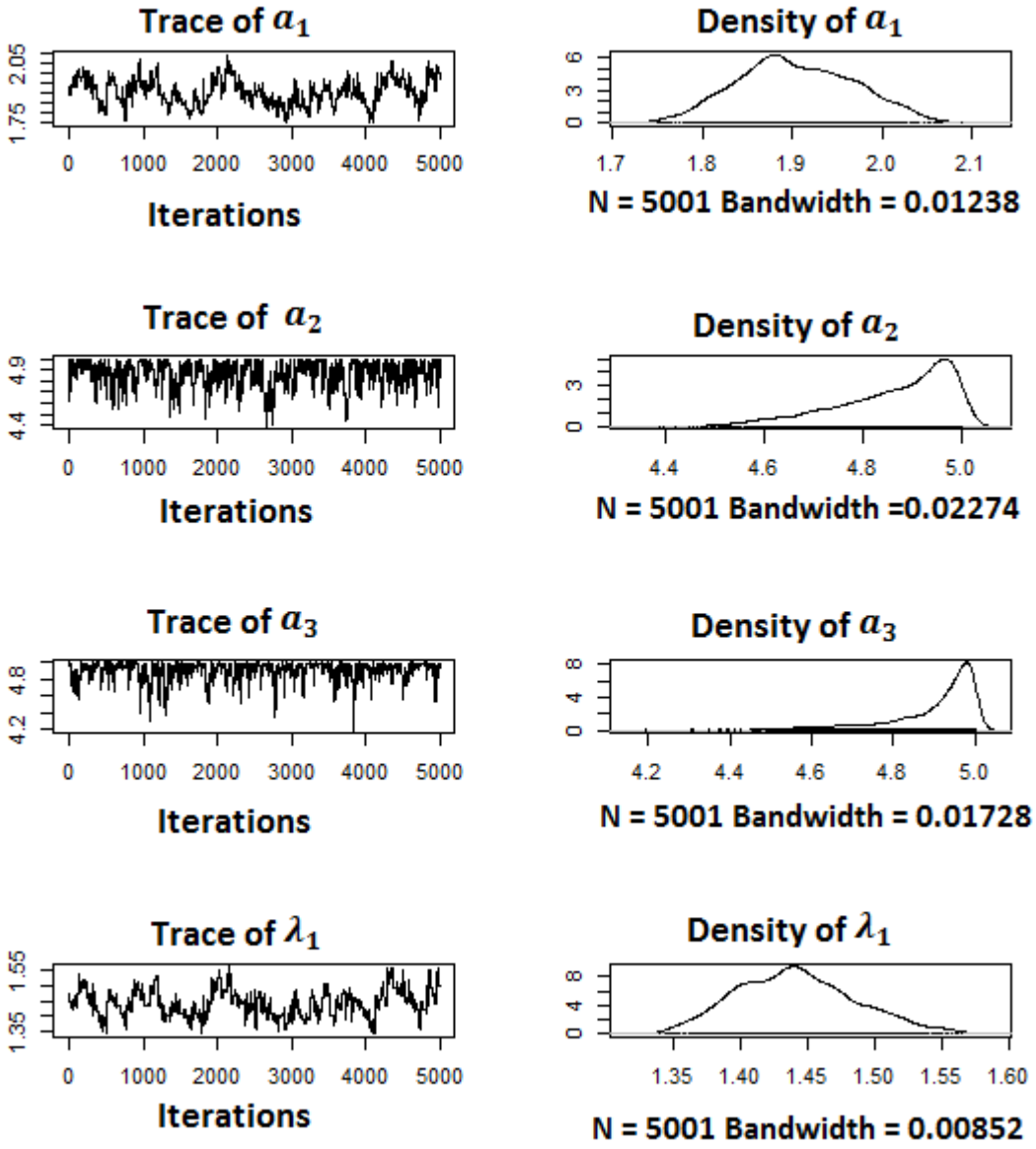

Figure B2.1: MCMC traces and density plots of the parameter estimates for the linear time-dependent hazard rate model.

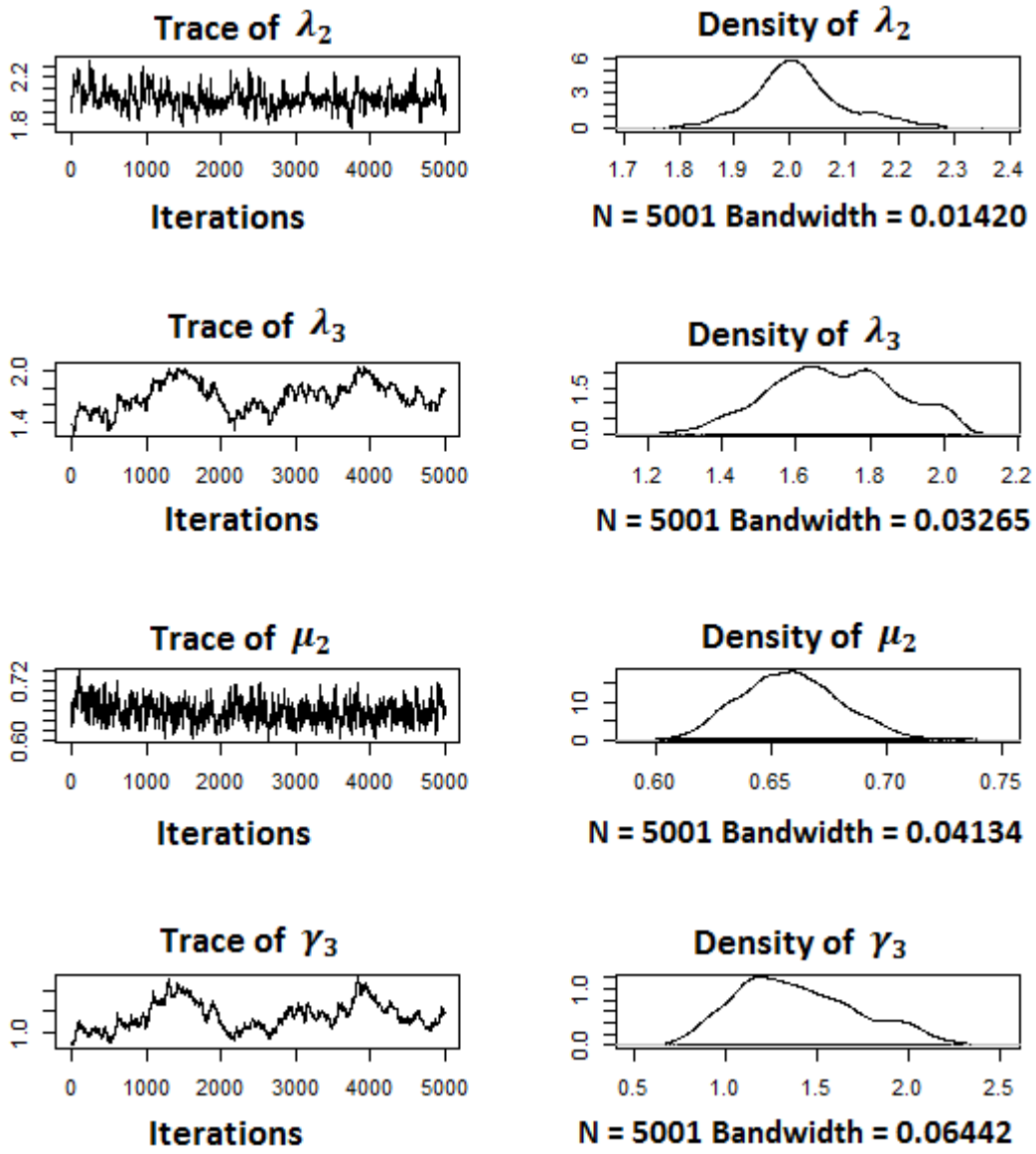

Figure B2.2: MCMC traces and density plots of the parameter estimates for the linear time-dependent hazard rate model.

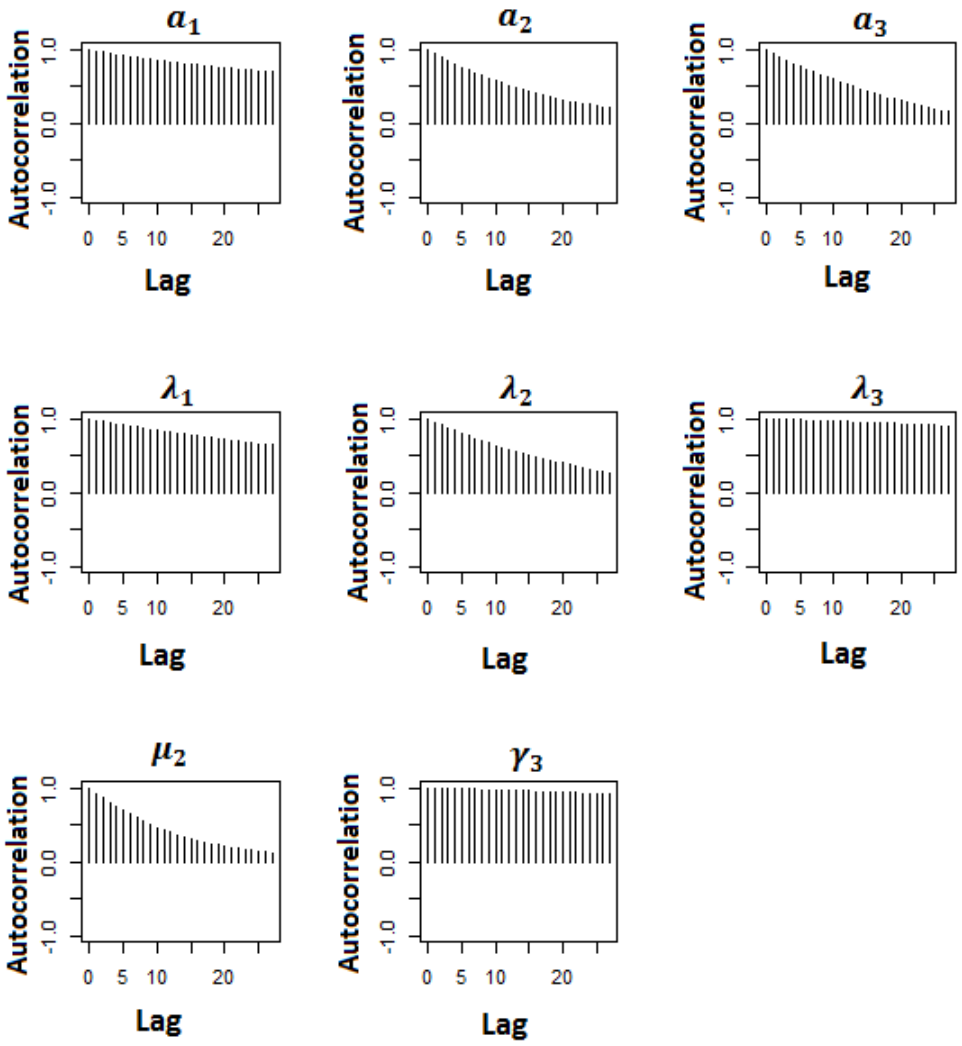

Figure B2.3: Autocorrelation plots of the parameter (a_j, λ_j, $j = 1, 2, 3, \mu_2$ and γ_3) estimates for the linear time-dependent hazard rate model.

B.3 Data for cattle parasitic nematode (Section 6.1).

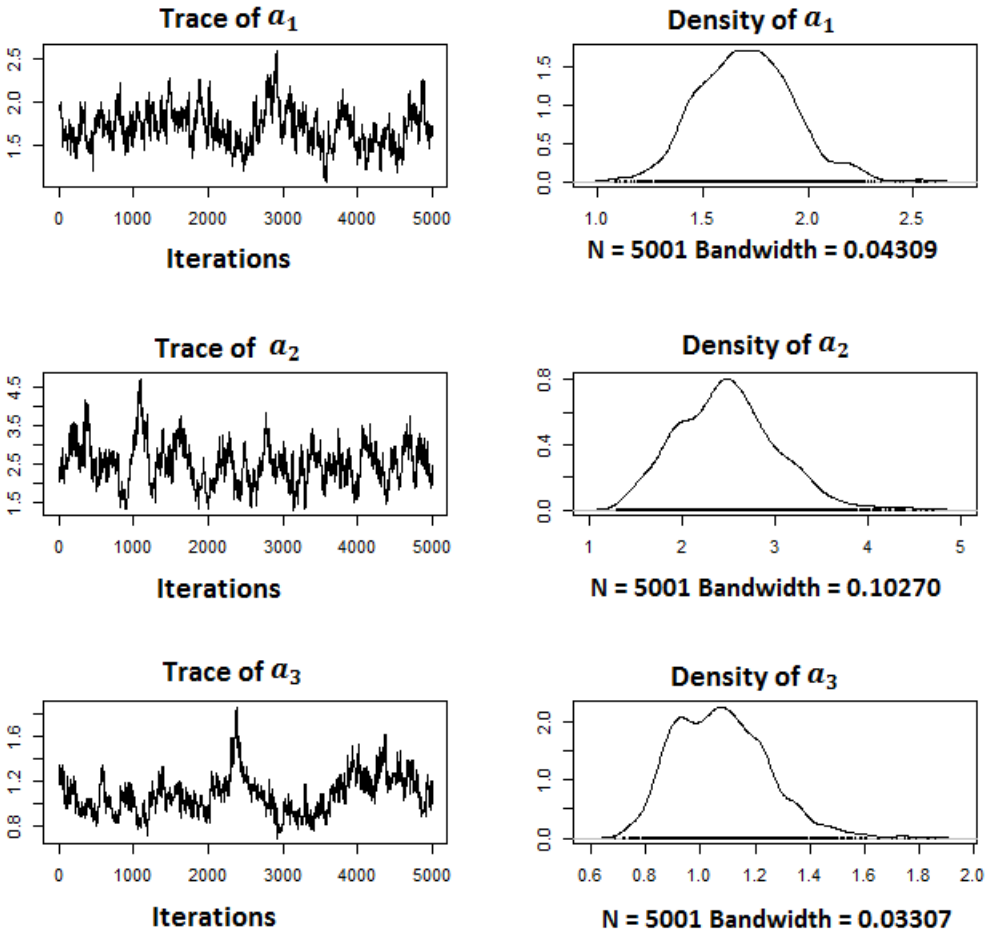

Figure B3.1: MCMC trace and density plots of the shape parameter (a_j, $j = 1, 2, 3$) estimates for parasitic nematode data.

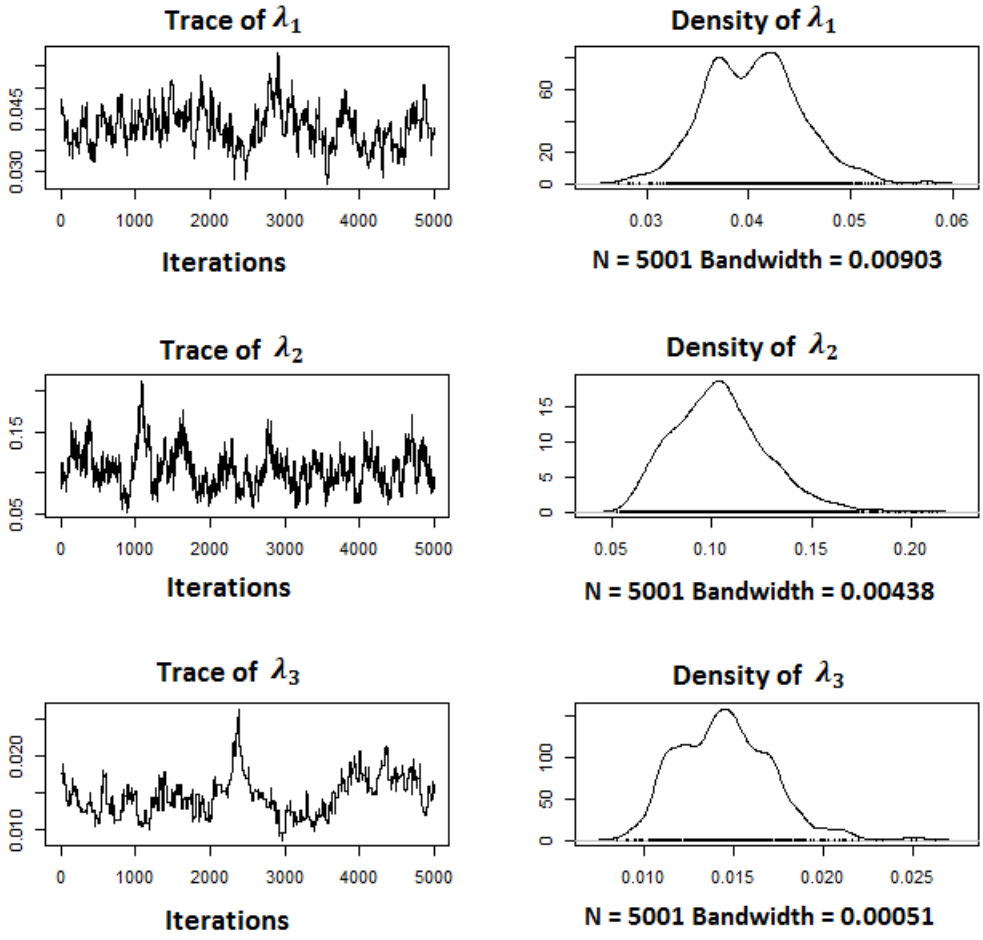

Figure B3.2: MCMC trace and density plots of the rate parameter (λ_j, $j = 1, 2, 3$) estimates for parasitic nematode data.

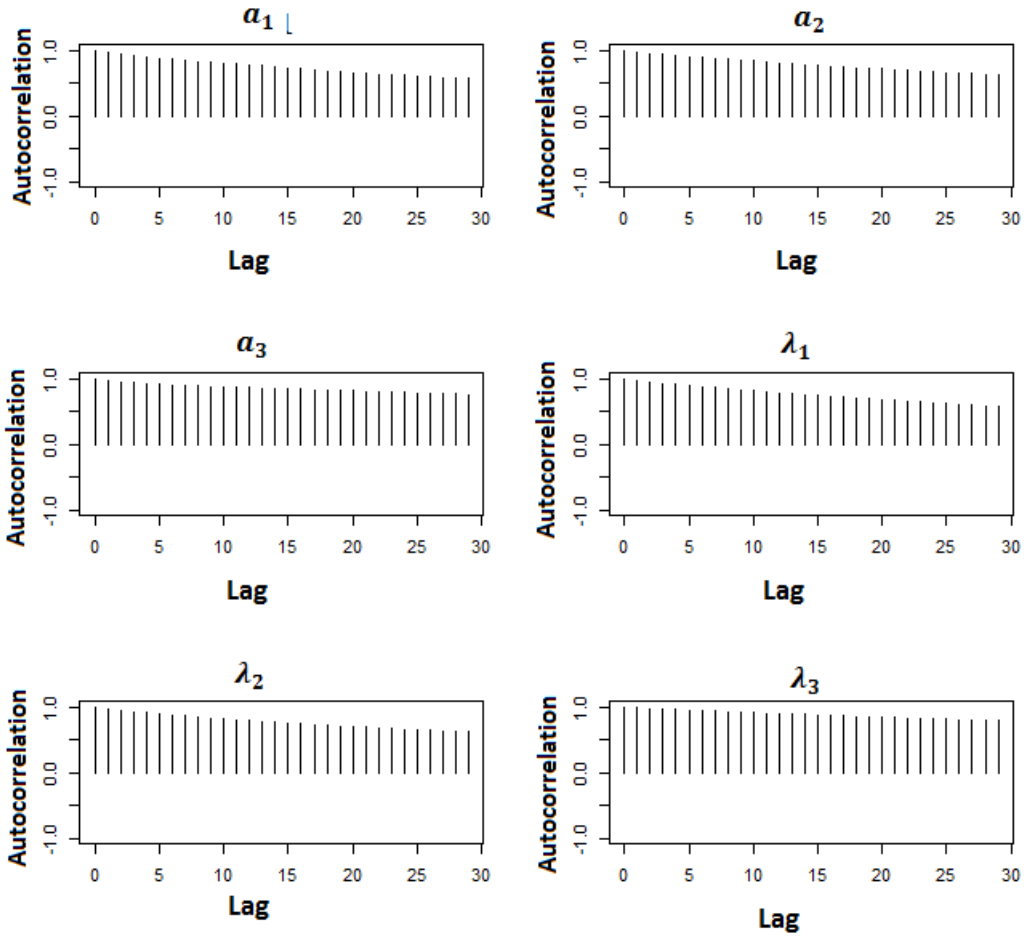

Figure B3.3: Autocorrelation plots of the maturation parameter $(a_j, \lambda_j,\ j = 1, 2, 3)$ estimates for parasitic nematode data.

B.4 Data for breast development of New Zealander schoolgirls (Section 6.2).

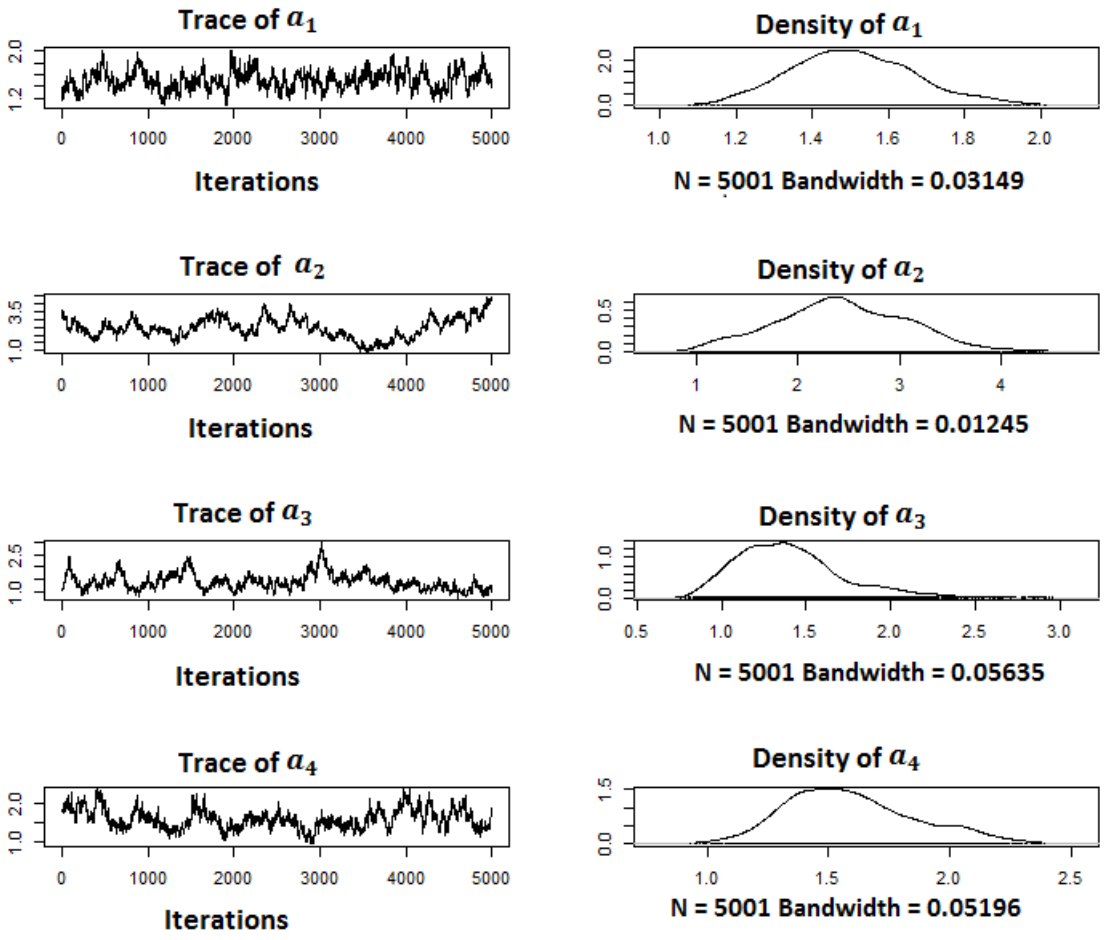

Figure B4.1: MCMC trace and density plots of the shape parameter (a_j, $j = 1, 2, 3, 4$) estimates for breast development of New Zealander schoolgirls data.

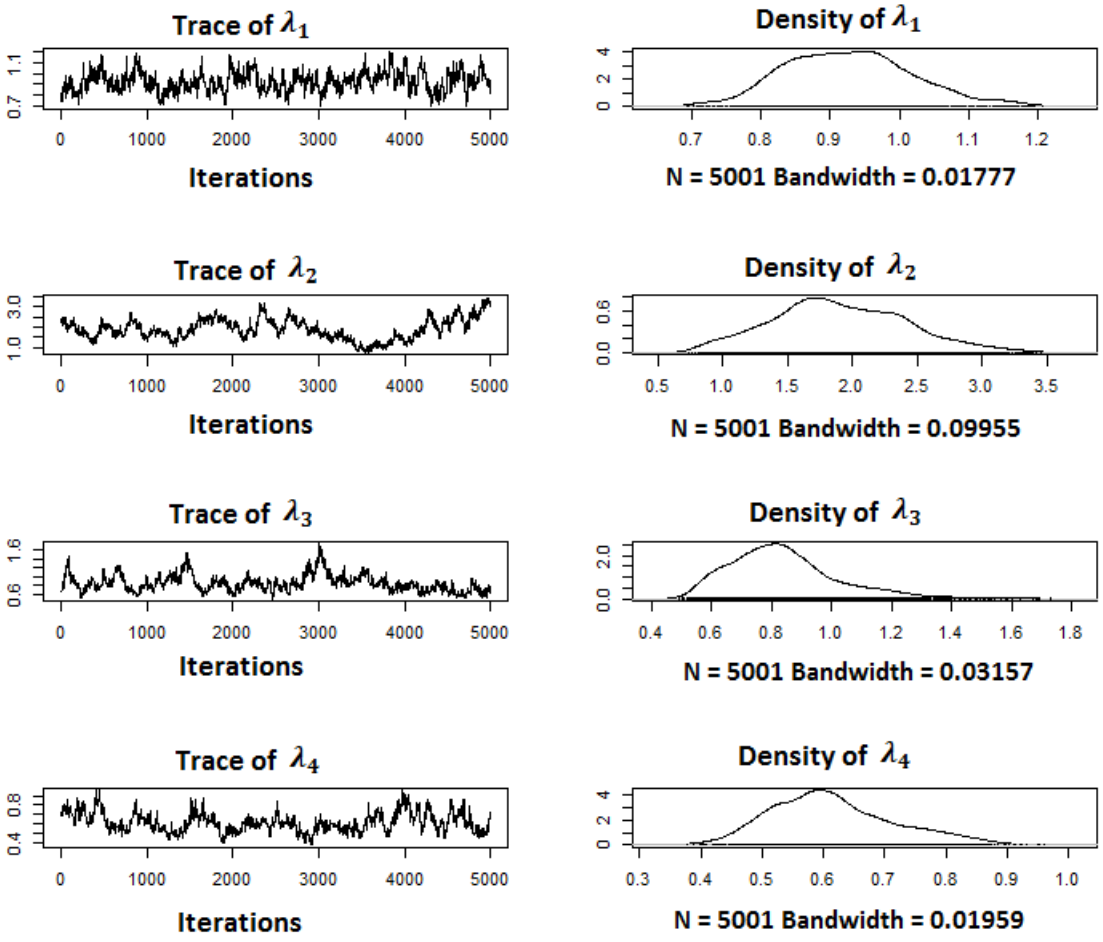

Figure B4.2: MCMC trace and density plots of the rate parameter (λ_j, $j = 1, 2, 3, 4$) estimates for breast development of New Zealander schoolgirls data.

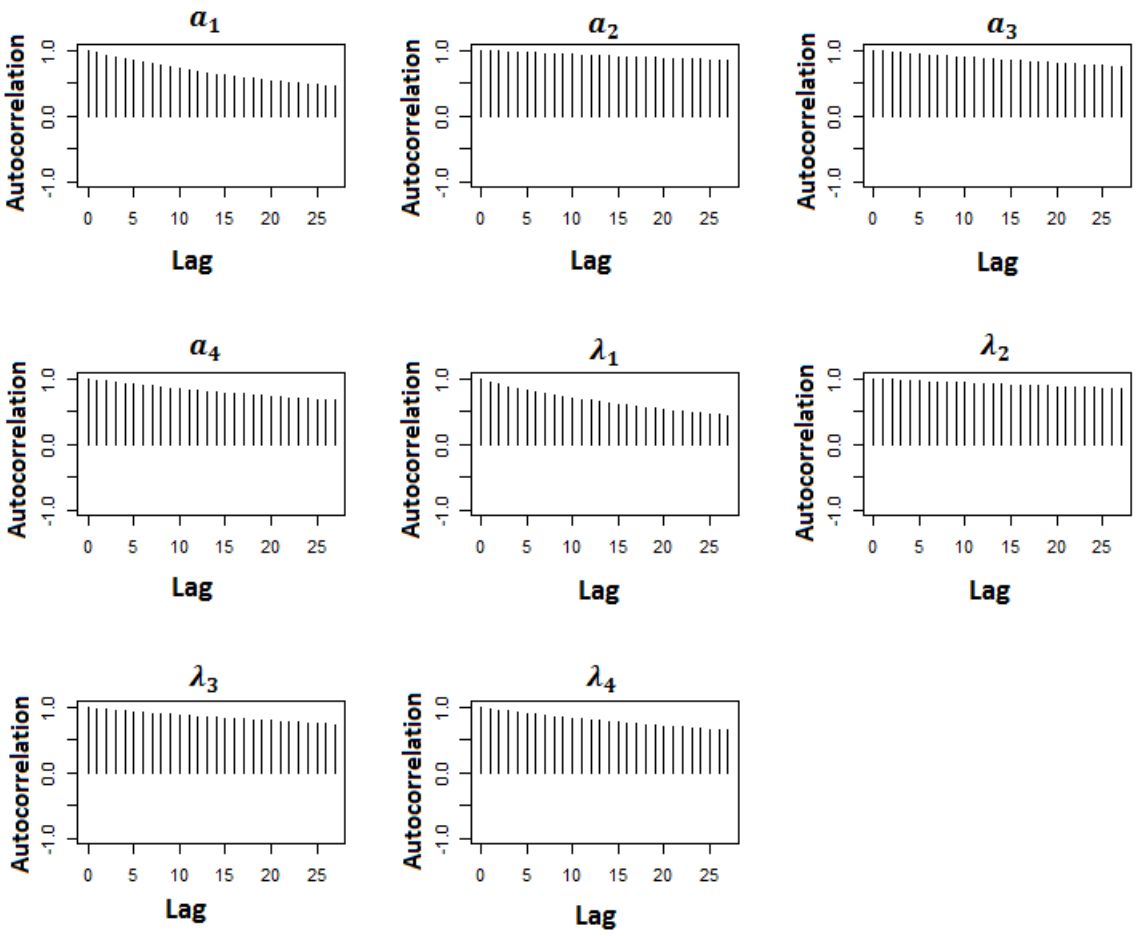

Figure B4.3: Autocorrelation plots of the maturation parameter $(a_j, \lambda_j,\ j = 1, 2, 3, 4)$ estimates for breast development of New Zealander schoolgirls data.

C. Plots from Gelman and Rubin diagnostic tests.

C.1 Model with stage-wise constant hazard rates (Section 5.2).

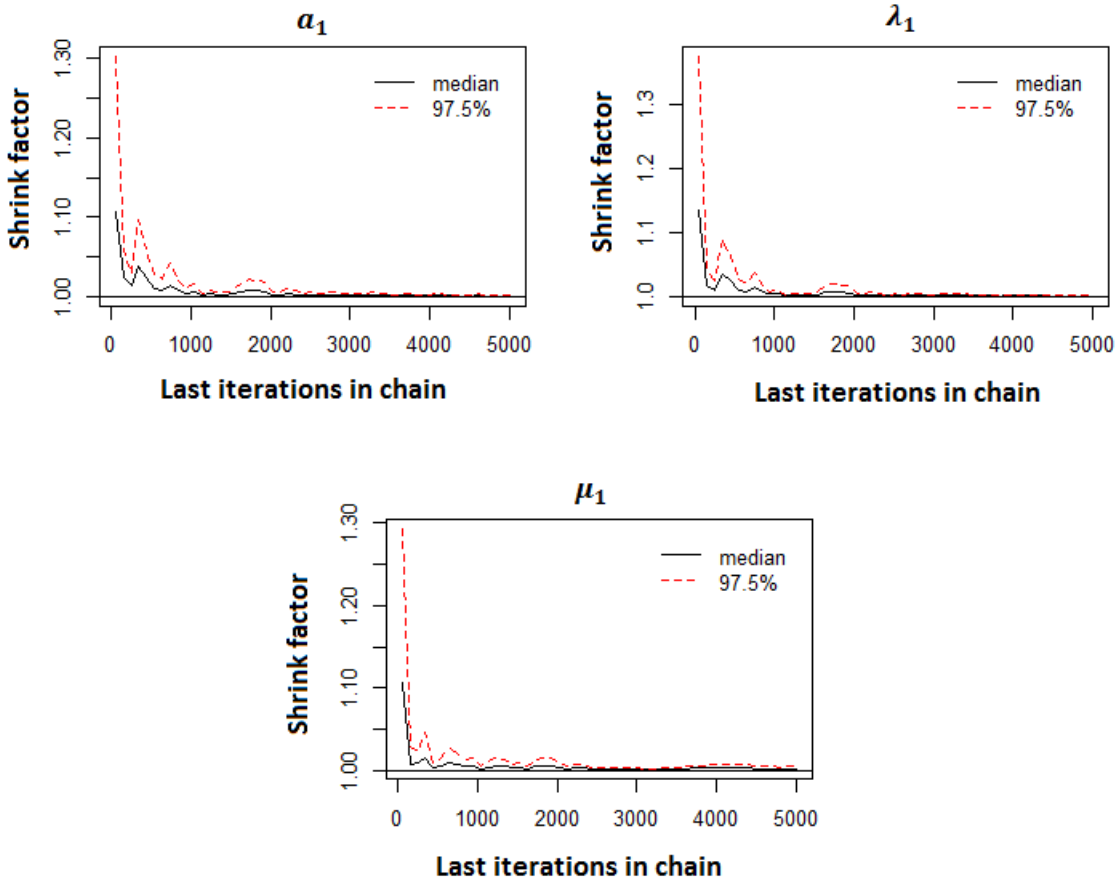

Figure C1.1: The potential rate reduction factor plots from Gelman and Rubin diagnostic of parameter (a_1, λ_1, μ_1) estimates at stage 1 for stage-wise constant hazard rate model for five Markov chains of length 10,000 iterations.

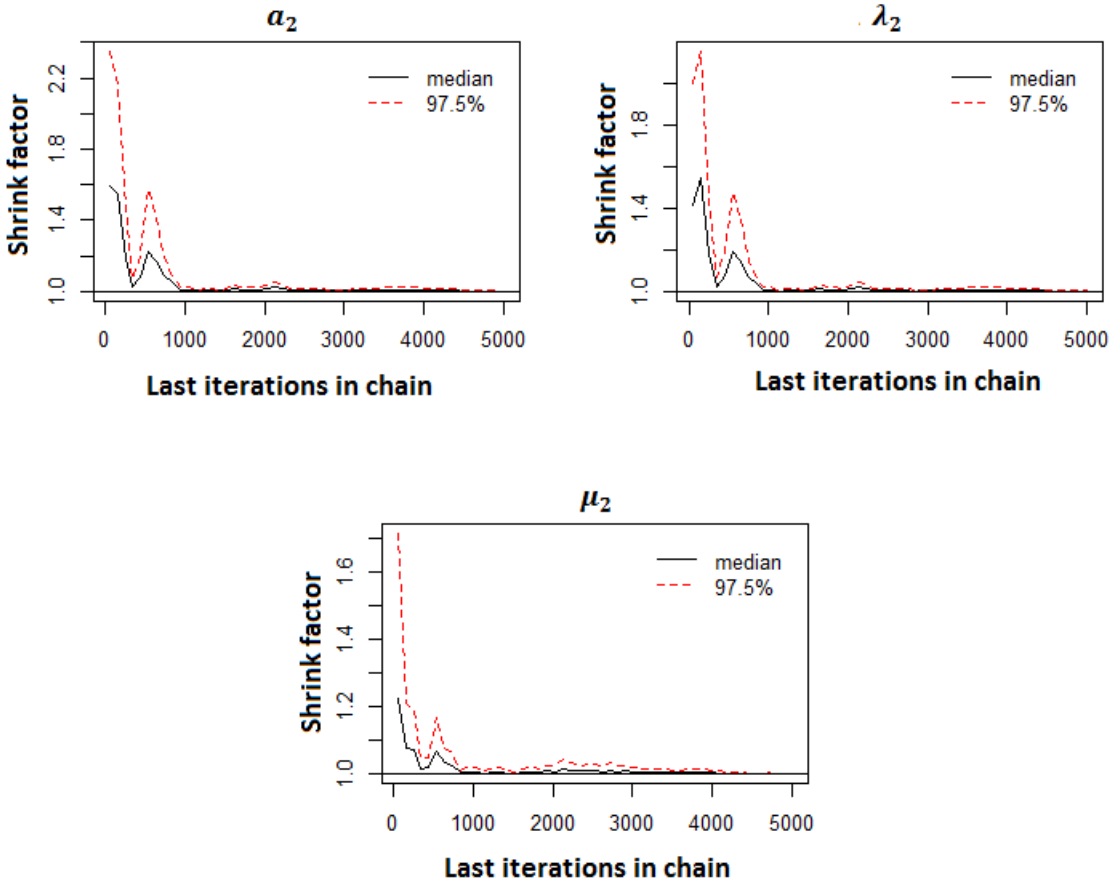

Figure C1.2: The potential rate reduction factor plots from Gelman and Rubin diagnostic of parameter (a_2, λ_2, μ_2) estimates at stage 2 for stage-wise constant hazard rate model for five Markov chains of length 10,000 iterations.

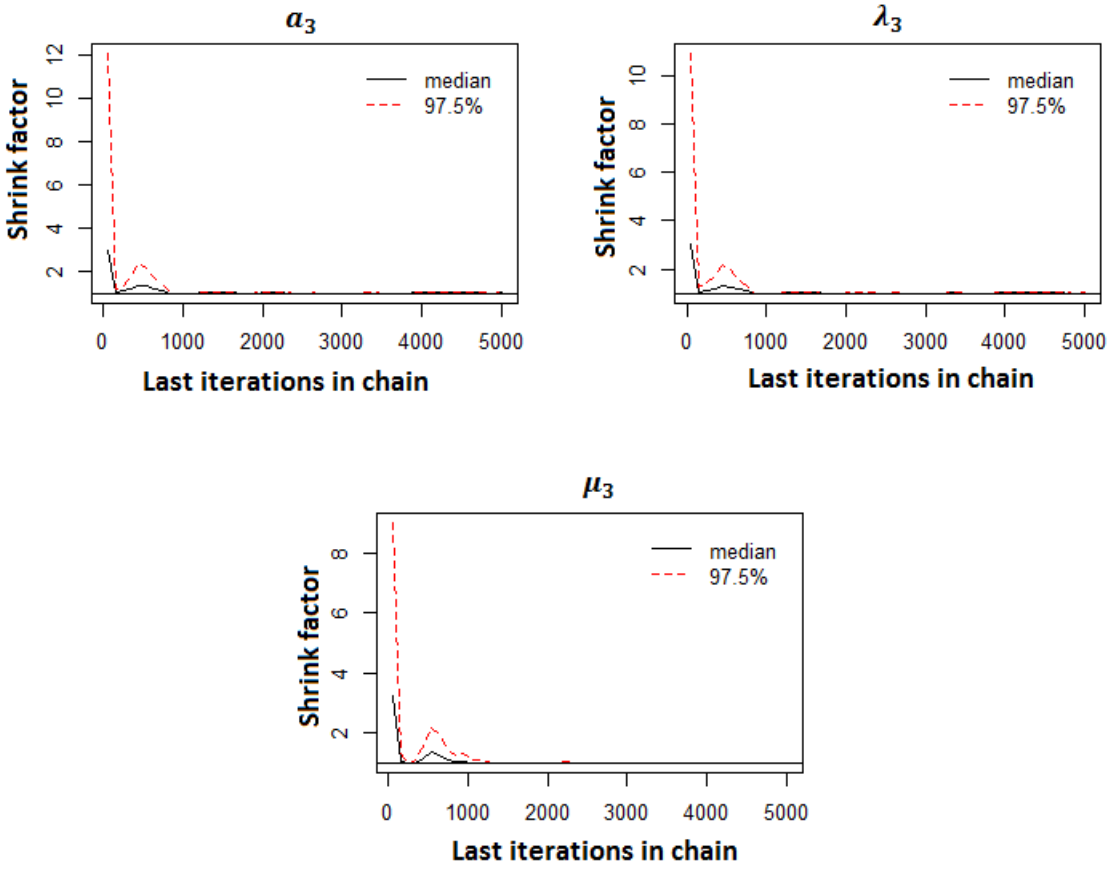

Figure C1.3: The potential rate reduction factor plots from Gelman and Rubin diagnostic of parameter (a_3, λ_3, μ_3) estimates at stage 3 for stage-wise constant hazard rate model for five Markov chains of length 10,000 iterations.

C.2 Model with linear time-dependent hazard rates (Section 5.3).

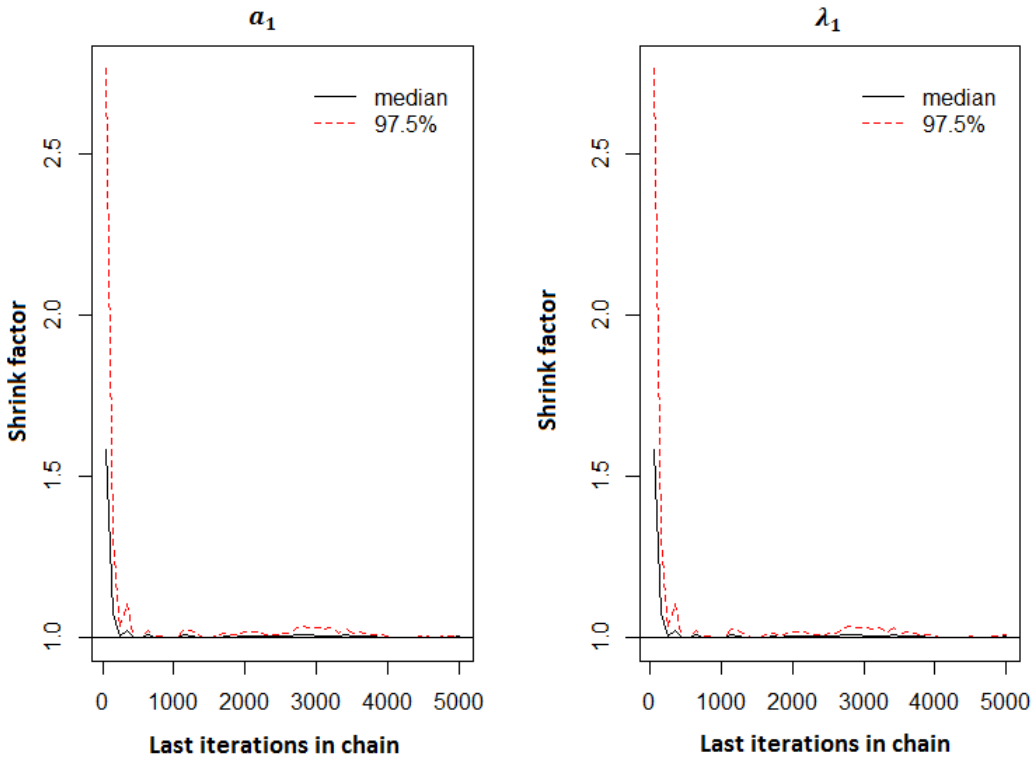

Figure C2.1: The potential rate reduction factor plots from Gelman and Rubin diagnostic of shape and rate (a_1, λ_1) estimates at stage 1 for the linear time-dependent hazard rate model for five Markov chains of length 10,000 iterations.

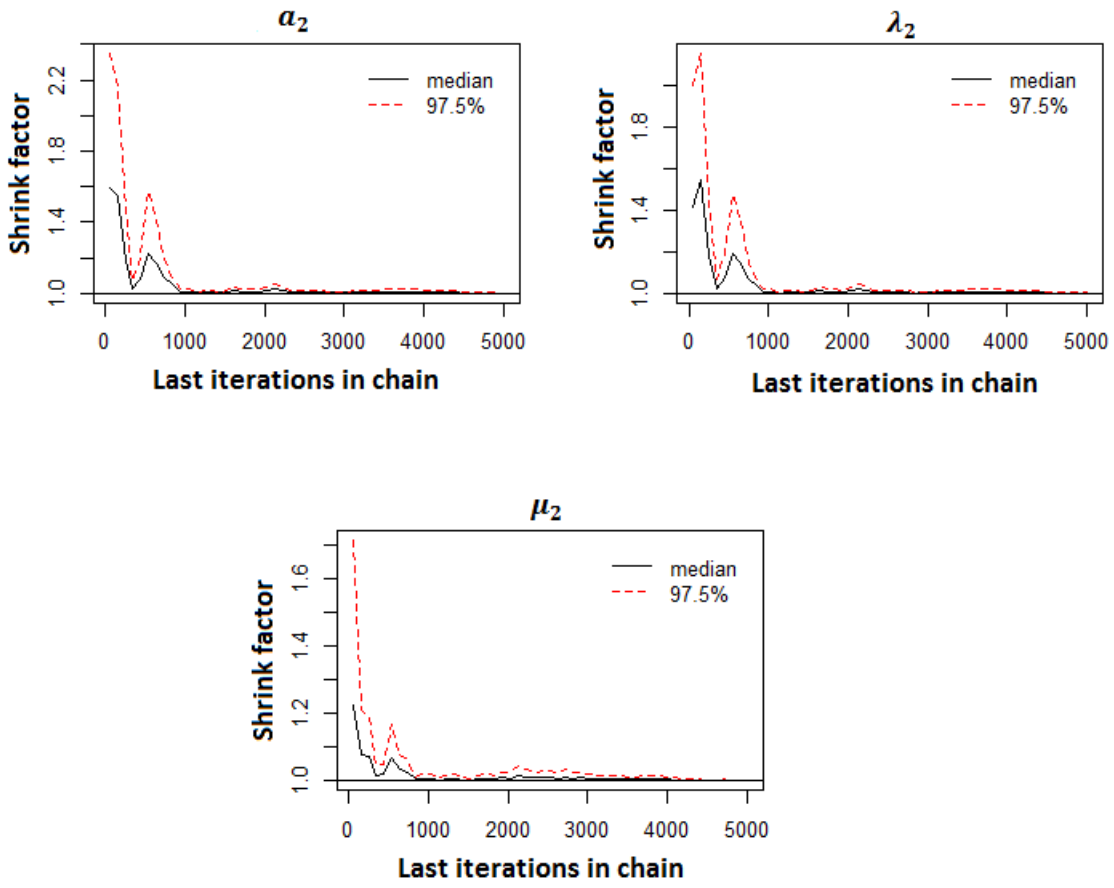

Figure C2.2: The potential rate reduction factor plots from Gelman and Rubin diagnostic of shape, rate and hazard rate (a_2, λ_2, μ_2) estimates at stage 2 for the linear time-dependent hazard rate model for five Markov chains of length 10,000 iterations.

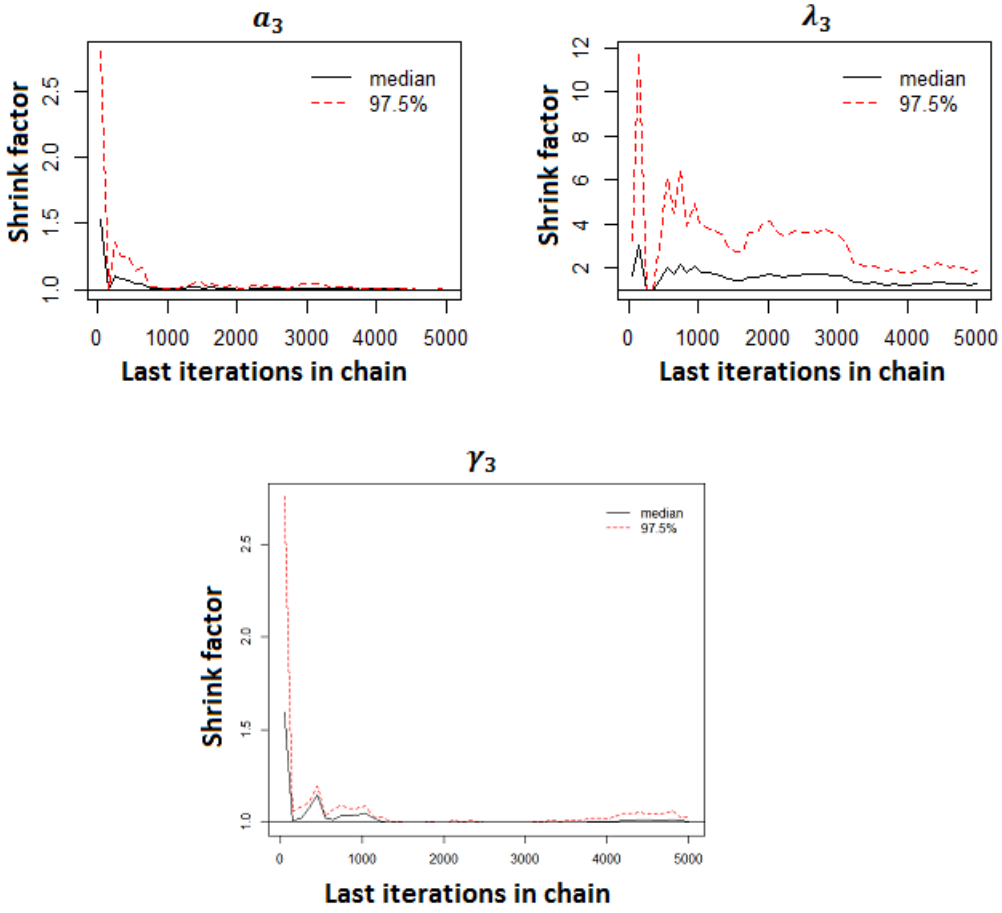

Figure C2.3: The potential rate reduction factor plots from Gelman and Rubin diagnostic of shape, rate and slope $(a_3, \lambda_3, \gamma_3)$ estimates at stage 3 for the linear time-dependent hazard rate model for five Markov chains of length 10,000 iterations.

C.3 Data for cattle parasitic nematode (Section 6.1).

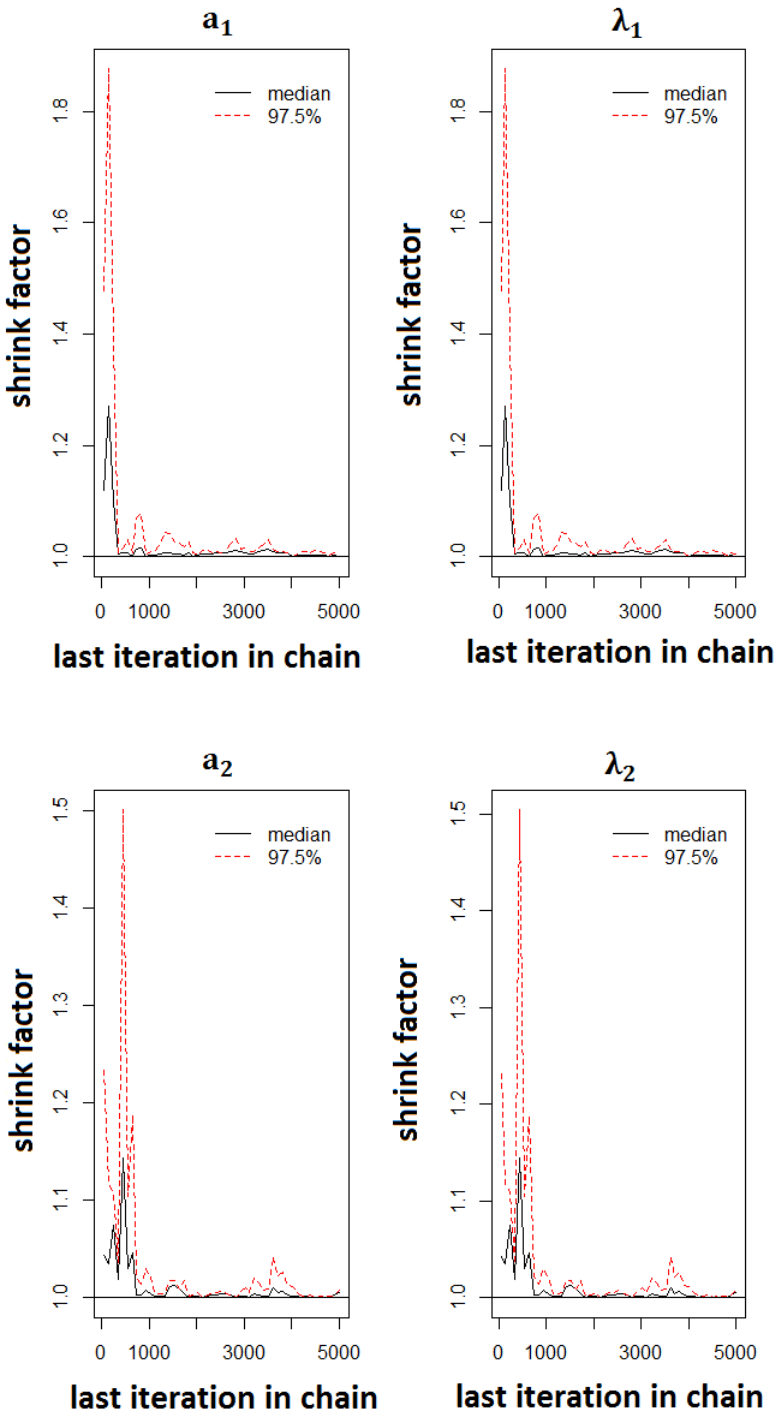

Figure C3.1: The potential rate reduction factor plots from Gelman and Rubin diagnostic of shape and rate $(a_j, \lambda_j,\ j = 1, 2)$ estimates at stage 1 and stage 2 for parasitic nematode data for five Markov chains of length 10,000 iterations.

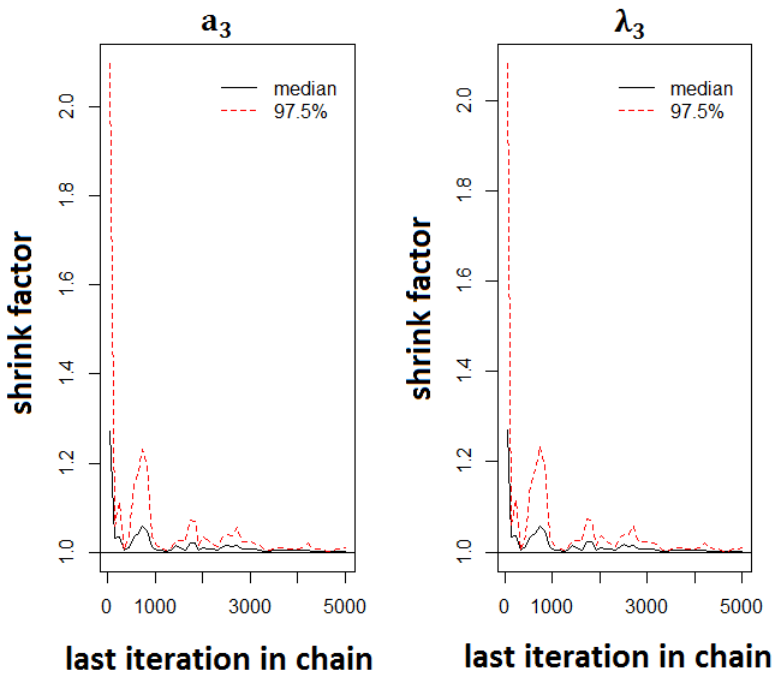

Figure C3.2: The potential rate reduction factor plots from Gelman and Rubin diagnostic of shape and rate (a_3, λ_3) estimates at stage 3 for parasitic nematode data for five Markov chains of length 10,000 iterations.

C.4 Data for breast development of New Zealander schoolgirls (Section 6.2).

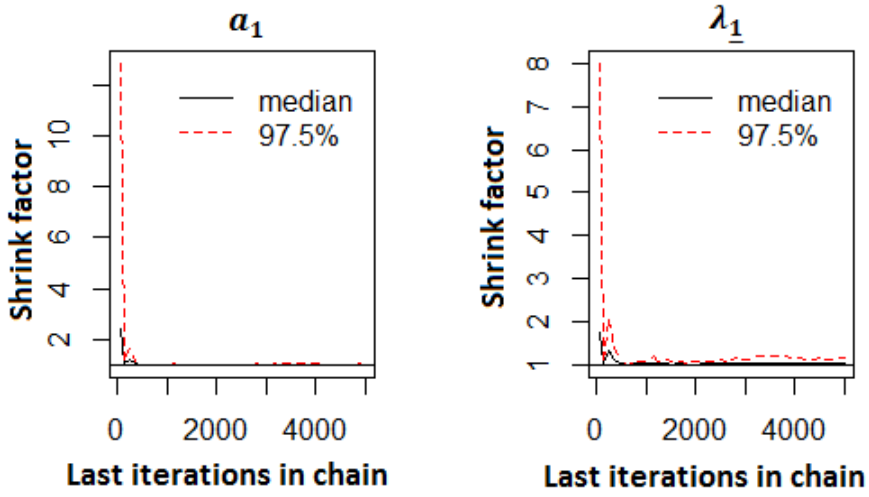

Figure C4.1: The potential rate reduction factor plots from Gelman and Rubin diagnostic of shape and rate (a_1, λ_1) estimates at stage 1 for breast development of New Zealander schoolgirls data for five Markov chains of length 10,000 iterations.

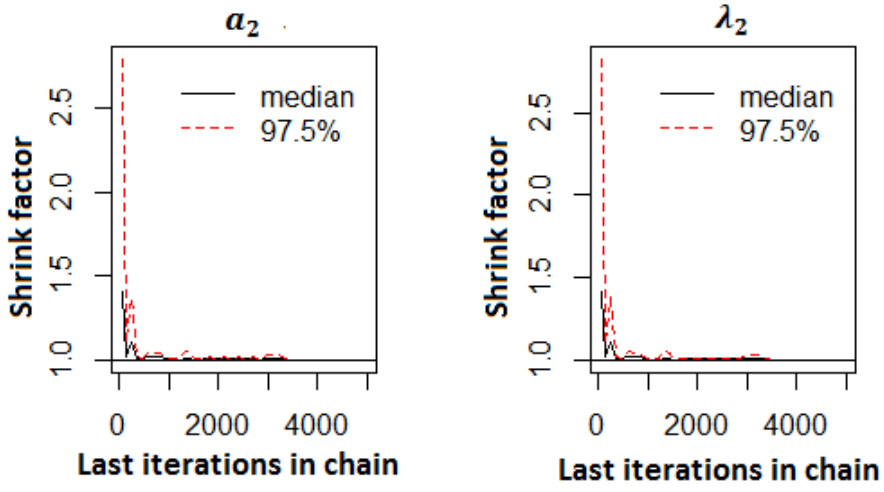

Figure C4.2: The potential rate reduction factor plots from Gelman and Rubin diagnostic of shape and rate (a_2, λ_2) estimates at stage 2 for breast development of New Zealander schoolgirls data for five Markov chains of length 10,000 iterations.

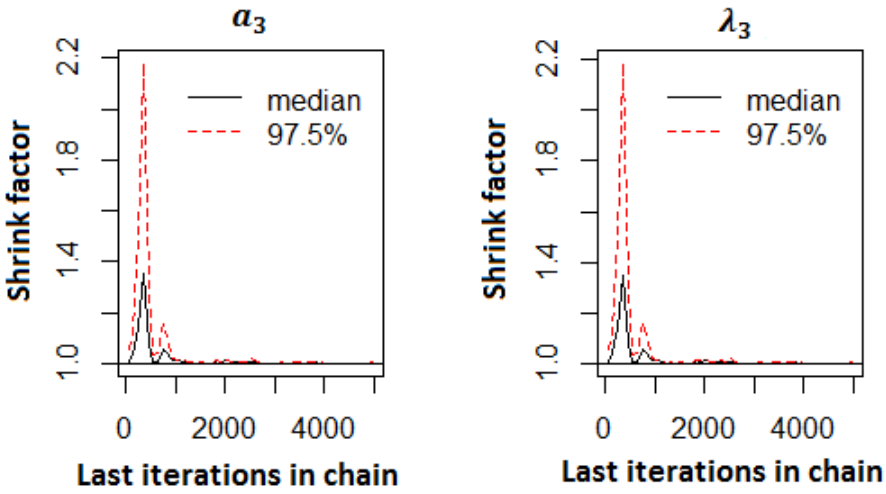

Figure C4.3: The potential rate reduction factor plots from Gelman and Rubin diagnostic of shape and rate (a_3, λ_3) estimates at stage 3 for breast development of New Zealander schoolgirls data for five Markov chains of length 10,000 iterations.

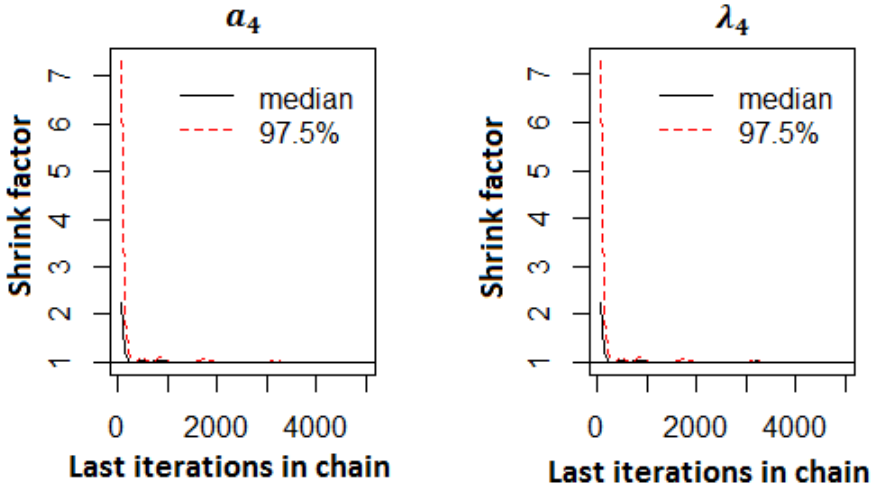

Figure C4.4: The potential rate reduction factor plots from Gelman and Rubin diagnostic of shape and rate (a_4, λ_4) estimates at stage 4 for breast development of New Zealander schoolgirls data for five Markov chains of length 10,000 iterations.

D. Trace plots, the distributions and autocorrelation plots with the MH algorithm based on deterministic transformations.

D.1 Model with stage-wise constant hazard rates (Section 5.2).

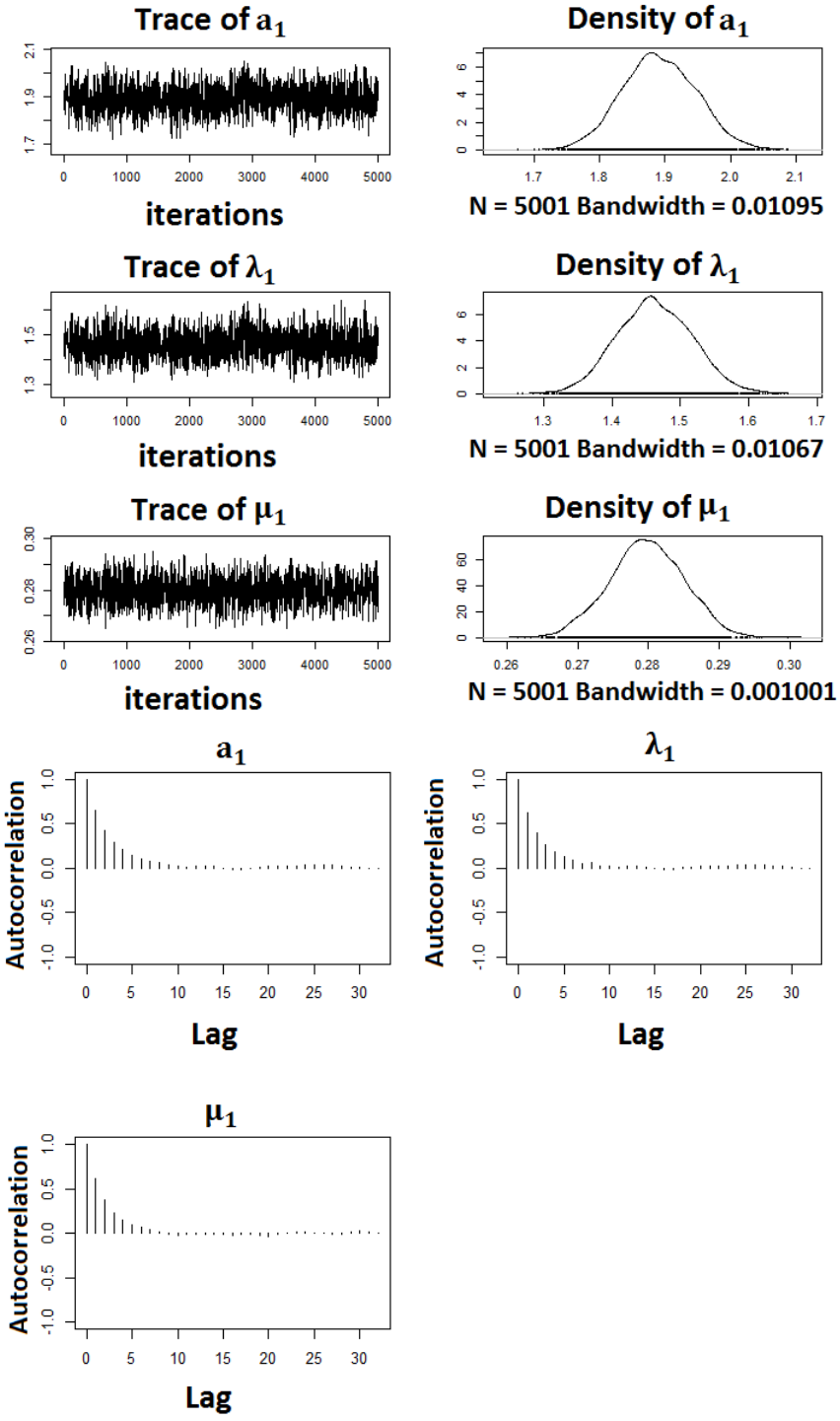

Figure D1.1: MCMC trace plots, density and autocorrelation plots of shape, rate and hazard rate (a_1, λ_1, μ_1) estimates at stage 1 for stage-wise constant hazard rate model.

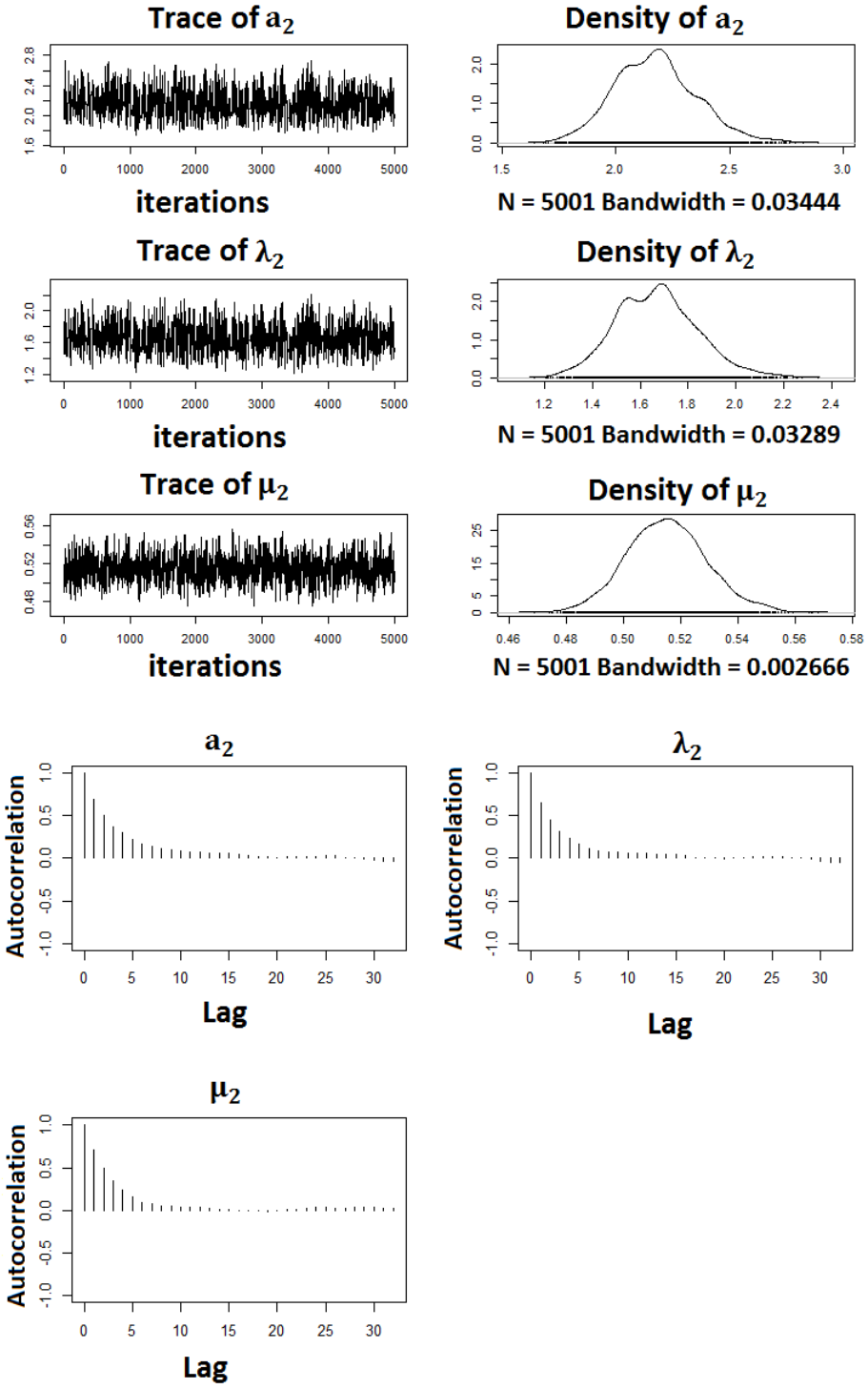

Figure D1.2: MCMC trace plots, density and autocorrelation plots of shape, rate and hazard rate (a_2, λ_2, μ_2) estimates at stage 2 for stage-wise constant hazard rate model.

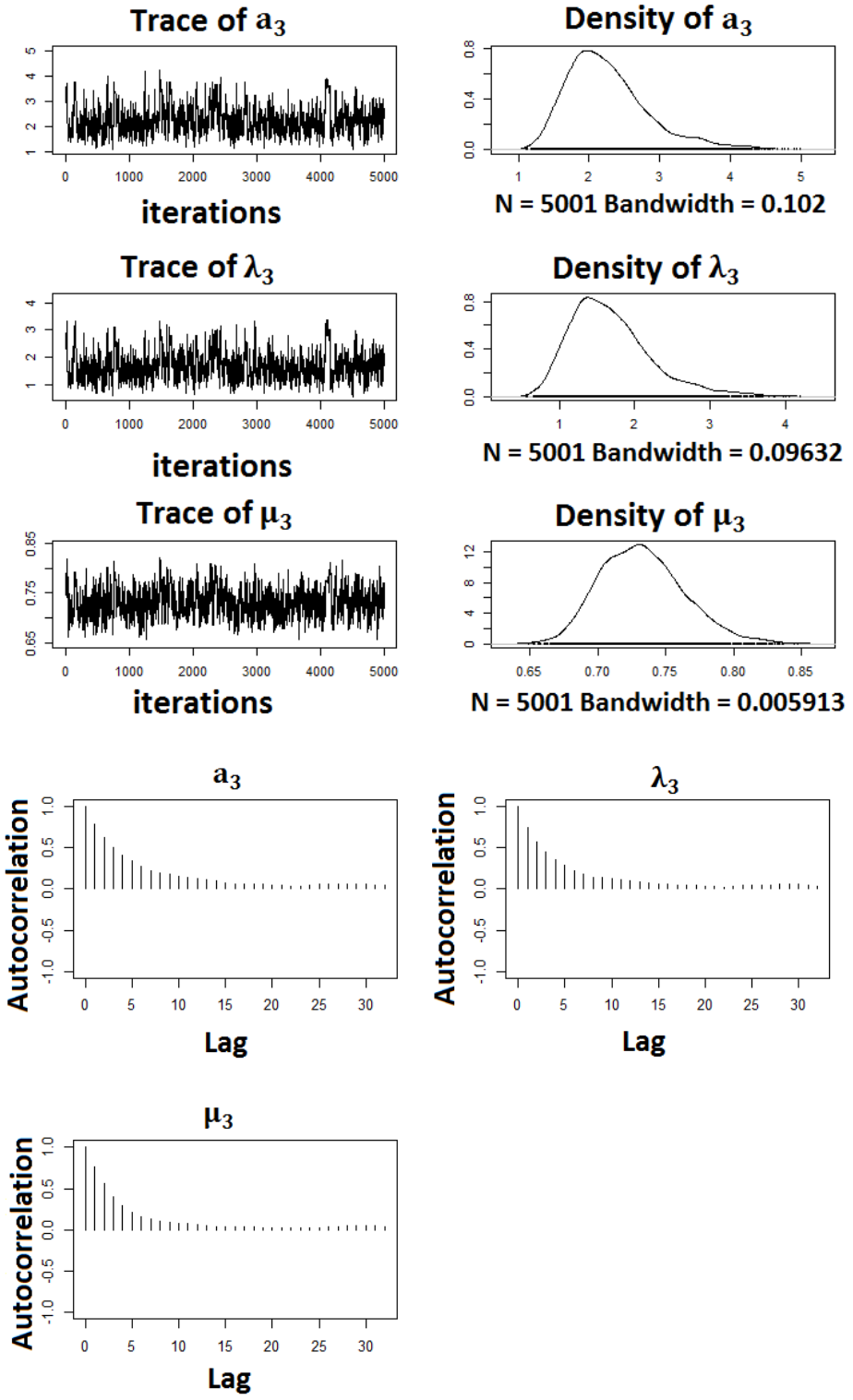

Figure D1.3: MCMC trace plots, density and autocorrelation plots of shape, rate and hazard rate (a_3, λ_3, μ_3) estimates at stage 3 for stage-wise constant hazard rate model.

D.2 Model with linear time-dependent hazard rates (Section 5.3).

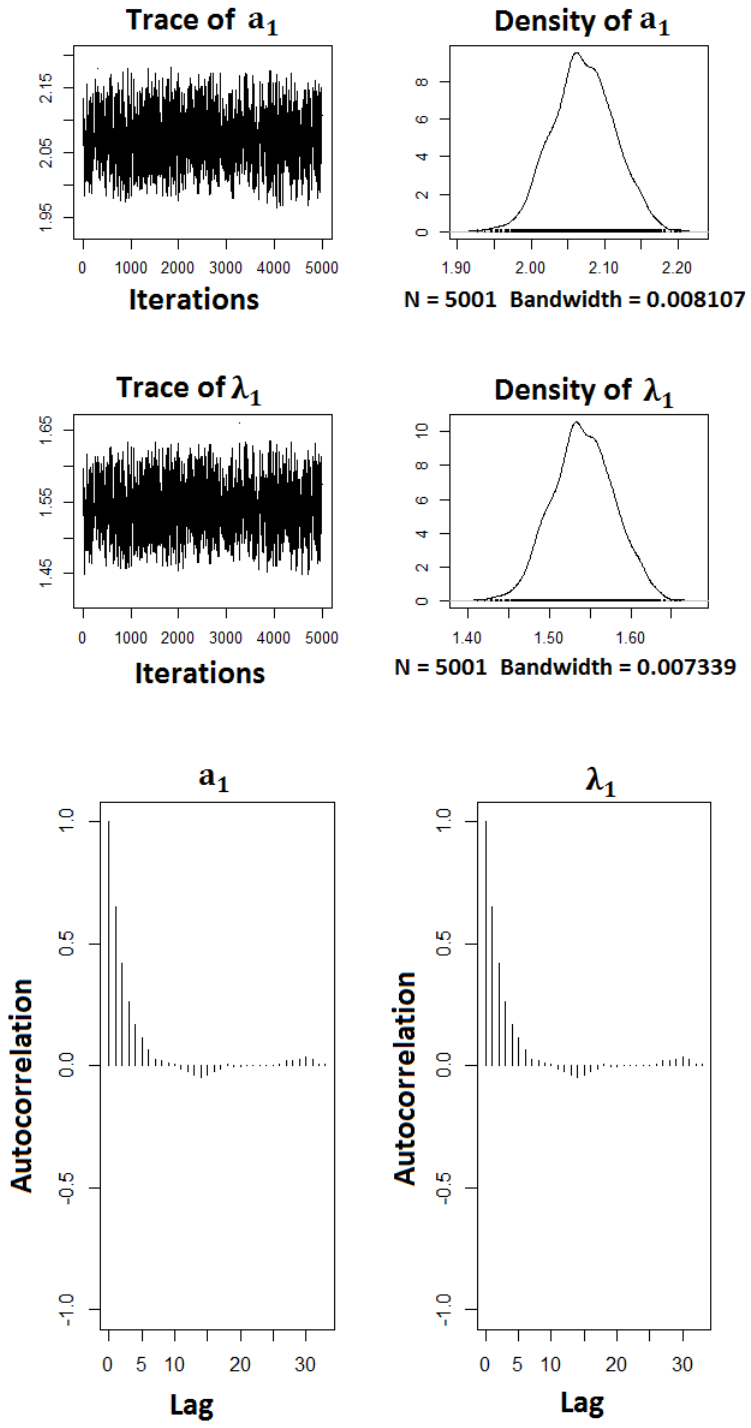

Figure D2.1: MCMC trace and density plots of shape and rate (a_1, λ_1) estimates at stage 1 for the linear time-dependent hazard rates model.

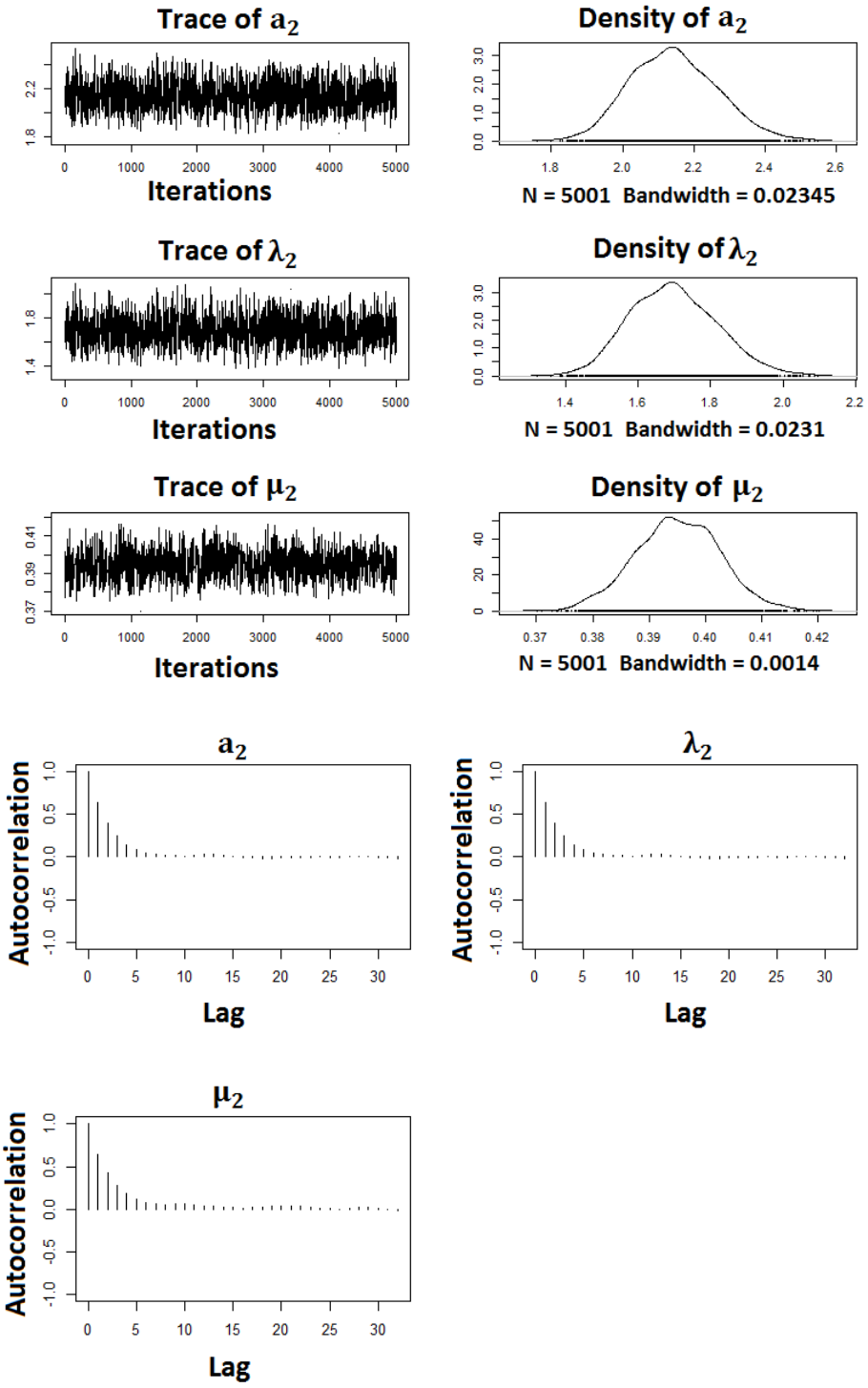

Figure D2.2: MCMC trace and density plots of shape, rate and hazard rate (a_2, λ_2, μ_2) estimates at stage 2 for the linear time-dependent hazard rates model.

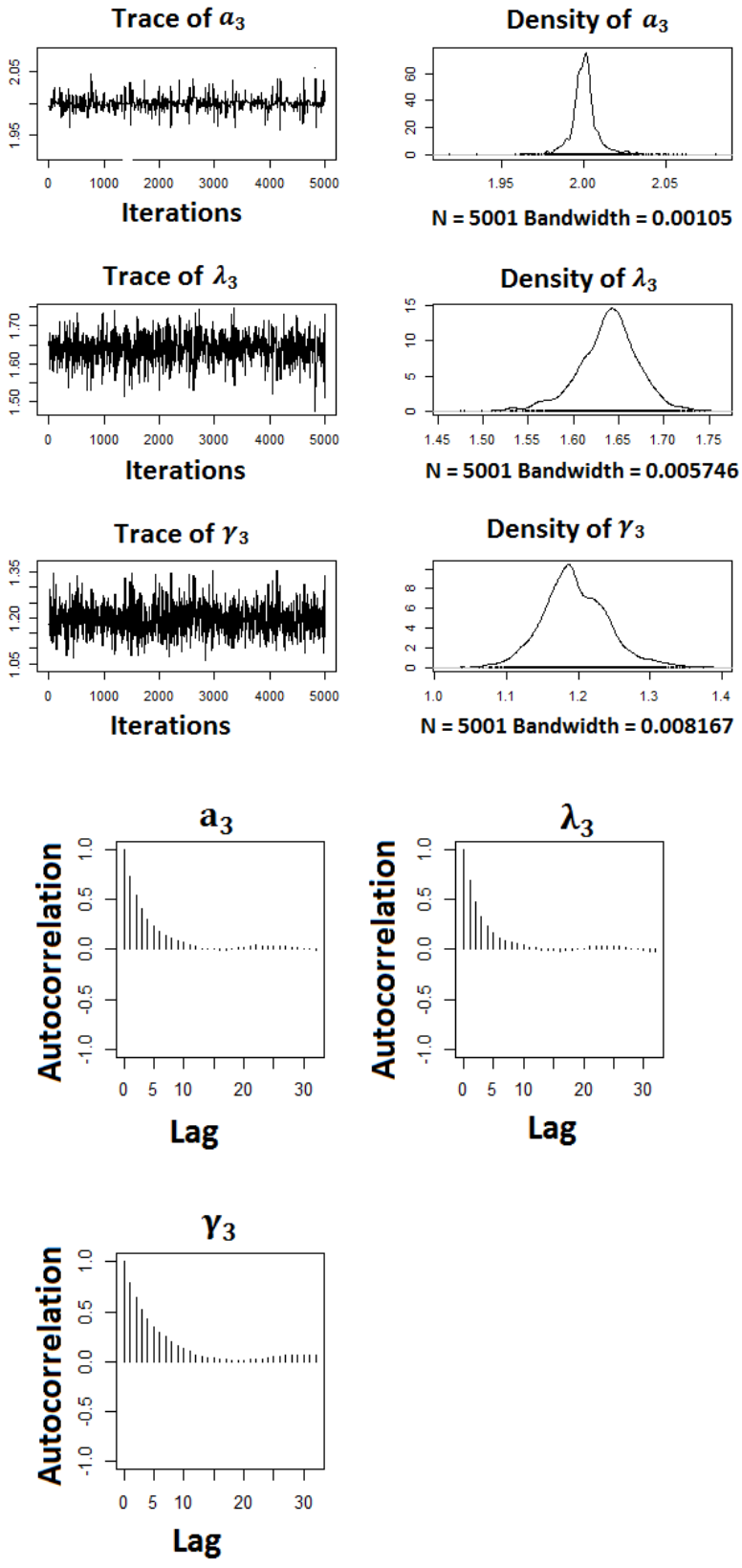

Figure D2.3: MCMC trace and density plots of shape, rate and slope $(a_3, \lambda_3, \gamma_3)$ estimates at stage 3 for the linear time-dependent hazard rates model.

D.3 Data for cattle parasitic nematode (Section 6.1).

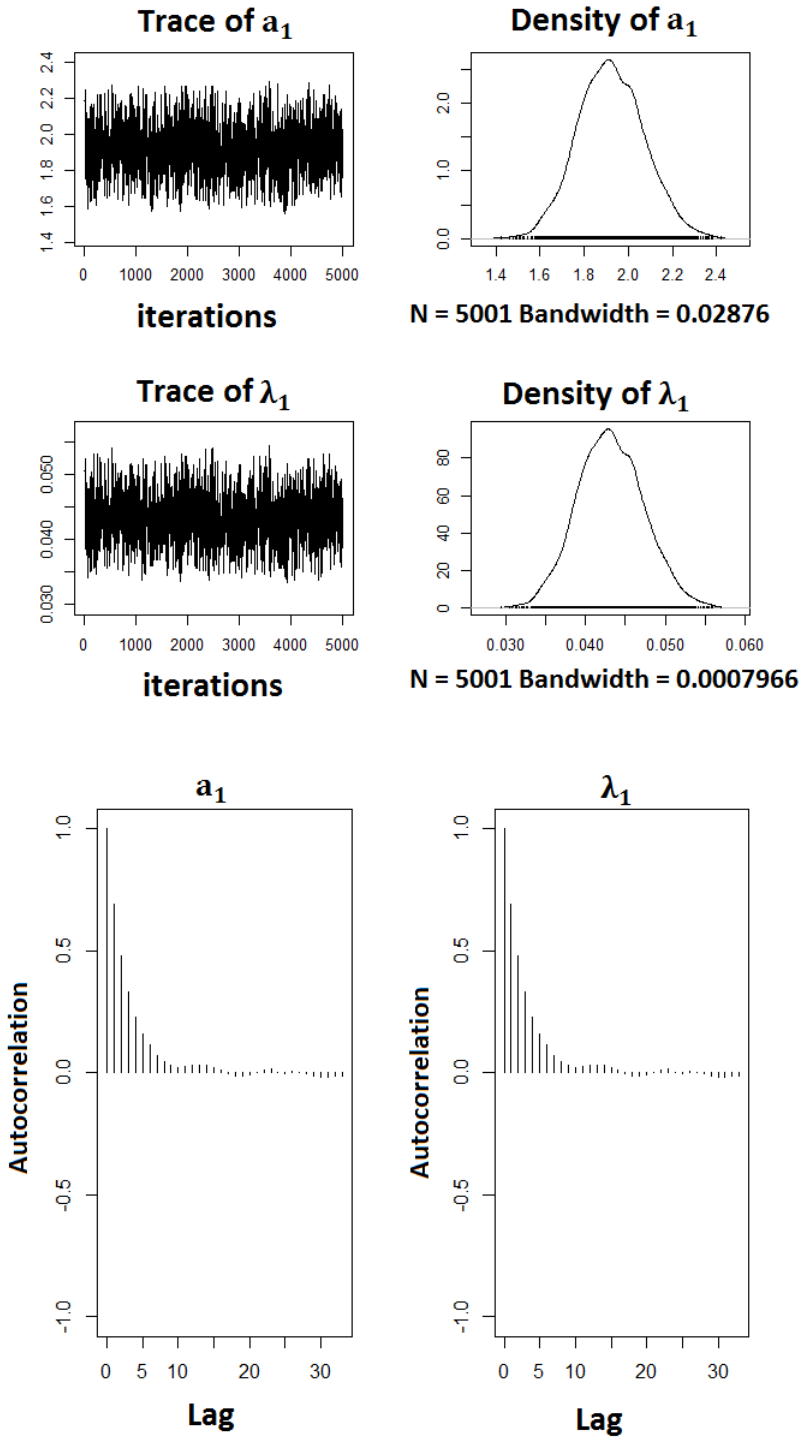

Figure D3.1: MCMC trace plots, density and autocorrelation plots of shape and rate (a_1, λ_1) estimates at stage 1for parasitic nematode data.

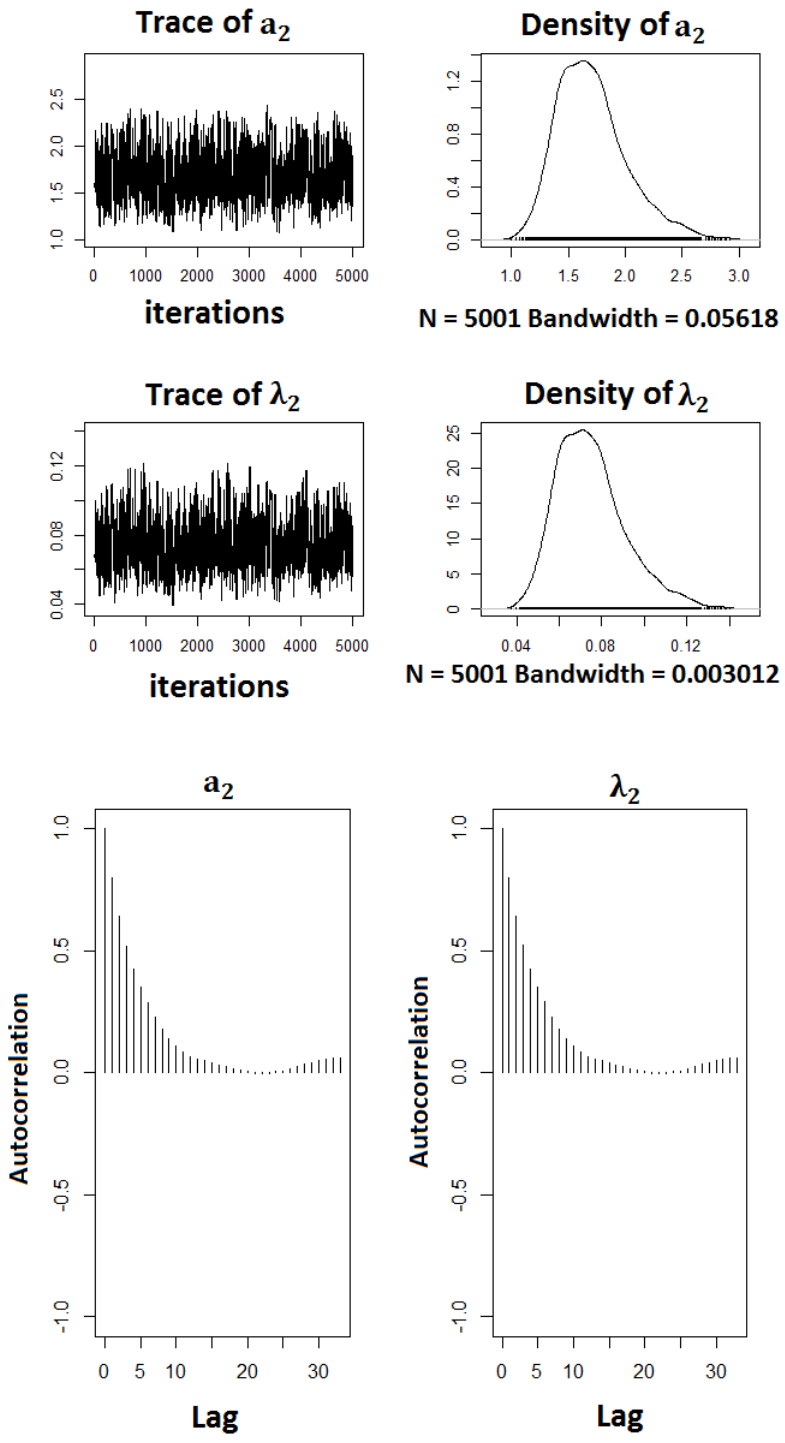

Figure D3.2: MCMC trace plots, density and autocorrelation plots of shape and rate (a_2, λ_2) estimates at stage 2 for parasitic nematode data.

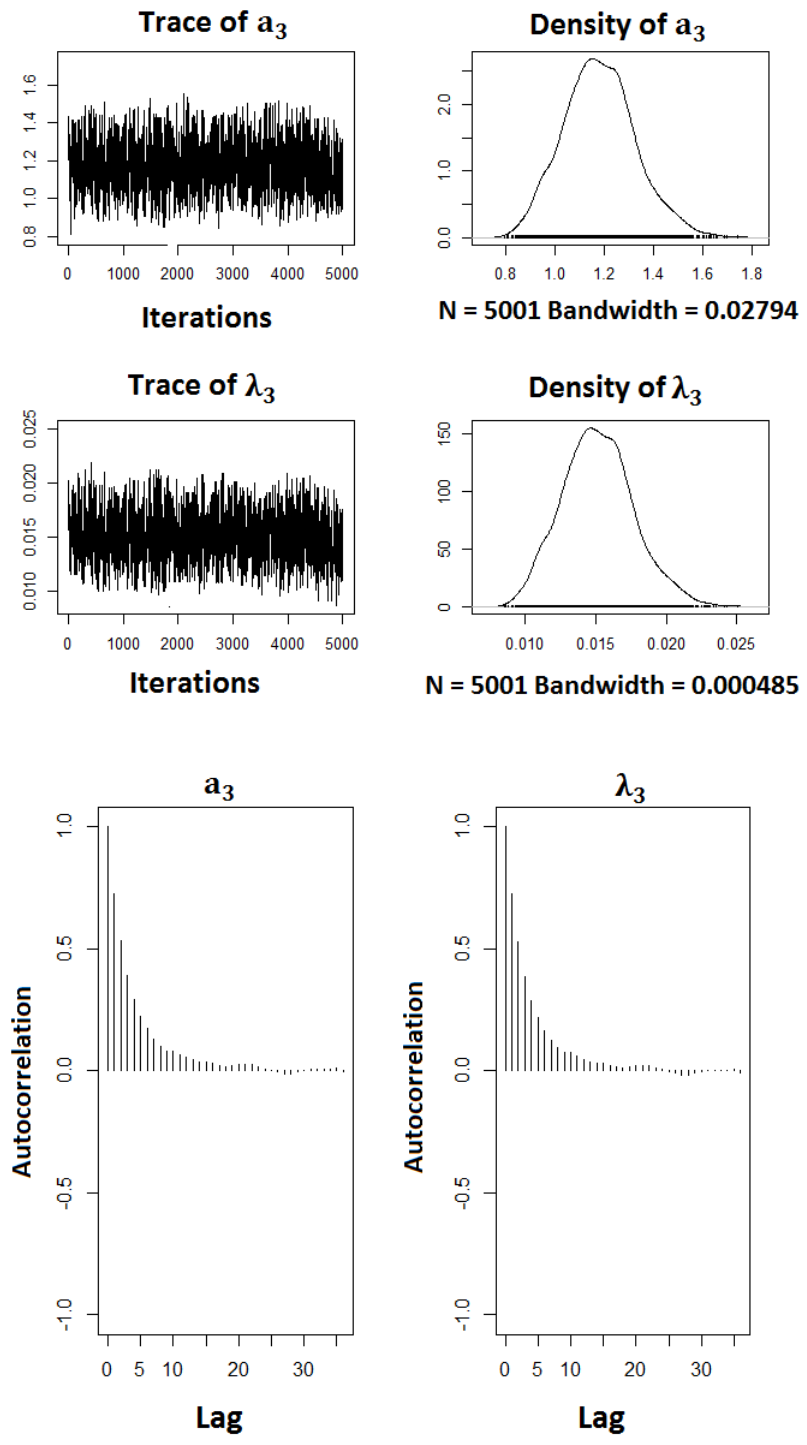

Figure D3.3: MCMC trace plots, density and autocorrelation plots of shape and rate (a_3, λ_3) estimates at stage 3 for parasitic nematode data.

D.4 Data for breast development of New Zealander schoolgirls (Section 6.2).

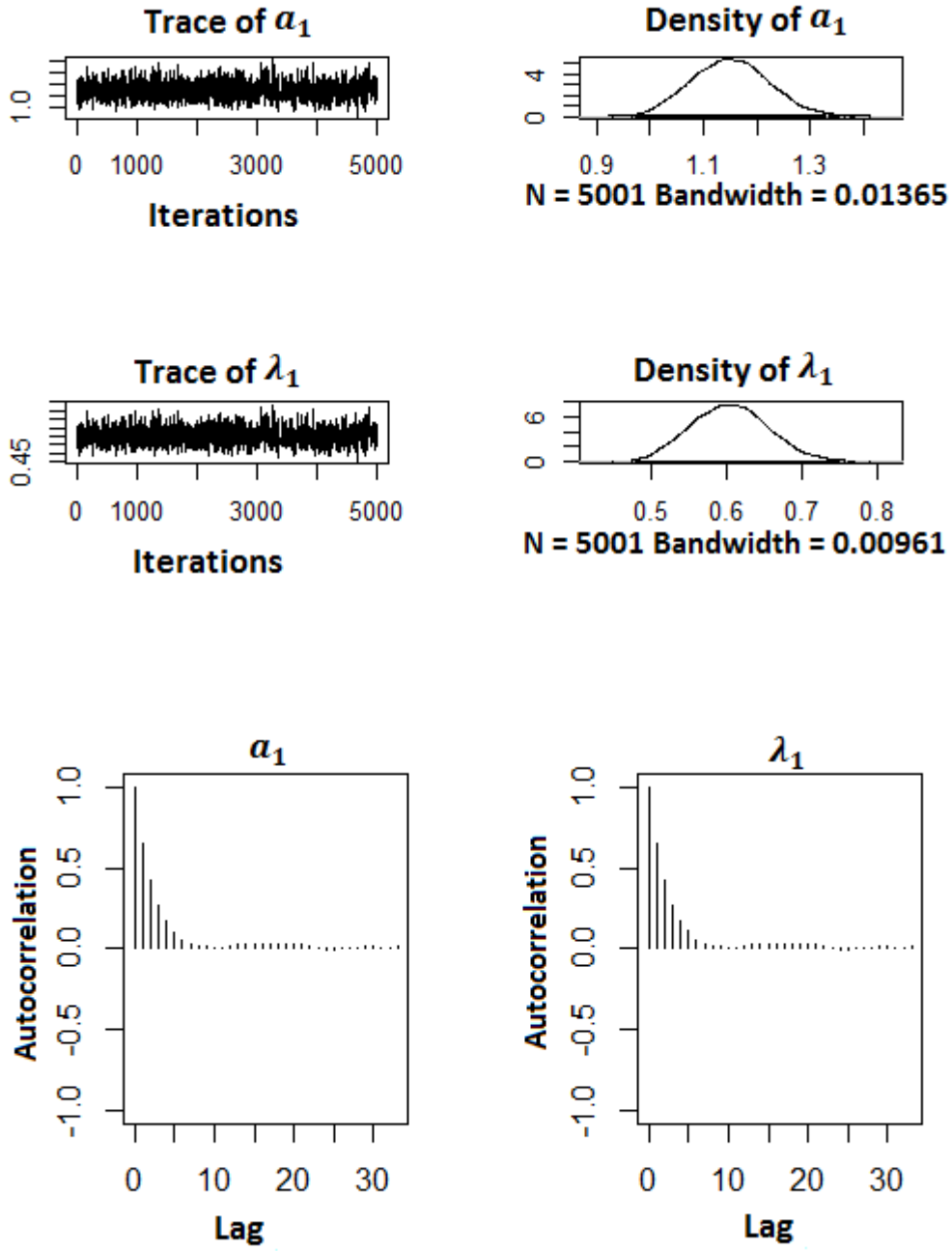

Figure D4.1: The trace, density and autocorrelation plots of shape and rate (a_1, λ_1) estimates at stage 1for breast development of New Zealander schoolgirls.

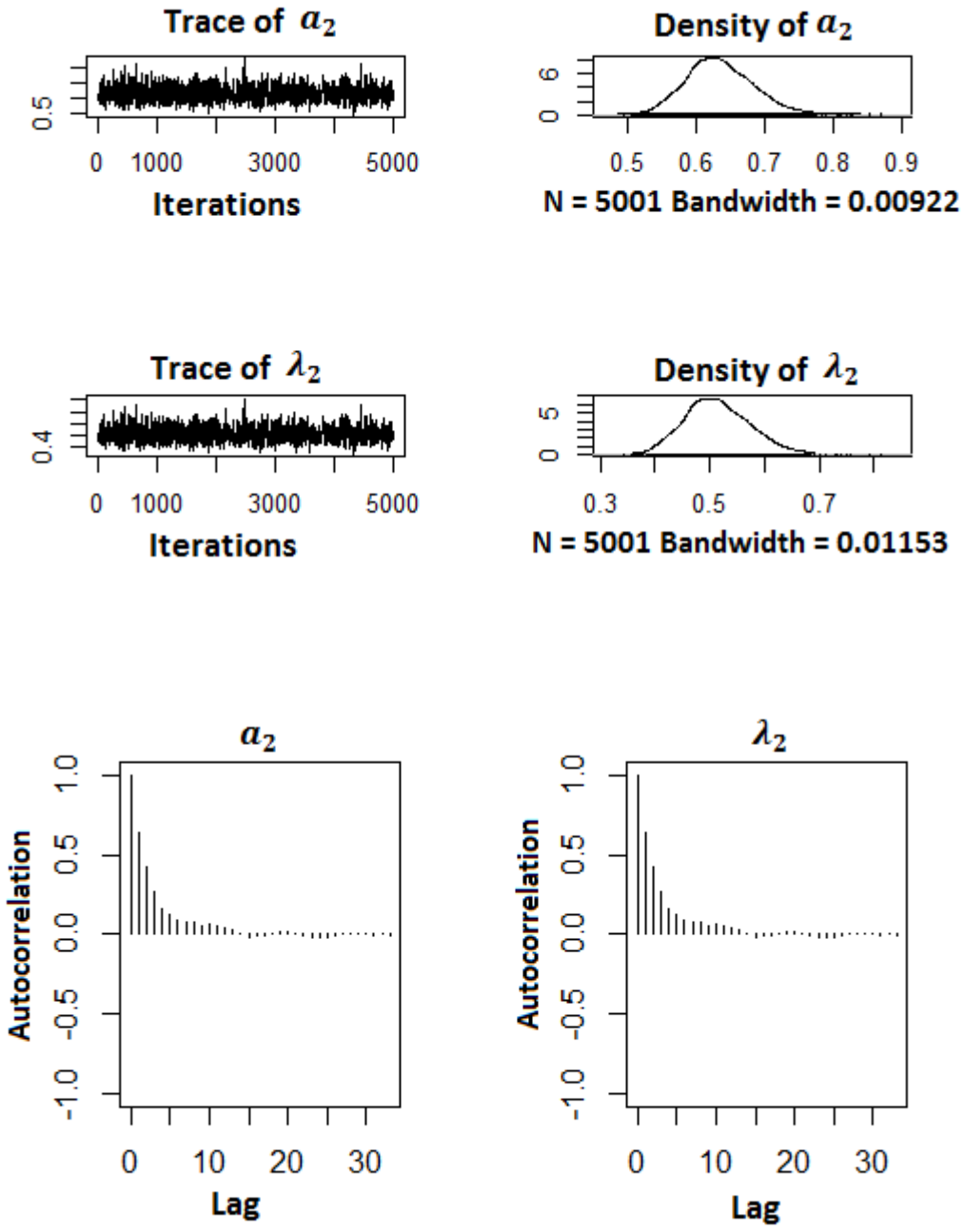

Figure D4.2: The trace, density and autocorrelation plots of shape and rate (a_2, λ_2) estimates at stage 2 for breast development of New Zealander schoolgirls.

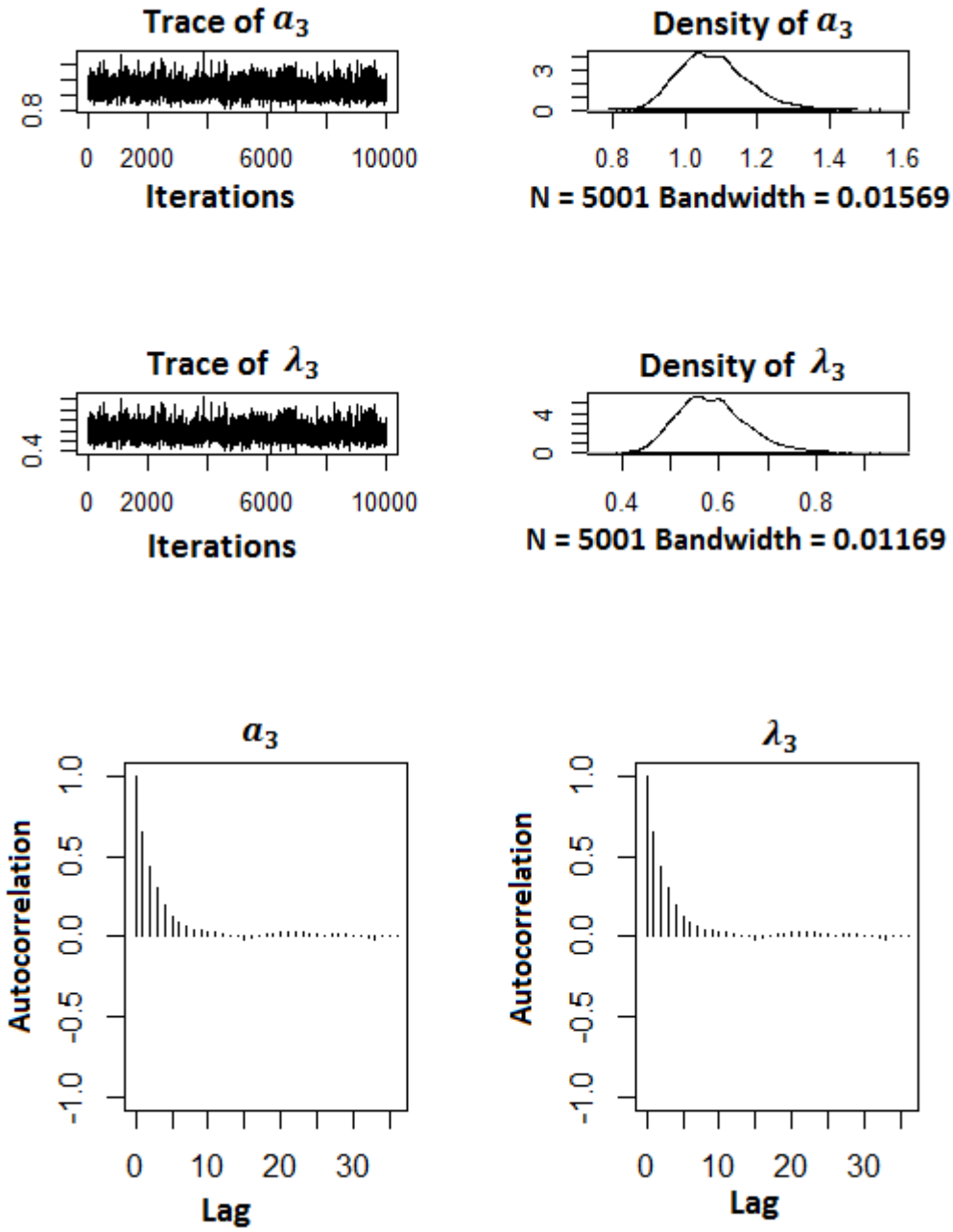

Figure D4.3: The trace, density and autocorrelation plots of shape and rate (a_3, λ_3) estimates at stage 3 for breast development of New Zealander schoolgirls.

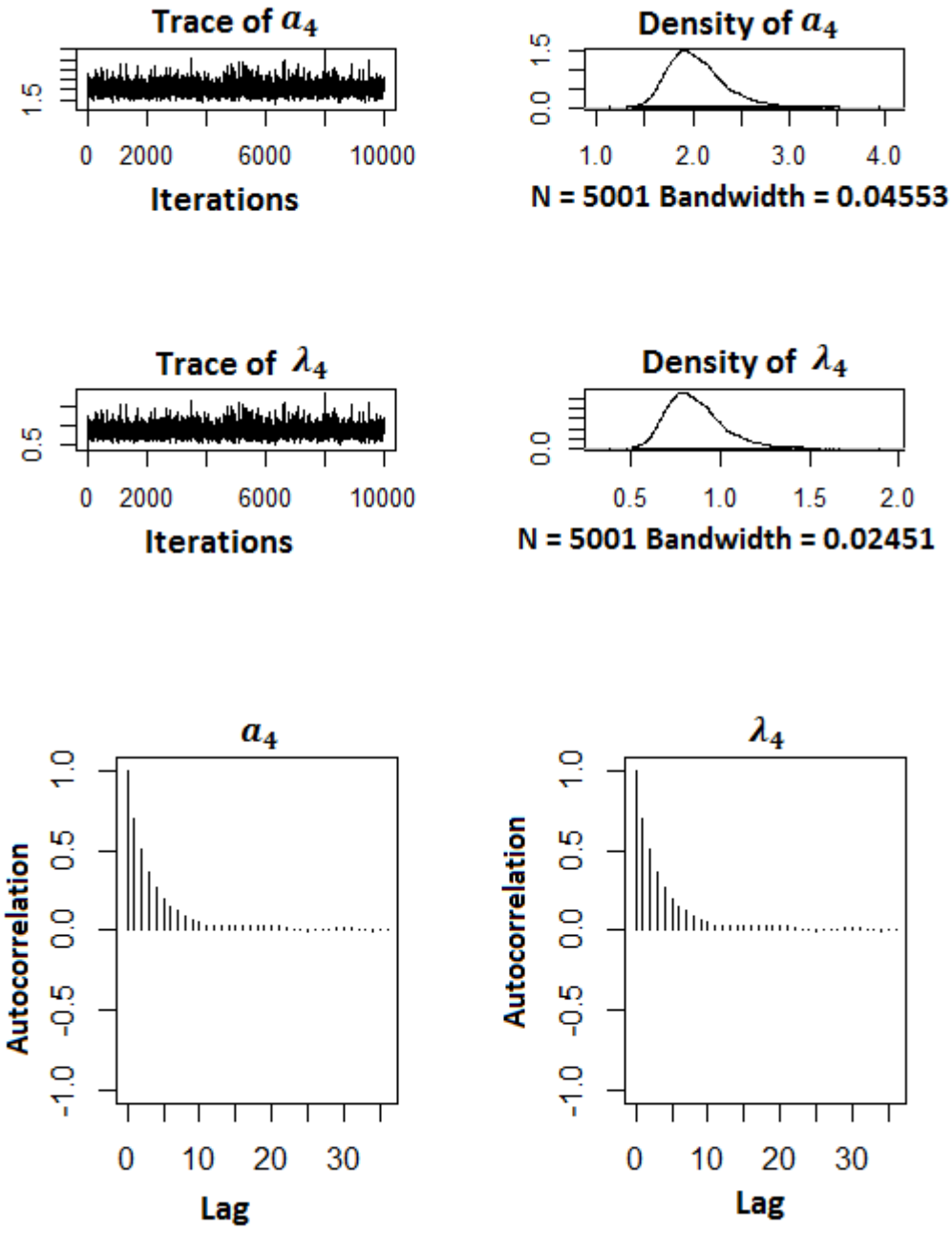

Figure D4.4: The trace, density and autocorrelation plots of shape and rate (a_4, λ_4) estimates at stage 4 for breast development of New Zealander schoolgirls.

Bibliography

[1] O. Aalen, O. Borgan, and H.K. Gjessing. *Survival and Event History Analysis*. Springer, New York, 2008.

[2] P. Amore. Asymptotic and exact series representations for the incomplete gamma function. *EPL (Europhysics Letters)*, 71(1), 2005. ISSN 0295-5075.

[3] F. Ball and R.K. Milne. On choosing values of the transform variables in empirical transform based infrence. Technical report, Department of Mathematics, University of Western Australia, 1996.

[4] I. Barra, L. Hoogerheide, S.J. Koopman, and A. Lucas. Joint independent Metropolis-Hastings methods for nonlinear non-Gaussian state space models. Technical report, Tinbergen Institute, Amsterdam, 2012.

[5] D. J. Bartholomew. *Stochastic Models for Social Processes*. John Wiley & Sons, New York, 2nd edition, 1973.

[6] J.T. Bellows and M.H. Birley. Estimating developmental and mortality rates and stage recruitment from insect stage-frequency data. *Researches on Population Ecology*, 23(2):232–244, 1981.

[7] T.G. Benton, T.C. Cameron, and A. Grand. Population responses to perturbations: predictions and responses from laboratory mite population. *Journal of Animal Ecology*, 73(5):983–995, 2004.

[8] J.M. Borwein and O. Chan. Uniform bounds for the complementary incomplete gamma function. *Mathematical Inequalities and Applications*, 12:115–121, 2009.

[9] S. Brooks, A. Gelman, G. Jones, and X.L. Meng. *Handbook of Markov Chain Monte Carlo*. Chapman and Hall/CRC Press, New York, 2011.

[10] G. Casella and R.L. Berger. *Statistical Inference*. Duxbury Press, Belmont, California, 1st edition, 1990.

[11] G. Casella and R.L. Berger. *Statistical Inference*. Duxbury Press, Pacific Grove, California, 2nd edition, 2002.

[12] H. Caswell. *Matrix Population Models*. Wiley Online Library, 2001.

[13] M. Chiani, D. Dardari, and M.K. Simon. New exponential bounds and approximations for the computation of error probability in fading channels. *IEEE*, 2(4):840–845, 2003.

[14] M.K. Cowles and B.P. Carlin. Markov chain Monte Carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association*, 91:883–904, 1996. ISSN 0162-1459.

[15] D.R. Cox and D. Oakes. *Analysis of Survival Data*. Chapman and Hall/CRC Press, New York, 1984.

[16] R.V. Craiu and J.S. Rosenthal. Bayesian computation via Markov chain Monte Carlo. *Annual Review of Statistics and Its Application*, 1:179–201, 2014. ISSN 2326-8298.

[17] P. De Valpine. Stochastic development in biologically structured population models. *Ecology*, 90(10):2889–2901, 2009. ISSN 0012-9658.

[18] P. De Valpine and J. Knape. Estimation of general multistage models from cohort data. *Journal of Agricultural, Biological, and Environmental Statistics*, 20(1):140–155, 2015. ISSN 1085-7117.

[19] P. De Valpine, K. Scranton, J. Knape, K. Ram, and N. J. Mills. The importance of individual developmental variation in stage structured population models. *Ecology Letters*, 17(8):1026–1038, 2014. ISSN 1461-0248.

[20] S. Dutta and S. Bhattacharya. Markov chain Monte Carlo based on deterministic transformations. *Statistical Methodology*, 16:100–116, 2014. ISSN 1572-3127.

[21] P.D. Feigin, R.L. Tweedie, and C. Belyea. Weighted area techniques for explicit parameter estimation in multi-stage models. *Australian Journal of Statistics*, 25(1): 1–16, 1983.

[22] Andrew Gelman and Donald B. Rubin. Inference from iterative simulation using multiple sequences. *Statistical science*, pages 457–472, 1992. ISSN 0883-4237.

[23] John Geweke. *Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments*, volume 196. Federal Reserve Bank of Minneapolis, Research Department Minneapolis, MN, USA, 1991.

[24] G.H. Givens and J.A. Hoeting. *Computational Statistics*. John Wiley & Sons, New York, 2012.

[25] J.A. Hoeting, R.L. Tweedie, and C.S. Olver. Transform estimation of parameters for stage-frequency data. *Journal of American Statistical Association*, 98:503–514, 2003.

[26] C.A. Hopkins, G. Subramanian, and H. Stallard. The development of Hymenolepis diminuata in primary and secondary infections in mice. *Parasitology*, 64(3):401–412, 1972.

[27] P. Hougaard. *Analysis of Multivariate Survival Data*. Springer, New York, 2000.

[28] J. D. Kalbfleisch and R. L. Prentice. *The Statistical Analysis of Failure Time Data*. John Wiley & Sons, New York, 1980.

[29] J. D. Kalbfleisch and R. L. Prentice. *The Statistical Analysis of Failure Time Data*. John Wiley & Sons, New York, 2002.

[30] R.A. Kempton. Statistical analysis of frequency data obtained from sampling an insect population grouped by stages. *Statistical Distributions in Scientific Work.*, pages 401–418, 1979.

[31] N. Keyfitz and H. Caswell. *Applied Mathematical Demography*. Springer, New York, 2005.

[32] W. Kimmerer and A. Gould. A Bayesian approach to estimating copepod development times from stage frequency data. *Limnology and Oceanography: Methods*, 8(4): 118–126, 2010. ISSN 1541-5856.

[33] J. Knape and P. De Valpine. Monte carlo estimation of stage structured development from cohort data. *Ecology*, 97(4):992–1002, 2016. ISSN 1939-9170.

[34] J. Knape, K.M. Daane, and P. De Valpine. Estimation of stage duration distributions and mortality under repeated cohort censuses. *Biometrics*, 70(2):346–355, 2014. ISSN 1541-0420.

[35] S.J. Koopman, A. Lucas, and M. Scharth. Numerically accelerated importance sampling for nonlinear non-Gaussian state-space models. *Journal of Business & Economic Statistics*, 33(1):114–127, 2015. ISSN 0735-0015.

[36] A.F. Laurence and B.J.T. Morgan. Selection of the transformation variable in the Laplace transform methods of estimation. *Australian Journal of Statistics*, 29(2): 113–127, 1987.

[37] J.F. Lawless. *Statistical Models and Methods for Lifetime Data*. Springer, New York, 2011.

[38] E.T. Lee and J.W. Wang. *Statistical Methods for Survival Data Analysis*. John Wiley & Sons, New York, 2003.

[39] B.F.J. Manly. *Stage-structured Populations: Sampling, Analysis and Simulation*. Chapman and Hall, New York, 1990.

[40] S.P. Meyn and R.L. Tweedie. *Markov Chains and Stochastic Stability*. Springer, New York, 2012.

[41] P.A. Murtaugh, S.C. Emerson, P.B. McEvoy, and K.M. Higgs. The statistical analysis of insect phenology. *Environmental entomology*, 41(2):355–361, 2012. ISSN 0046-225X.

[42] E. Neuman. Inequalities and bounds for the incomplete gamma function. *Results in Mathematics*, 63(3):1209–1214, 2013. ISSN 1422-6383.

[43] M.D. Ohman. Estimation of mortality for stage-structured Zooplankton populations: What is to be done? *Journal of Marine Systems*, 93:4–10, 2012. ISSN 0924-7963.

[44] H. Pham and A. Branford. Exploring parameter relations for multi-stage models in stage-wise constant and time dependent hazard rates. *Australian & New Zealand Journal of Statistics*, 58(3):357–376, 2016.

[45] M. Plummer, N. Best, K. Cowles, and K. Vines. Package coda. *URL http://cran. r-project. org/web/packages/coda/coda. pdf, accessed January*, 25, 2015.

[46] H. Qasrawi. A study of the energy flow in a natural population of the grasshopper Chorthippus parallelus zett (Acrididae). *PhD Thesis, Department of Biological Sciences, University of Exeter*, 1966.

[47] K.L.Q. Read and J.R. Ashford. A system of models for the life cycle of a biological organism. *Biometrika*, 55(1):211–221, 1968. ISSN 0006-3444.

[48] C. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer, New York, 2nd edition, 2004.

[49] C. Robert and G. Casella. *Introducing Monte Carlo Methods with R*. Springer, New York, 2009.

[50] J.S. Rosenthal. A review of asymptotic convergence for general state space Markov chains. *Far East Journal of Theoretical Statistics*, 5(1):37–50, 2001.

[51] H.J. Schuh and R.L. Tweedie. Parameter estimation using transform estimation in time-evolving models. *Mathematical Biosciences*, 45(1):37–67, 1979. ISSN 0025-5564.

[52] K. Scranton, J. Knape, and P. De Valpine. An approximate Bayesian computation approach to parameter estimation in a stochastic stage-structured population model. *Ecology*, 95(5):1418–1428, 2014. ISSN 0012-9658.

[53] A.Y. Shestopaloff and R.M. Neal. Efficient Bayesian inference for stochastic volatility models with ensemble MCMC methods. Technical report, Department of Statistical Sciences, University of Toronto, 2014.

[54] B. Sroysang. Inequalities for the incomplete gamma function. *Mathematica Aeterna*, 3(4):245–248, 2013.

[55] T.M. Therneau and P.M. Grambsch. *Modeling Survival Data Extending The Cox Model*. Springer, New York, 2000.

[56] K. Thomas and F. Ollevier. Hatching, survival, activity and penetration efficiency of second-stage larvae of Anguillicola crassus (Nematoda). *Parasitology*, 107(02): 211–217, 1993. ISSN 1469-8161.

[57] L. Tinerney. Markov chains for exploring posterior distribution. *The Annuals of Statistics*, 22(4):1701–1728, 1994.

[58] R.R. Young, N. Anderson, D. Overend, R.L. Tweedie, K.W.J. Malafant, and G.A.N. Preston. The effect of temperature on times to hatching of eggs of the nematode Ostertagia circumcincta. *Parasitology*, 81(03):477–491, 1980. ISSN 1469-8161.