

FLEXIBLE BAYESIAN MODELLING OF COVARIATE
EFFECTS ON SURVIVAL

JULIANA IWORIKUMO CONSUL

Thesis submitted for the degree of
Doctor of Philosophy



*School of Mathematics & Statistics
Newcastle University
Newcastle upon Tyne
United Kingdom*

November, 2016

I dedicate this thesis to the Almighty God!

Acknowledgements

My first word goes to God. To God be the glory for taking me thus far. I am extremely grateful to my supervisor, Dr. Malcolm Farrow for his insight, patience, guidance, kindness and helpful discussions which led to the completion of this thesis. I am very glad to have had a supervisor like him. My sincere appreciation goes to my lovely husband who is my best friend, Mr Digieneni Consul and children Lilly and Esther Consul for demonstrating great patience as the period of my research took most of the time we were meant to stay together. Thank you for your patience! I look forward to returning to my everyday life. I would like to express my gratitude to Dr. Evans Osaisai for recommending me to the Petroleum Technology Development Fund (PTDF) who partially sponsored me. Special thanks are due to my mother and siblings for their prayers, encouragement and financial support. I would like to express my special appreciation to my late father who contributed so much to my education but could not wait to see the end of the thesis. I want to thank staff and my fellow graduate students in the School of Mathematics and Statistics for their academic support. I will specially thank my friends, Mary, Maryam and Refah for their encouragement and support towards the completion of this thesis.

Abstract

Proportional hazards models are commonly used in survival analysis. Typically a baseline hazard function is combined with hazard multipliers which depend on covariate values through a logarithmic link function and a linear predictor. Models have been developed which allow flexibility in the form of the baseline hazard. However, the form of dependence of the hazard multipliers on covariates is usually specified. The aim of this research is to introduce flexibility into the form of the dependence of the hazard function on the covariates by removing the assumptions of parametric forms which are usually made. Given sufficient data, this will allow the model to adapt to the true form of the relationship and possibly uncover unexpected features.

The Bayesian approach to inference is used. The choice of a suitable prior distribution allows a compromise which relaxes the assumption of a parametric form of relationship while imposing enough structure to exploit the information in finite data sets by specifying correlations in the prior distribution between log-hazards for neighbouring covariate profiles. The choice of prior distribution can therefore be important for obtaining useful posterior inferences.

A generalised piecewise constant hazard model is introduced, in which quantitative covariates, as well as time, are categorised. Thus, the time and covariate space is divided into cells, within each of which the hazard is a constant. Two forms of prior distribution are considered, one based on a parametric model and the other using a Gaussian Markov random field. When the number of covariates is large, this approach leads to a very large number of cells, many of which might not represent any observed cases. Therefore, we consider an alternative approach in which a Gaussian process prior for the log-hazards over the covariate space is used. The posterior distribution is computed only at the observed covariate profiles.

The methodology developed is applicable to a wide range of survival data and is illustrated by applications to two data sets referring to patients with non-Hodgkin's lymphoma and leukaemia respectively.

Contents

1	Introduction and Outline of Thesis	1
1.1	Introduction	1
1.2	Why the Bayesian approach to inference?	4
1.3	Aims of the research	4
1.4	Outline of the thesis	5
2	Examples of Data	8
2.1	Introduction	8
2.2	Scotland and Newcastle Lymphoma Group (SNLG) data	8
2.2.1	Background and data collection	8
2.2.2	Non-Hodgkin’s Lymphoma (NHL)	9
2.2.3	Diffuse large B-cell lymphoma	9
2.2.4	An overview of explanatory variables of the SNLG data set	10
2.3	Leukaemia Data	15
2.3.1	Background	15
2.3.2	An overview of explanatory variables of the leukaemia data	15
2.4	Summary	16
3	Introduction to survival analysis	17
3.1	Introduction	17
3.2	Survival analysis	18
3.2.1	Censored data	18
3.2.2	The Survival Function	18
3.2.3	The Hazard Function	19
3.3	Relating covariates to the survival distribution	20
3.3.1	Proportional hazards model	20
3.3.2	Accelerated life model	21
3.4	Parametric approach to survival modelling	21
3.4.1	Weibull distribution	22

3.4.2	Likelihood functions	23
3.5	Semi-parametric hazard models	24
3.5.1	The piecewise constant hazard (PCH) model	24
3.5.2	Time varying covariate effects	25
3.5.3	Piecewise constant hazard model with time varying covariate effects	25
3.6	Frailty	26
3.7	Introduction to prognostic indices	29
3.8	Review of prognostic indices across lymphoma subtypes	30
3.8.1	International Prognostic Index (IPI)	30
3.8.2	Age-adjusted International Prognostic index (AaIPI)	30
3.8.3	Follicular lymphoma International Prognostic Index (FLIPI)	31
3.8.4	Revised International Prognostic Index (R-IPI)	31
3.8.5	Mantle cell International Prognostic Index (MIPI)	32
3.8.6	National Comprehensive Cancer Network International Prognostic Index (NCCN-IPI)	33
3.9	Summary	33
4	Introduction to Bayesian inference	35
4.1	Introduction	35
4.2	Bayes' Theorem and Bayesian Inference	37
4.3	Monte Carlo integration	38
4.4	Prior distributions	39
4.4.1	Prior elicitation	39
4.4.2	Centering of covariates	41
4.4.3	Structures for prior distributions	41
4.5	Prior and posterior predictive distribution	51
4.6	Markov chain Monte Carlo (MCMC) techniques	52
4.6.1	Introduction	52
4.6.2	Metropolis and Metropolis-Hastings algorithm	52
4.6.3	Choice of proposal distribution	53
4.6.4	Gibbs Sampler	54
4.6.5	Metropolis within Gibbs	55
4.6.6	Sampling from a multivariate normal distribution using the Cholesky decomposition	56
4.6.7	Initial values	56
4.6.8	Burn-in and convergence	57
4.6.9	Thinning	57
4.6.10	Efficiency and accuracy	58

4.7	Numerical methods for improving mixing in MCMC	59
4.7.1	Sampling the parameters using the idea of forward backward recursion	59
4.7.2	Gaussian approximation of the proposal distribution of the parameters	60
4.8	Missing data	61
4.8.1	Types of Missingness	61
4.9	Summary	63
5	Introduction to Bayesian inference in survival analysis	64
5.1	Introduction	64
5.2	Censoring and data augmentation	65
5.3	Proportional hazards and prior distribution for the coefficients of covariates	66
5.3.1	Construction of prior distribution for the coefficients of a linear predictor	66
5.4	Prior distributions for linear predictors	72
5.5	Bayesian survival analysis with Weibull lifetimes	77
5.5.1	Model, prior and likelihood	77
5.5.2	Example: Metropolis-Hastings sampling with random walk proposal	81
5.6	Application: Bayesian survival modelling to SNLG data set	83
5.6.1	Prior distributions	84
5.6.2	Posterior distributions	86
5.6.3	Calculation of the prognostic index for the Weibull survival modelling to SNLG data set	87
5.7	Bayesian survival modelling using the piecewise constant hazard model . . .	87
5.7.1	Choosing the cut points of time intervals for the piecewise constant hazard model	88
5.7.2	Construction of the likelihood for piecewise constant hazard model .	88
5.7.3	Construction of prior distribution for the parameters of piecewise constant hazard model	89
5.7.4	Construction of prior distribution for the frailty variance	90
5.8	Application: Bayesian survival modelling using the piecewise constant hazard model	90
5.9	A Prognostic index based on the piecewise constant hazard model	94
5.9.1	Introduction	94
5.9.2	Computation of survival probability at a fixed time	95
5.9.3	Practical computation of predictive median survival time	96
5.9.4	An iterative algorithm for finding the predictive median survival time	96
5.9.5	Application: Finding the predictive median survival time and survival probability at some fixed time	97

5.10	Summary	99
5.11	Flexible models for survival data	101
6	Generalised piecewise constant hazard model	103
6.1	Introduction	103
6.2	Generalised piecewise constant hazard model (GPCH model)	105
6.3	GPCH model with priors based on parametric model	107
6.3.1	Construction of prior distribution based on parametric model	107
6.3.2	MCMC sampling of the log-hazards (η) using priors based on parametric model	109
6.4	Gaussian process prior	110
6.5	Gaussian Markov random field priors	112
6.6	GPCH model with Gaussian Markov random field prior	115
6.6.1	Construction of Gaussian Markov random field prior	115
6.6.2	Frailty	115
6.6.3	MCMC sampling of the log-hazards (η) using Gaussian Markov random field prior	116
6.7	Splitting the log-hazards	118
6.8	Sampling the principal components of the log-hazards	121
6.9	Some comments on the computations using the GPCH model	123
6.10	Summary	124
7	Continuous parameter space model	125
7.1	Introduction	125
7.2	Basic ideas and notation	125
7.2.1	Unconstrained η (Gaussian process prior)	126
7.3	Bayesian modelling of the log-hazards (η) using Gaussian process prior and Weibull lifetime distribution	127
7.3.1	Constructing prior distribution for the parameters of the continuous parameter space model	128
7.3.2	Sampling from the posterior distribution of the log-hazards	129
7.4	Summary	131
8	Applications to Example Data sets and use of alternative computational methods	132
8.1	Introduction	132
8.2	Application 1: GPCH model with priors based on parametric model	133
8.3	Some alternative numerical approaches in GPCH model with priors based on parametric model	140

8.3.1	Sampling principal components using priors based on parametric model	140
8.3.2	Gaussian approximation of the fcd as a proposal distribution using priors based on parametric model	141
8.4	Splitting the log-hazards	142
8.5	Application 2: GPCH model with priors based on Gaussian Markov random field	145
8.6	Some alternative numerical approaches in GPCH model with Markov random field prior	148
8.6.1	Sampling principal components using Markov random field prior	149
8.6.2	Gaussian approximation of the proposal distribution using Markov random field prior	149
8.7	Splitting the log-hazards	149
8.8	Application 3: Continuous parameter space model with Gaussian process prior	149
8.8.1	Construction of prior distribution for the log-hazards	151
8.8.2	Posterior distributions	152
8.9	Some alternative numerical approaches in continuous parameter space model	152
8.9.1	Improving mixing in continuous parameter space model by sampling the principal components of the log-hazards	152
8.9.2	Improving mixing in continuous parameter space model by Gaussian approximation fcd of the log-hazards	154
8.10	Splitting the log-hazards and sampling from the block of the log-hazards in the continuous parameter space model	156
8.11	Calculation of effective sample size	157
8.12	General discussion on methods and alternative numerical approaches	160
8.13	Prognostic index for the generalised piecewise constant hazard model	164
8.14	Prognostic index for the continuous parameter space model	165
8.15	Summary	165
9	Conclusion and Future Work	166
9.1	Summary of thesis	166
9.2	A review of objectives of Research	168
9.3	Model choice	171
9.4	Future work	173
9.4.1	Flexible modelling with more than one event times	173
9.4.2	Flexible modelling of covariate effects using the idea of K nearest neighbours	173

9.4.3	Flexible modelling of covariate effects using Bayesian networks . . .	175
A	Appendix	176
A.1	General Appendix	176
A.1.1	Software	176
A.1.2	Using RJAGS for survival data with right censoring	177
A.1.3	Setting up the data for piecewise constant hazard model using JAGS	177
A.2	Some common distributions	178
A.2.1	Beta distribution	178
A.2.2	Gamma distribution	178
A.2.3	Dirichlet distribution	178
A.2.4	Binomial distributions	178
A.2.5	Multinomial observation	179
A.3	Appendix to Chapter 2	179
A.3.1	Partial display of SNLG data	179
A.3.2	Partial display of leukaemia data	179
A.4	Appendix to Chapter 5	180
A.4.1	Bayesian survival modelling to SNLG data set	180
A.4.2	Function to calculate the prognosis indices using the SNLG data set	182
A.4.3	Bayesian survival modelling using the piecewise constant hazard model and including frailties using the SNLG data	185
A.4.4	Bayesian survival modelling using the piecewise constant hazard model and including frailties using the leukaemia data	192
A.4.5	An iterative algorithm for finding the predictive median survival time and fixed time survival probability	194
A.5	Appendix to Chapter 8	201
A.5.1	Functions for sampling the log-hazards using priors based on the parametric model in the GPCH model	201
A.5.2	Functions for sampling the log-hazards using priors (using the second form of covariance matrix) based on the parametric model in the GPCH model	202
A.5.3	Functions consisting of all functions used in sampling the log-hazards with priors based on the parametric model	203
A.5.4	Functions for splitting the log-hazards and sampling using forward backward algorithm with priors based on the parametric model in the GPCH model	216
A.5.5	Functions for splitting the log-hazards and sampling using Cholesky decomposition	220

A.5.6	Functions for sampling the principal components of the log-hazards with priors based on the parametric model in the GPCH model . . .	221
A.5.7	Functions for sampling the log-hazards when we approximate the proposal distribution with priors based on the parametric model in the GPCH model	223
A.5.8	Functions consisting of all functions used in sampling the log-hazards with Markov random field prior in the GPCH model	224
A.5.9	Functions for sampling the log-hazards with Markov random field prior in the GPCH model	229
A.5.10	Functions for sampling the principal components of the log-hazards using Markov random field priors in the GPCH model	231
A.5.11	Functions for splitting the log-hazards and sampling using forward backward algorithm with Markov random field priors in the GPCH model	231
A.5.12	Functions for splitting the log-hazards and sampling using Cholesky decomposition	231
A.5.13	Functions for sampling the log-hazards when we approximate the proposal distribution with Markov random field priors	231
A.5.14	Functions consisting of all functions used in sampling the log-hazards using Gaussian process prior in the continuous parameter space model	232
A.5.15	Functions for sampling log-hazards with Gaussian process prior in the continuous parameter space model using SNLG data set	242
A.5.16	Functions for sampling the principal components of the log-hazards with Gaussian process priors in the continuous parameter space model	245
A.5.17	Functions for splitting and sampling the log-hazards using Gaussian process prior in continuous parameter space model	246
A.5.18	Functions for sampling the log-hazards with Gaussian process priors using Gaussian approximation of the proposal distribution in the continuous parameter space model	249
A.5.19	Function to calculate the prognosis indices using the generalised piecewise constant hazard model with the leukaemia data set	251
A.5.20	Function to calculate the prognosis indices using the continuous parameter space model with the SNLG data set	253
A.5.21	Function to calculate effective sample size	253
A.6	List of notations	254

List of Figures

5.1	Density plots of the 3 years survival probability with (a) mean -1.81 and standard deviation 0.77 (b) mean 0 and standard deviation 0.77 (c) mean -1.81 and standard deviation 1 (d) mean 2 and standard deviation 0.77 (e) mean -1.81 and standard deviation 1.5 (f) mean -1.465 and standard deviation 0.77	74
5.2	Density plots of the median survival time using different means and variances for η (a) mean -1.81 and standard deviation 0.77 (b) mean -1 and standard deviation 0.77 (c) mean -1.81 and standard deviation 1 (d) mean -1 and standard deviation 1	76
5.3	Time plots of the posterior means and ± 2 standard deviation intervals for the coefficients of age, sex and baseline with including frailty (left column) and without frailty (right column) using the SNLG data set	93
5.4	Time plots of the posterior means and ± 2 standard deviation intervals for the coefficient of age, sex and baseline with including frailty (left column) and without frailty (right column) in the model using the leukaemia data set	94
5.5	Plot of the predictive median survival time against the 3 years survival probability using a piecewise constant hazard model for the SNLG data set	98
5.6	Plot of the predictive median survival time against the 500 days survival probability using a piecewise constant hazard model for the leukaemia data set	99
5.7	Plot of the predictive median survival time against the 3 years survival probability using a piecewise constant hazard model for the SNLG data set on a restricted axis	100
5.8	Plot of the predictive median survival time against the 500 days survival probability using a piecewise constant hazard model for the leukaemia data set on a restricted axis	101
6.1	Directed acyclic graph to illustrate a 2-dimensional case of GMRF	114
6.2	Directed acyclic graph to explain the idea of splitting the log-hazards	119

8.1	Trace plot and density plot (with dash lines representing the prior density and straight lines representing the posterior density) for three log-hazards using the GPCH model and priors based on the parametric model for the leukaemia data set	137
8.2	Time plots of the posterior means and ± 2 standard deviation intervals for three log-hazards using the GPCH model where priors are based on parametric model with overlapping log-hazards for the leukaemia data set	138
8.3	Time plots of the posterior means and ± 2 standard deviation intervals for three log-hazards using the GPCH model where priors are based on parametric model without overlapping log-hazards for the leukaemia data set	139
8.4	Trace plot for three log-hazards when the proposal distribution was approximated using the GPCH model and priors based on the parametric model for the leukaemia data set	143
8.5	Trace plot and density plot (with dash lines representing the prior density and straight lines representing the posterior density) for three log-hazards using the GPCH model with Markov random field priors for the leukaemia data set	147
8.6	Time plots of the posterior means and ± 2 standard deviation intervals for three log-hazards using the GPCH model with Markov random field priors without overlapping log-hazards for the leukaemia data set	148
8.7	Trace plots for three log-hazards with approximated proposal distribution with Markov random field priors for the leukaemia dataset	150
8.8	Trace plot and Density plot for three log-hazards using the continuous parameter space model with Gaussian process priors for the SNLG data set	153
8.9	Trace plot when we sample the principal components of the log-hazards using the continuous parameter space model with Gaussian process priors for the SNLG data set	154
8.10	Trace plot for three log-hazards with Gaussian approximation of the proposal distribution using the continuous parameter space model with Gaussian process priors for the SNLG data set	157
8.11	Contour plots of the posterior (a) and proposal density (b) of the log-hazards using only sex as covariate	158
8.12	Trace plot of the log-hazards using only sex as covariate	159

List of Tables

2.1	Ann Arbor staging system	11
2.2	ECOG performance scale	11
2.3	Table showing the number of missing covariates with the corresponding number of patients	14
3.1	Summary of the IPI point, risk group and 5-year survival	30
3.2	Summary of the AaIPI point, risk group and 5-year survival	31
3.3	Summary of FLIPI score, risk group and 5-year survival	31
3.4	Summary of R-IPI score, risk group and 4-year survival	32
3.5	Summary of MIPI risk groups, the percentage of the patients and the me- dian overall survival	32
3.6	Summary of the 8 scoring points for the NCCN-IPI	33
3.7	Summary of 4 distinct risk groups with comparison with the IPI with 5- years overall survival (OS) and progression free survival (PFS) with NCCN and BCCA cohort	34
4.1	Scheme with level 1 as baseline	46
4.2	Scheme of orthogonal contrasts constrain to sum to zero	46
4.3	A scheme when the number of levels is not a power of 2	47
4.4	Example of a scheme number of levels is 5	47
5.1	Covariates and notations	83
5.2	The prior means and standard deviations for the unknown parameters . . .	85
5.3	The posterior means and standard deviations (std) for the unknown param- eters	86
5.4	The prior means and standard deviations (std) for the coefficients of the covariates	91
5.5	Posterior numerical summaries of some selected parameters in each interval in the SNLG model	92

5.6	Posteriors numerical summaries of some selected parameters in each interval in the leukaemia model	92
8.1	Ordering of covariates for the leukaemia data set for GPCH model	133
8.2	The prior means and variances for the coefficients of the ordinal covariates of the leukaemia data set	134
8.3	Covariance matrix of the parameters of the ordinal covariates of the leukaemia data set	135
8.4	Partial display of the correlation between ordinal covariate vectors with priors based on parametric model	135
8.5	Partial display of the correlation between ordinal covariate vectors with (an alternative) prior based on parametric model	136
8.6	Partial display of the prior covariance matrix of the log-hazards with priors based on parametric model	136
8.7	Partial display of the prior covariance matrix of the log-hazards with (an alternative) priors based on parametric model	137
8.8	Posterior means and standard deviation of some chosen parameters using the GPCH model with priors based on parametric model	140
8.9	Posterior means and standard deviation of some chosen parameters when we split the log-hazards using the GPCH model with priors based on parametric model	145
8.10	Partial display of the correlation between ordinal covariate vectors using the Markov random field prior with the leukaemia data set	146
8.11	Partial display of the prior covariance matrix of the log-hazards using the Markov random field prior with the leukaemia data set	146
8.12	Posterior means and standard deviation of some chosen parameters using the GPCH model with Markov random field prior	151
8.13	Partial display of the correlation between covariate vectors of the SNLG data set	151
8.14	Partial display of the prior covariance matrix of the log-hazards of the SNLG data set	152
8.15	Posterior means and standard deviation of some chosen parameters using the continuous parameter space model and SNLG data set	160
8.16	Computational and statistical efficiency	161
A.1	A Partial display of SNLG data	179
A.2	A Partial display of leukaemia data	179
A.3	Glossary of notations	255
A.4	Glossary of abbreviations	256

Chapter 1

Introduction and Outline of Thesis

1.1 Introduction

This research is concerned with the methodology of flexible Bayesian modelling of covariate effects in survival. A standard problem in survival analysis is to make inference for covariate effects and baseline hazards from life time or survival data. The proportional hazard model is well known to explore the relationship between the covariates and survival (Cox, 1972). The key assumption of the proportional hazard model is that the hazard of any individual is a fixed proportion of the hazard of any other individual. This implies that the hazard ratio is dependent only on the covariates and not on the time. The baseline hazard function is combined with hazard multipliers which depend on covariate values through a logarithmic link function and a linear predictor. Most modelling in survival analysis is usually done using a proportional hazard model.

Typically in survival analysis, assumptions of three kinds are made. The first is the proportional hazard assumption. The second is the form of the baseline hazard. The third is the form of the relationship between the covariate values and the hazards. Each of these assumptions might not be appropriate and hence we can relax them to provide a more flexible model. We can relax the first assumption by having a non-proportional hazard model and we discuss one way to do this in Chapter 3 and 5 where it is done with the piecewise constant hazard model. The form of the baseline hazard can also be relaxed using the piecewise constant hazard model. The third assumption is the main focus of this research. Thus, this research is aimed at introducing flexibility into the form of the dependence of the hazard function on the covariates by removing the assumptions of parametric forms which are usually made. There is the need to develop flexible models that will capture the underlying shape of the data and allow the model to adapt to the true form of the relationship and possibly uncover unexpected features. We discuss two ways of relaxing the form of the relationship between the covariate values and the hazards using

the generalised piecewise constant hazard and the continuous parameter space model in Chapter 6 and 7 respectively.

Parametric models for survival data may have the advantage of being simple to handle because the visualization of the hazard function is much easier (Collett, 1994). Hence, most analysts start out by fitting a parametric proportional hazard model to describe the features of the data. The baseline hazard in these parametric models might not have a simple form. One limitation of the assumed proportional hazard model which is the key to its analytical simplicity in estimating the parameters is that the parameters stay constant over time thereby forcing the covariates effects to have the same effect at all points in time. The popular Cox semi-parametric model (Cox, 1972) is semi-parametric in the sense that there is a parametric model for the dependence of the hazard multipliers on the covariates but no parametric form is specified for the baseline hazard. Similarly, the piecewise constant hazard model can be used as a relaxation of the assumption of a particular form of baseline hazard. It allows the form of the baseline hazard to change. In addition, the coefficients of the covariates can change over time and this allows for non-proportionality of hazards whereby the proportional hazard (or accelerated life) assumption will be inappropriate. Cortese *et al.* (2009) discussed the limitations of the Cox model as not capturing some important aspects of the data such as time varying covariate effects and the need for more flexibility by allowing a semi parameteric Cox model using generalised additive models. For example, the combinations of covariate values which suggest a high risk of early death may not be the same as those which suggest a relatively short survival once the initial phase has been survived. Gamerman (1991) and Wilson & Farrow (2010) discussed modelling non-proportionality of hazards using the piecewise constant hazard model with priors for the coefficients in the linear predictor taking the form of the system evolution in a dynamic linear model. Sasieni (1992) extended the proportional hazard model by allowing for the non-proportionality of hazards by incorporating time varying regression coefficients. Zhao (2010) used a mixture model approach to non-proportionality of hazards.

Typically, covariates contribute linearly to the logarithm of the hazard multiplier. This may not be appropriate as a pure linear predictor may not be sufficient to capture complex relationship between covariates and survival. Some non-linear function of the covariates may be more appropriate. There may be interaction effects between covariates. This suggests the possibility of using some kind of non-parametric regression on the covariates rather than a specified (typically linear) functional form. There are various possibilities for different forms of dependence on covariates, such as splines, basis function regression, Bayesian classification and regression trees (C&RT) (Denison *et al.*, 2002). Royston & Parmar (2002) developed a flexible parametric model which was based on the assumption of proportional hazards of covariate effects by using regression splines which are flexible

mathematical functions. Zhou *et al.* (2014) checked the assumption of linear covariate effects on survival using restricted cubic splines and Martingale residuals. Ripley *et al.* (2004) demonstrated how a neural network can be used to allow for non-linear predictors and covariate effects varying over time. Therefore, this research proposes some methods for introducing flexibility in incorporating the covariates in the survival model since the linear predictor may not be suitable. The Bayesian approach to inference will be adopted.

Existing approaches to creating flexibility in the form of dependence of the hazard on covariates are linked with the form of the baseline hazard and non-proportionality of hazards. Much effort has been focused on incorporating flexible baseline hazards and time dependent covariate effect in survival analysis. Ibrahim *et al.* (2001) focused on the use of gamma, beta and Dirichlet processes to model baseline hazards and time dependent covariate effects.

Like most analysts, we start by discussing the proportional hazard model using a parametric model for the hazard function and we illustrate inference for the covariate effects and the baseline hazards using some example life time data. We will also fit a piecewise constant hazard model which relaxes the assumption of a particular form for the baseline hazard by having sub-divided time. Here, there is flexibility in modelling of the baseline hazard and we check for time dependent covariate effects which is a form of non-proportionality of hazards. In this research, we will demonstrate using the generalised piecewise constant hazard model and continuous parameter space model as two ways of introducing flexibility to covariate effects. In the generalised piecewise constant hazard model, we categorise the quantitative covariates and make them ordinal and thus, we divide the covariate space into cells. We will approach this either by having a prior based on the parametric models (with main and interaction effects) or a Gaussian Markov random field. In the continuous parameter space model, we give a Gaussian process prior.

Two sets of data with large numbers of patients are available for illustrations in this research. One of the sets of data was provided by the Scotland and Newcastle Lymphoma Group (SNLG) (Proctor, 2000). The data include survival times of patients with non-Hodgkin's lymphoma and a large number of covariates with some missing values. Earlier work, notably the PhD project of Xioahui Zhao (Zhao, 2010), has mainly concentrated on methods for dealing with missing data, both in the analysis of the data set and in calculating prognostic indices for new patients. The other set of data was provided by the North West Leukaemia Register in the United Kingdom (Henderson *et al.*, 2002). This data set has records of incidence and subsequent survival status of all leukaemia cases in northwest England and includes survival times and a small number of covariates without missing values.

1.2 Why the Bayesian approach to inference?

The Bayesian approach to inference will be of interest in creating flexibility in the form of the dependence of the hazard on covariates. The choice of a suitable prior distribution allows a compromise which relaxes the assumption of a parametric form of relationship while imposing enough structure to exploit the information in finite data sets by specifying correlations in the prior distribution between log-hazards for neighbouring covariate profiles. The choice of prior distribution can therefore be important for obtaining useful posterior inferences.

1.3 Aims of the research

The main aims of this research are

- providing a Bayesian framework for flexibility in incorporating covariates into the model.
- building a model where there is no restriction in the way that the covariates enter the model.
- developing flexible models for covariate effects that could greatly improve the accuracy of predicting the survival.

In this research, we look at interesting questions to do with finding a “good” model for the effects of the covariates. The specific objectives of this research are to answer the following questions

- Can we fit survival models and make useful inferences without specifying the form of the dependence on the covariates?
- Can we, at the same time, not specify the form of the baseline hazard?
- Can we allow non-proportionality of hazards?
- What are the practicalities involved and how can we overcome computational difficulties?
- How do we specify meaningful prior distributions?
- How can we construct a prognostic index?

1.4 Outline of the thesis

The two data sets used as illustrations in this research are introduced in Chapter 2. The background knowledge about lymphoma and leukaemia will be provided along with information on the Scotland and Newcastle Lymphoma Group (SNLG) data set collection process. We will also give an overview of the explanatory covariates of both data sets.

Chapter 3 introduces survival analysis with censored data. We discuss the proportional hazard and accelerated life models as ways of relating the covariates to the survival distribution using a parametric approach like the Weibull distribution as an illustration. The piecewise constant hazard model is discussed as an alternative to the usual parametric model. We introduce frailty and illustrate with an example to show that it could cause non-proportionality when included in the model. We review some important lymphoma prognostic indices. These indices include the most widely used IPI (Shipp *et al.*, 1993) and FLIPI (Solal-Celigny, 2004) and others are AaIPI (Shipp *et al.*, 1993), RIPI (Sehn *et al.*, 2007) and MIPI (Hoster *et al.*, 2008).

Chapter 4 reviews Bayesian inference. We introduce the use of Bayes theorem and some Markov chain Monte Carlo (MCMC) techniques which include the Metropolis-Hastings algorithm, Gibbs sampling and Metropolis within Gibbs sampling. We provide some information on the construction and elicitation of priors. The discussion on the construction of priors for the model parameters will include quantitative, binary covariates and, unordered and ordered factors. When fitting with MCMC methods, we may observe considerable autocorrelation and poor mixing because of the correlations between model parameters. Some approaches to improving mixing in MCMC when samples are strongly correlated are suggested. The numerical approaches include the forward backward algorithm and the use of a Gaussian approximation to find a suitable proposal distribution.

In Chapter 5, we will apply Bayesian inference to survival analysis using the two data sets as illustrations. We illustrate the construction of priors for the model parameters for quantitative, binary and categorical covariate parameters using the two data sets. The usual assumptions of a proportional hazard model are illustrated using the non-Hodgkins lymphoma data in using a Bayesian approach and a prognostic index is produced following Zhao (2010). We discuss the piecewise constant hazard model as a model which allows the relaxation of the form of the baseline hazard and one possibility of non-proportionality of hazards if the covariate effects change over time. We illustrate the piecewise constant hazard models with time varying covariate effects using both the non-Hodgkins lymphoma and leukaemia data sets. The predictive median survival time and the fixed time survival probability are suggested as ways of constructing the prognostic indices when using the piecewise constant hazard model.

In Chapter 6, the generalised piecewise constant hazard model will be discussed as

an extension of the piecewise constant hazard model where we introduce flexibility into the form of the dependence of the hazard function on the covariate effects model by categorising the covariates. Each combination of covariate profile with a time interval corresponds to a node in a multidimensional rectangular lattice. We refer to these nodes as cells. Associated with each cell is an unknown hazard value. We discuss the generalised piecewise constant hazard model using two forms of prior distribution. One of the priors will be based on a parametric model and the other uses a Gaussian Markov random field. The structure of the former will allow for the use of prior beliefs about main and interaction effects. In the Gaussian Markov random field prior, the log-hazard in each cell depends on those in its neighbouring cells but is conditionally independent of all other log-hazards given its neighbouring log-hazards. The generalised piecewise constant hazard model will be illustrated using the leukaemia data set. We discuss numerical approaches of improving the mixing of the MCMC by sampling the principal components of the log-hazards since the principal components are uncorrelated in the prior which will improve mixing. Another approach to improving mixing was splitting the log-hazard of an individual into two parts where one part carries the dependence between the log-hazards and the other part was the logarithm of the frailty of the individual. Thus, we sample the part that carries the dependence. We make some comments on the computations using the generalised piecewise constant hazard model and some alternatives to the approach.

Chapter 7 presents the continuous parameter space model using a Gaussian process prior over the covariate space of the log-hazards as an alternative to the generalised piecewise constant hazard model where we leave the continuous covariates as they are. Each observed covariate profile thus has a separate log-hazard and, in the prior, these are given a correlation based on a distance measure in the covariate space. We suggested improving the mixing using the numerical approaches discussed in Chapter 6.

In Chapter 8, we apply the methodologies mentioned in Chapter 6 and Chapter 7 to the two sample data sets. We model the log-hazards representing the possible combination of covariates using the generalised piecewise constant hazard model using priors based on the parametric model and Gaussian Markov random field. We suggested some numerical approaches to improving mixing in MCMC. We will also illustrate using the continuous parameter space model using a Gaussian process prior and Weibull lifetime where we increase the number of parameters so that there is exactly one parameter for each distinct covariate profile in order not to have the constraints implied by the standard parametric model. We illustrate improving poor mixing in the MCMC by sampling the principal components of the log-hazards which are independent in the prior, sampling of the log-hazards as a block by approximating the joint full conditional distribution as the proposal distribution and by splitting the log-hazards into two parts to include log-normal frailties in the model then sampling using the Cholesky decomposition. In the latter case, we

separate the mean log-hazards excluding the frailties and we sample log-hazards one at a time using the Metropolis-Hastings steps and then we sample the vector of mean log-hazards not including frailties since it will have a multivariate normal full conditional distribution.

Finally, Chapter 9 will summarise the project, make some conclusions and give ideas for future work.

The methods developed are implemented in R functions (R Development Core Team, 2008).

Chapter 2

Examples of Data

2.1 Introduction

This chapter introduces the two data sets used for illustrations in this research. The background information, the process of data collection, definition and types of variables of the data sets will be discussed.

2.2 Scotland and Newcastle Lymphoma Group (SNLG) data

2.2.1 Background and data collection

The Scotland and Newcastle Lymphoma Group (SNLG) was formed in 1977. It built up a database on about 18,000 patients with lymphoma within Northern England and Scotland. The SNLG data set was collected over a period of 10-years (1992 to 2002). The SNLG includes specialists from the disciplines of medicine, surgery, pathology, haematology, radiology, medical and clinical oncology and other specialities. The data were provided for this project by Professor Stephen J. Proctor, Dr. Michal Sieniawski and Ms Jo White. These data are collected using a process known as Population Adjusted Clinical Epidemiology (PACE) which was used by the Northern Regional Haematology Group (NRHG) (Proctor, 2000). The data collected have been used by different groups of people working in individual centres with the aim of promoting collaboration in the assessment of patients with the disease, developing the procedures for assessing the extent of the disease, and to promote co-operative prospective clinical trials.

The data set was an unselected data collection on all cases of lymphoma from North England and Scotland which enabled the SNLG to become a leader in population lymphoma studies internationally. In this research, we will concentrate on the non-Hodgkin's lymphoma (NHL), a subset of the SNLG data set. NHL is one of the most common types of lymphoma.

2.2.2 Non-Hodgkin's Lymphoma (NHL)

Non-Hodgkin's Lymphoma (NHL) is the 6th most common cancer in the UK. NHL begins in the lymphocyte cells which are found in the lymph nodes and other lymphoid tissues like the bone marrow and spleen of the body (American Research Society, 2013). Cancer Research UK (2012) recorded around 12,200 cases of NHL each year in adults. About 4 out of every 100 cancers diagnosed are NHL. American Research Society (2013) also records that NHL accounts for about 4% of all cancer cases and 95% of the NHL cases occur in adults in the United States. The lymphatic system consists of lymph vessels that run throughout the body. It is part of the immune system and fights infections and other diseases. Hence, the lymph nodes are the most common places where the NHL is first found. It can also occur in other organs of the body (Cancer Research UK, 2012).

NHL begins when the lymphocytes (white blood cells) of the body behave abnormally by dividing themselves continuously before they are fully matured and hence, produce too many cells which fail to die off naturally as they ought to. These abnormal cells do not protect the body but rather form a mass of tissue.

NHL is normally classified by the type of cell affected. There are two main types of lymphocytes cell. These are the B lymphocytes (B-cells) and T lymphocytes (T-cells). We have B-cell lymphoma or T-cell lymphoma. The B-cells produce proteins which protect the body against bacteria (American Research Society, 2013) while the T-cells help to destroy the infected cells and release substances to attract other types of white blood cells which take part in digesting these infected cells. The B-cell lymphoma is mostly common with aged or older NHL patients while T-cell lymphoma is more common in teenagers and younger adults (Cancer Research UK, 2012).

2.2.3 Diffuse large B-cell lymphoma

Diffuse large B-cell lymphoma (DLBCL) is the most common type of NHL. It is a cancer of the B-cells. The diffuse large B-cell lymphoma occurs when B lymphocytes continue to develop and get matured and larger than normal and stop responding to signals that limit the growth and reproduction of cells.

The first sign of the condition is often a painless swelling in the neck, armpit or groin, caused by enlarged lymph nodes. Sometimes lymph nodes in more than one part of the body are affected. In some people, DLBCL does not start in the lymph nodes but rather develops in another part of the body like the stomach, lungs, skin, thyroid, brain or even spinal cord. These are called extranodal disease. The lymphoma may spread to various organs in the body such as the liver, lungs or bones. Some people may experience a loss of appetite and tiredness. Some other symptoms which are known as B symptoms, include night sweats, unexplained high temperatures and weight loss. DLBCL can be diagnosed

by removing an enlarged lymph node (biopsy) and examining the cells under a microscope. Some additional tests include blood tests, x-rays, scans and bone marrow samples which are done to find out more about the type and extent of spread of lymphoma.

DLBCL can occur at any time between adolescence and old age but is most commonly diagnosed around an average age of 65 years (American Research Society, 2013). DLBCL is slightly more common in men than in women (Tilly & Dreyling, 2009). The causes of DLBCL are unknown. Like other cancers, DLBCL is not infectious and can not be passed on to other people (Vickers, 1999).

2.2.4 An overview of explanatory variables of the SNLG data set

One of the data sets used as an example in this research is the SNLG data set. The time to death or last follow - up which is the overall survival (OS), the time to first relapse (TFR) and the censoring indicator which was “1” for death and “0” for censoring were given in the data. In this section, we will present fourteen prognostic variables chosen following the clinical research by Prof Proctor, Dr Sieniawski and Mrs White. These variables used were continuous, binary or categorical.

Age: This explanatory variable refers to the patient’s age at diagnosis. NHL can happen at any age but Cancer Research UK (2012) recorded that about 6 in 10 of all cases of diagnosed NHL are aged at least 65 years. Age was recorded as a continuous variable. The average age in the DLBCL data set is 62 years and the standard deviation is 14.2.

Sex: This variable describes the gender of patients. Sex is a 2-level categorical variable. The value 1 was used for male and 2 was for female. There were 704 males and 687 females in the data. Cancer Research UK (2012) recorded that NHL was slightly more common in men than in women.

Stage: This involves the extent of spread of the disease. It is usually advised that further tests be done to assess the extent of spread (staging) of the disease, size of the tumour and whether the lymph nodes containing the cancer has spread from the original site to other parts of the body once NHL is diagnosed. Staging involves a number of tests like physical examination, biopsies of enlarged lymph nodes or other abnormal areas, blood tests, imaging tests such as computer tomography (CT) scans, bone marrow aspiration and biopsy (not always carried out) and lumbar puncture (may not be carried out) (Cancer Research UK, 2012).

The Ann Arbor staging system in Table 2.1 (Goffinet *et al.* (1973); Hirsh (1990)) is the first and most often used staging system to describe the extent of spread and location of NHL in adults. Stage is an ordinal categorical variable. The stages are described using Roman numerals I, II, III and IV (1-4). Stage I, II, III and IV are also called Early, Locally Advanced, Advanced and Widespread disease respectively (Lymphoma.org, 2016). In this

thesis, the stages were recorded with the values 1,2,3 and 4. The Ann Arbor staging system was used.

Table 2.1: Ann Arbor staging system

Stage	Description
I	The lymphoma is found in one lymph node region or one region of a single organ outside the lymph system .
II	The lymphoma is found in two or more lymph node regions and on one of the sides of the diaphragm alone.
III	The lymphoma is found in lymph nodes regions on both sides of the diaphragm.
IV	The lymphoma has spread to one or more organs or tissues outside the lymphatic system, with or without involving a nearby lymph node.

Eastern Cooperative Oncology Group (ECOG) performance scale: This variable is a scale and criterion used by doctors and researchers to assess the progress and how the disease affects the daily living abilities of the patient (Oken *et al.*, 1982). This is an ordinal categorical variable. ECOG has a scale from 0 to 5. The value 5 indicates that the patient is dead so the ECOG scale is normally taken from 0 to 4. The ECOG performance is given in Table 2.2. In this thesis, the variable ECOG was recorded with

Table 2.2: ECOG performance scale

Performance	Scale definition
0	The individual is fully active and has no restrictions on performance.
1	The individual is restricted from performing strenuous physical activities but he or she is allowed to carry out light work.
2	The individual is capable of taking care of himself or herself but unable to carry out any work activities.
3	The individual is capable of partially taking care of himself or herself but confined to bed or chair more than 50% of waking hours.
4	The individual is completely disabled and confined to bed or chair. He or she can not take care of himself or herself.

the values 1,2,3,4 and 5.

Serum Lactate Dehydrogenase (LDH): This variable is one of the group of enzymes found in the blood and other body tissues. It is involved in the production of energy in cells. An increase in the amount of LDH in the blood may be a sign of tissue damage. LDH can be used as either a continuous or categorical variable (Longmore & Longmore, 2007). In this thesis, the normal range of LDH was used as values between 70 to 250

IU/L and the value was not recorded when it was “normal”. Endrizzi *et al.* (1982) used three levels of LDH. Level 1 (normal range) with LDH value less than 250 international units per litre (IU/L), level 2 (moderately increased) with LDH between 250IU/L and 500IU/L and level 3 (highly increased) with LDH more than 500IU/L. Butt *et al.* (2002) recorded the normal range of LDH as between 100IU/l and 250IU/L. The lower limit level and upper limit level varies in different hospital because of the use of different calibrations. The difference between the LDH level and LDH upper limit was divided by their corresponding upper limit $((\text{LDH.level} - \text{LDH.upper.limit})/\text{LDH.upper.limit})$. However, because the lower and upper limit varies between hospital, Endrizzi *et al.* (1982) system was modified as will be discussed below.

Suppose we let the observed LDH be L and the corresponding upper limit for the hospital where the measurement was made be L_u . To allow for the calibration differences, we give the value “1” if $L < L_u$, “2” if $L_u < L < 2L_u$ and “3” if $L > 2L_u$. Equivalently, we give “1” if $\log \frac{L}{L_u} < 0$, “2” if $0 < \log \frac{L}{L_u} < 0.7$ and “3” if $\log \frac{L}{L_u} > 0.7$ (where $\log 2$ is ≈ 0.7). Hence, our coding 1, 2, 3 corresponds to level 1, level 2 and level 3 respectively with allowance for calibration. Zhou *et al.* (2014) in the development of the National Comprehensive Cancer Network International Prognostic Index (NCCN-IPI), categorised LDH ratio into > 1 , > 2 and > 3 (using normal LDH (ratio ≤ 1) as a reference). We will note that the actual values of LDH could be used in cases where they are available.

Haemoglobin (HB): This variable is a protein that enables the red blood cells to carry oxygen from the lungs to the rest of the body. The amount of haemoglobin determines how much oxygen the red blood cells are capable of carrying to other cells. Low or reduced HB may be due to low red cell mass and could lead to inadequate concentration of oxygen within the vital organs. The reference normal levels of haemoglobin has a range from 130 to 180 grams per litre (g/dl) for men and 115 to 160 grams per litre (g/dl) for women (Longmore & Longmore, 2007). Smokers often show an increase in their haemoglobin level. The HB level in the data used in this thesis has continuous values between 70 and 180 (g/dl). Haemoglobin will be used as a continuous variable.

White Blood Cell (WBC): These are also called leukocytes. The WBC make up part of the immune system and defend the body against infection and foreign materials. The WBC just like other blood cells are produced in the bone marrow. If the WBC of a patient is above its upper limit of normality, this signifies the presence of infection, inflammatory processes and cancer in the bodies. On the other hand, a low WBC signifies there are many factors of viral infection, HIV infection or marrow infiltration (Longmore & Longmore, 2007). WBC is measured in $10^9/l$. LaFleur-Brooks (1994) classified patients with White Blood Cell Count between $4 \times 10^9/l$ and $11 \times 10^9/l$ as those with healthy condition. Johnston (1996) recorded the normal range of WBC as values between $3.9 \times 10^9/l$ and $11.3 \times 10^9/l$. In this data set, WBC will be a continuous variable with values

between $1 \times 10^9/l$ and $30 \times 10^9/l$.

Binary variables

Some of the covariates used were represented as binary variables. In principle, the variables serum albumin, alkaline phosphatase and blood urea nitrogen are continuous. However, in the NHL data, measurement of the variables in the normal range were recorded as “normal” while measurements outside normal range were not recorded. Hence, in this thesis, these binary variables were coded 1 for “normal range”, 2 for “abnormal”, and 9 for “missing”.

Serum Albumin (Albumin): This is the most abundant type of protein in the blood. It promotes the transfer of nutrients and wastes to and from the blood and cells. It is manufactured in the liver and decreases in chronic liver disease. It also reflects one’s general nutritional status. Albumin has a normal range of concentrations in blood of 35 to 50 grams/litre (g/l) (Longmore & Longmore, 2007). Low albumin suggest malignancy while high albumin indicate dehydration.

Alkaline Phosphatase (AP): This reflects the alkaline blood pH. An increased alkaline phosphatase indicates an alkaline blood system and inefficient mineral transfer to the cells. A decreased alkaline phosphatase expresses an exhausted adrenal system and an acid blood system (common in chronic disease). The AP has a normal range of 50 to 150IU/L for non pregnant adults (Longmore & Longmore, 2007).

Blood Urea Nitrogen (urea): This is the waste from the liver which is processed by the kidneys. Blood urea nitrogen also reflects carbohydrate storage in the liver associated with the kidneys. It tends to increase in dehydration and in kidney or heart failure. An increase in blood urea nitrogen can also indicate that the liver is inactive. A decrease in blood urea nitrogen indicates that the pancreas and/or adrenal gland is inactive. The normal range of blood urea nitrogen for an adult is between 2.5 and 6.7 mmol per litre (Longmore & Longmore, 2007).

Extranodal without Bone Marrow (extranod): This occurs when the lymphoma starts in areas outside the lymph nodes (Cancer Research UK, 2012). Such lymphomas can develop in other tissues around the body such as the kidneys and gut. The value 1 is used to indicate a patient with extranodal disease without an indication from the bone marrow otherwise the value 2 or the value 9 was used if missing.

Bulk Disease (Bulk): This is used to describe a lymph node mass measuring 10 cm (4 inches) or more in its greatest dimension or at least one-third as wide as the chest as determined by computed tomography(CT) scan (Longmore & Longmore, 2007). However, some authors suggested a variety of sizes of nodal mass. Federico *et al.* (2009) suggested 6cm as the best size for the definition of bulky disease for follicular lymphoma and Cheson *et al.* (2014) suggested 6 to 10cm for DLBCL patients who were treated with rituximab. In this thesis, the value 1 is used to indicate a patient with bulk disease otherwise the

value 2 or the value 9 was used if missing.

Bone Marrow Involvement (marrow): Bone marrow is a spongy material in the center of bones where blood cells are made (Cancer Research UK, 2012). Bone marrow biopsy (BMB) is often performed even when the likelihood of its involvement is low (Cheson *et al.*, 2014). The positron emission tomography (PET)-computed tomography (CT) is more sensitive than BMB. If the PET or CT is performed, the BMB is no longer required. In this thesis, the value 1 is used to indicate that a patient has shown the evidence of the lymphoma disease by the cause of bone marrow otherwise the value 2 or the value 9 was used if missing.

B-symptoms (Bsy): This is an indicator variable to show whether a patient has specific symptoms. A patient is said to have B-symptoms if he or she has weight loss, fevers or night sweats. The presence or absence of B-symptoms has a significance for the prognosis and can be seen in the staging of NHL. In this thesis, the value 1 was used to indicate the absence of B-symptoms while the value 2 was used to indicate the presence of B-symptoms and 9 was used to indicate missing values.

In the SNLG data set, the covariates age, sex and stage were the only completely observed covariates. There are missing data in the other covariates. Out of the 1391 patients, there are only 636 patients having all 14 complete covariates. Table 2.3 summarises the number of missing covariates with the corresponding number of patients.

Table 2.3: Table showing the number of missing covariates with the corresponding number of patients

Number of missing covariates	Number of patients
0	636
1	449
2	188
3	57
4	22
5	18
6	12
7	8
8	0
9	1

A cyclophosphamide, doxorubicin, vincristine, and prednisone (CHOP) chemotherapy has been the standard therapy for diffuse large B-cell lymphoma. All patients in the SNLG data set are known to have received CHOP. The addition of rituximab to CHOP has been shown to improve outcome in elderly patients with DLBCL (Sehn, 2005). There is also a smaller data set of 309 patients who were treated with rituximab (Zhou *et al.*, 2014). The

variables used in both data sets were exactly the same.

2.3 Leukaemia Data

2.3.1 Background

The leukaemia data set from the North West Leukaemia Register in the United Kingdom (Henderson *et al.*, 2002) was also used in this thesis. This database hold records of incidence and subsequent survival status of all leukaemia cases in northwest England. Results were recorded for 1043 patients who were diagnosed with acute myeloid leukaemia between 1982 and 1998.

Leukaemia is a cancer of the blood-forming cells. Like lymphoma and other cancers, it develops when some cells in the body start to grow out of control. These cells might be dividing too quickly and build up in large numbers. Leukaemia starts from the bone marrow and sometimes spreads from there into the bloodstream and to other parts of the body. Leukaemia is described as either lymphoid or myeloid. This is dependent on the blood-forming cell in which it develops. People with leukaemia experience some common symptoms like tiredness, inability to work and breathlessness on exertion which are caused by lack of red blood cells. They also have an increased risk of developing infections due to lack of white blood cells. Sometimes, leukaemia patients have bruises and bleeding from some part of the body such as the nose due to lack of platelets which normally help with blood clotting. Some other symptoms are like those of lymphoma. Fast developing leukaemia is called acute leukaemia otherwise it is called chronic leukaemia.

Leukaemia could be diagnosed by conducting some blood and bone marrow test to check the counts of the different blood cells. Sometimes, bone marrow biopsy can be carried out like in the case of lymphoma.

2.3.2 An overview of explanatory variables of the leukaemia data

The data set includes four covariates which are the age, sex, white blood cell count (WBC) and a measure of deprivation of the area of residence which uses the Townsend score (Thunhurst, 1988). There were 879 (84%) patients who died and 164 (16%) patients who were censored. The data set includes time in days until the death of the patient. The censoring indicator was “1” for death and “0” for censoring. The variables used in this leukaemia data set are discussed below.

Age: This is the age (in years) of the patient. Leukaemia can occur at any age but it is more common between the ages of 2 and 5 years. The risk of leukaemia also increases again in individuals aged over 45 years. Acute leukaemia is most common in adults. The average age in the leukaemia data set is 60.7 years and the standard deviation is 18.3 years

Sex: This is the sex of the patient. Yamamoto & Goodman (2007) record that leukaemia occurs more frequently in males than females. Female was indicated as 0 while male as 1. There were 547 males and 496 females in the data.

White blood cell (WBC): This is the white blood cell count (1 unit is $50 \times 10^9/l$) (see Section 2.2.4)

Deprivation score (Depsc): The deprivation score is a measure of poverty/affluence for the residential location of the patient. It uses the Townsend score (Thunhurst, 1988) on a scale from -7 to 10 (lower values indicate more severe deprivation).

The treatment of leukaemia depends on the age, kind of leukaemia and the general health of the patient. Leukemia treatments include, chemotherapy, immunotherapy, radiation therapy, stem cell transplantation and surgery. The first treatment offered to people diagnosed with leukaemia is chemotherapy. This involves destroying both the good and bad cells from the bone marrow. The good cells are grown back in the marrow which recovers the blood counts. The kind of leukaemia determines the number of courses of chemotherapy treatment. Some patients could have bone marrow and stem cell transplantation. This treatment involves destroying cancerous bone marrow cells using higher doses of chemotherapy or radiotherapy. These cells are replaced with healthy bone marrow or stem cells which help patients build a new and healthy immune system. Most patients now receive rituximab like the lymphoma patients.

2.4 Summary

In this chapter, we have given details of the background information of the data sets that will be used as illustrations in the project. We also gave an overview of the types of covariates in each of the data sets.

Chapter 3

Introduction to survival analysis

3.1 Introduction

This chapter describes some basic aspects of survival analysis, based on previously published work.

Survival analysis involves the modelling of time to event data which may include dealing with “incomplete” or “censored” data. For further background details see, for example Hosmer & Lemeshow (1999), Lee & Wang (2003) and Kleinbaum & Klein (2005).

Section 3.3 reviews preliminary work that has to do with relating covariates to the survival distribution using the proportional hazard and the accelerated life models. The form of the survival distribution could either be specified as a parametric or non-parametric form. Section 3.4 and 3.5 will review parametric forms such as the Weibull distribution and semi-parametric which combines a parametric model for the covariate effects and a non-parametric form for the baseline distribution such as a piecewise constant hazard model. We discuss investigating non-proportionality in survival analysis using a piecewise constant hazard model with time varying covariate effects. The hazard or survival function of an individual may depend on some unmeasured or unknown set of covariates or risk factors called frailties. We discuss with an illustration that frailties can cause a form of non-proportionality of hazard in survival modelling. Section 3.7 introduces prognostic indices as the logarithm of the hazard multiplier and as being used to predict the outcome in patients. We give a review of some prognostic indices across lymphoma subtypes in Section 3.8. Background knowledge about the International Prognostic Index (IPI) for aggressive lymphoma, Age-adjusted International Prognostic Index, Follicular Lymphoma International Prognostic Index, Revised International Prognostic Index, Mantle Cell International Prognostic Index and National Comprehensive Cancer Network International Prognostic Index will also be reviewed in Section 3.8.

3.2 Survival analysis

Survival analysis is a statistical methodology for analysing data for which the variable of interest is the time until an event occurs. The simplest case comes from how long people live until death. The event which terminates the time interval need not be death. Other examples of time to some events are time till failure of a machine, time till discharge from hospital after a surgery et.c. The event is often called a failure and the time taken till failure is called the failure or survival time. Survival analysis can also be used in many fields of study like medicine, biology, health, engineering, economics et.c.

Survival analysis is different from standard regression methods because times usually have skewed distributions and are always non-negative, variances usually depend upon covariates and survival times are often censored (see Section 3.2.1).

3.2.1 Censored data

Censoring is a nearly universal feature of survival data which occurs when we have some information about the individual but we do not know the exact survival or failure time. Censored data can also mean that the observations are incomplete or partially known. We could have left, right and interval censoring.

Right censoring occurs when the failure time is known to be larger than some given time. There are many reasons that make it difficult to get complete data in survival time. The most common reason is that the study finishes before all subjects experience the event of interest and we disregard when the subject really had the event. We assume that for an individual i , there is a lifetime T_i and a fixed censoring time C_i . If the exact lifetime of the individual, T_i is greater than or equal to C_i ($T_i \geq C_i$), the individual is a survivor and is right censored at C_i . Right censoring is common in almost all survival data.

In left censoring, the event of interest is known to have occurred before the censoring time. The exact failure time is less than the censoring time. The event of interest has already occurred for the individual before that person is observed in the study.

Interval censoring occurs when the interval of the failure time is known. Interval censoring occurs when the exact failure time lies between a lower and upper limit. Such censoring occurs when subjects have periodic follow-up and the event time is only known to fall in an interval.

3.2.2 The Survival Function

Let the survival time t_i of an individual i be a realisation of a non-negative random variable T_i with probability density function $f_i(t)$ and cumulative distribution function (cdf) $F_i(t)$.

Then, the lifetime distribution function of T_i , $F_i(t)$, is given by

$$F_i(t) = \Pr(T_i < t)$$

and the probability density function $f_i(t)$ is given by

$$f_i(t) = \frac{d}{dt}F_i(t)$$

The survival function $S_i(t)$ of an individual i can be defined as the probability that the individual survives longer than some specified time t where t ranges from 0 to ∞ (Lee & Wang, 2003). The survival function can be given as

$$S_i(t) = \Pr(T_i \geq t) = 1 - F_i(t) = \int_t^{\infty} f_i(t)dt$$

Survival functions decrease as the time t increases and hence, they are non-increasing. The survival function at time $t=0$ is 1. This means that, at the start of the study, no subject has reached the event yet so the probability of surviving past time 0 is one.

3.2.3 The Hazard Function

The Hazard function $h_i(t)$ of an individual i can be expressed mathematically as

$$\begin{aligned} h_i(t) &= \lim_{\Delta t \rightarrow 0} \frac{\Pr(T_i < t + \Delta t | T_i \geq t)}{\Delta t} \\ &= \lim_{\Delta t \rightarrow 0} \frac{F_i(t + \Delta t) - F_i(t)}{S_i(t)\Delta t} \\ &= \frac{f_i(t)}{S_i(t)} \end{aligned}$$

The cumulative hazard function $H_i(t)$ of an individual i is given by

$$H_i(t) = \int_0^t h_i(z)dz$$

The cumulative hazard function can be expressed in terms of the survival function since the hazard function can be written in terms of the survival function. It follows that

$$\begin{aligned} \frac{d}{dt} \log[S_i(t)] &= \frac{1}{S_i(t)} \times \frac{d}{dt}S_i(t) \\ &= -\frac{f_i(t)}{S_i(t)} \\ &= -h_i(t) \end{aligned}$$

We will have that

$$\log[S_i(t)] = -H_i(t)$$

Hence, the relationship between the survival function and the corresponding cumulative hazard function is given as

$$S_i(t) = \exp \{-H_i(t)\}$$

The relationship between the survival and hazard function in terms of integral is

$$S(t) = \exp \left\{ - \int_0^t h(u) du \right\}$$

and in terms of derivative is

$$h(t) = - \frac{1}{S(t)} \frac{dS(t)}{dt}$$

The hazard function is always non-negative and has no upper bound.

A survival model includes two features. One of them is how we relate the covariates to the distribution so that we can distinguish between individuals and the other is the form of the survival distribution.

3.3 Relating covariates to the survival distribution

The proportional hazard and accelerated life models are both ways of relating the covariates to the survival distribution.

3.3.1 Proportional hazards model

The proportional hazard model is the most commonly used method to relate the hazard function to the covariate values for an individual using the proportionality assumption. Cox (1972) proposed using the proportional hazard model in medical testing analysis and modelling the effect of explanatory variables on survival. Suppose we have S covariates for $s = 1, \dots, S$ and n individuals for $i = 1, \dots, n$. We denote the covariate vector for an individual i by $\underline{X}_i = (1, x_{i,1}, x_{i,2}, \dots, x_{i,S})$. These covariates may be continuous or discrete, categorical or even indicator variables (equal to 1 if present and 0 if absent). The proportional hazard model assumes that any two individuals i and j with the hazard function $h_i(t)$ and $h_j(t)$ at time t and covariate vectors $\underline{X}_i = (1, x_{i,1}, x_{i,2}, \dots, x_{i,S})$ and $\underline{X}_j = (1, x_{j,1}, x_{j,2}, \dots, x_{j,S})$, have their hazards related by

$$h_i(t) = \lambda_{i,j} \times h_j(t)$$

We have that $\lambda_{i,j}$ is a constant and does not depend on t . The proportional hazard model can also be written as:

$$h_i(t) = \lambda_i \times h_0(t) \tag{3.1}$$

where $h_0(t)$ is the baseline hazard function which is a function of time t but does not involve the covariates $\underline{X}_i = (1, x_{i,1}, x_{i,2}, \dots, x_{i,S})$. The quantity λ_i is the hazard multiplier which depends on (or is related to) the covariates of the individual i but not on the time variable t . We must have that $\lambda_i \geq 0$. This is usually done using a logarithmic link function to a linear predictor η_i . So,

$$\log \lambda_i = \eta_i = \beta_0 + \sum_{s=1}^S \beta_s x_{i,s} \quad (3.2)$$

where $x_{i,s}$ is the value of covariate s for subject i and β_0 is the baseline parameter. The linear predictor can then be used as a prognostic index.

3.3.2 Accelerated life model

The accelerated life model is another way of incorporating explanatory variables in a model where we relate the survival function to the covariates. Instead of scaling the hazard functions, we scale time in the survival function in the accelerated life model. In an accelerated life model, we assume that the survival function $S_i(t)$ for individual i with covariates vector $\underline{X}_i = (1, x_{i,1}, x_{i,2}, \dots, x_{i,S})$ takes the form

$$S_i(t) = S_0(\varphi_i t)$$

where $S_0(t)$ is the baseline survival function and φ_i is a positive constant which is called the acceleration factor which depends on the covariates of individual i . Hougaard (1999) compared the proportional hazard models with accelerated life models.

3.4 Parametric approach to survival modelling

We can choose either a parametric or a non-parametric form when specifying the form of the survival distribution. Models with a non-parametric form for the baseline distribution are often combined with a parametric model for the covariate effects giving a semi-parametric model.

We could specify the baseline hazard in a proportional hazard model or the baseline survival function in an accelerated life model by giving a standard parametric form, such as a Weibull distribution. This is called a parametric model. Parametric modelling offers relatively straightforward modelling and analysis techniques provided that the chosen distribution is appropriate and it makes efficient use of data. The commonly used parametric forms for lifetime distributions include the Weibull, Gompertz and log-logistic distributions.

3.4.1 Weibull distribution

The Weibull distribution (Weibull, 1951) is an important and widely used distribution in survival analysis. The Weibull distribution is a very flexible distribution for lifetimes data. It is a generalized form of the exponential distribution.

The distribution function of a Weibull distribution with scale parameter λ and shape parameter α is given by

$$F(t) = 1 - \exp \{-\lambda t^\alpha\}$$

It has a hazard rate that is monotone increasing if $\alpha > 1$, monotonic decreasing if $\alpha < 1$ or constant if $\alpha = 1$. The Weibull model reduces to an exponential model when $\alpha = 1$. The shape of the Weibull distribution depends upon the value of α which is the reason for referring to this parameter as the shape parameter. The probability density function of the Weibull distribution is given by

$$f(t) = \lambda \alpha t^{\alpha-1} \exp \{-\lambda t^\alpha\}$$

The scale parameter λ can be used to incorporate the covariates. The linear predictor of an individual (η_i) would be given as

$$\eta_i = g(\lambda_i)$$

where g is a known function called the link function which must be monotonic and differentiable. Usually,

$$\eta_i = \underline{X}_i^T \underline{\beta}$$

where \underline{X}_i^T is the transpose of the vector $\underline{X}_i = (1, x_{i,1}, x_{i,2}, \dots, x_{i,S})$ which are the covariates of the individual i and the vector of parameters $\underline{\beta} = (\beta_0, \beta_1, \dots, \beta_S)^T$

We could then have λ_i as

$$\begin{aligned} \lambda_i &= \exp \{\eta_i\} \\ &= \exp \{\underline{X}_i^T \underline{\beta}\} \end{aligned}$$

So, $g(\lambda_i) = \log \lambda_i$.

The survival function of the Weibull distribution with scale parameter λ and shape parameter α is given by

$$S_i(t|\lambda_i, \alpha) = 1 - F_i(t) = \exp \{-\lambda_i t^\alpha\} \quad (3.3)$$

The hazard function $h_i(t)$ of the Weibull distribution could also be determined since

$$\begin{aligned} h_i(t) &= \frac{f_i(t)}{S_i(t)} \\ &= \lambda_i \alpha t^{\alpha-1} \\ &= \lambda_i h_0(t) \\ h_0(t) &= \lambda_0 \alpha t^{\alpha-1} \end{aligned}$$

where $h_0(t)$ is the baseline hazard function. The simplicity of survival, hazard and probability density functions of the Weibull distribution has made it a very popular parametric model. It can also be used in either a proportional hazard or accelerated life model.

3.4.2 Likelihood functions

Suppose that we have n individuals with lifetimes governed by a survival function $S(t)$ with probability density function $f(t)$. We suppose that the i^{th} individual has an observation time t_i . The general form of the likelihood, where some observations are right censored is

$$L = \prod_{i \in E} f(t_i) \prod_{i \in C} S(t_i)$$

where C is the set of right censored individuals and E is the set of individuals that had the event. The probability density function is denoted by $f(t)$ and the survival function is also denoted by $S(t)$. The contribution of a right censored observation to the likelihood is $S(t_i)$. This is the probability that the patient is still alive at time t_i .

Let $D = (T, X, \delta, S, n)$ where:

X is a n by $(S + 1)$ matrix such that the i^{th} row of X is $(1, x_{i,1}, \dots, x_{i,S})$ given that $x_{i,1}, x_{i,2}, \dots, x_{i,S}$ are the covariate values for i^{th} individual,

n is the number of individuals,

S is the number of covariates used in the model,

$T = (t_1, t_2, \dots, t_n)^T$, where t_i is the event or censoring time for the i^{th} individual and

$\delta = (\delta_1, \dots, \delta_n)^T$, where δ_i is the event indicator which indicates whether the individual died or was right censored. We have

$$\delta_i = \begin{cases} 1 & \text{if the individual died} \\ 0 & \text{if censored} \end{cases}$$

Since $h(t) = \frac{f(t)}{S(t)}$, we have $f(t) = h(t)S(t)$. Therefore, the likelihood is

$$\prod h_i^{\delta_i}(t) S_i(t) = \left[\prod_E h_i(t) \right] \left[\prod_{E \cup C} S_i(t) \right]$$

Suppose that the lifetime random variable has a Weibull distribution with parameters (λ, α) . The probability density function of the i^{th} individual is denoted by

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

The survival function for the individual is given by

$$S(t|\lambda_i, \alpha) = \exp \{-\lambda_i t^\alpha\}$$

The likelihood contribution from the data is then given by

$$\begin{aligned} L(\beta, \alpha|D) &= \prod_{i \in E} f(t_i|\lambda_i, \alpha) \prod_{i \in C} S(t_i|\lambda_i, \alpha) \\ &= \prod_{i \in E} \lambda_i \alpha t_i^{\alpha-1} \prod_{i \in E \cup C} \exp \{-\lambda_i t_i^\alpha\} \\ &= \left[\prod_{i \in E} \lambda_i \right] \alpha^{n_D} \left[\prod_{i \in E} t_i^{\alpha-1} \right] \exp \left\{ - \sum_{i \in E \cup C} \lambda_i t_i^\alpha \right\} \end{aligned}$$

where n_D is the number of individuals in E .

3.5 Semi-parametric hazard models

A semi-parametric approach to specifying the hazard is often preferable to a fully parametric model (Sinha & Dey, 1997) to avoid specifying the time dependence parametrically and hence, mis-specifying the parametric form. We will discuss the piecewise constant hazard model and time varying covariate effects model as semi-parametric models.

3.5.1 The piecewise constant hazard (PCH) model

The piecewise constant hazard model is one of the most convenient and popular models for a semi-parametric approach to survival modelling. It is flexible and relaxes the assumption of a particular form for the baseline hazard by having sub-divided time where the baseline hazard $h_0(t)$ and the linear predictor are assumed constant in each interval. It also serves as a benchmark for comparisons with the fully parametric models for survival data. Breslow (1972, 1974) proposed the use of piecewise constant hazard model in survival data analysis and as a replacement of the proportional hazard model. The piecewise constant hazard model is sometimes called piecewise exponential because within each interval, the conditional distribution is exponential. The advantage of the piecewise constant hazard model is that the overall shape of the hazard function does not have to be imposed in advance. Kalbfleisch & Prentice (1973) suggested that the cut points should be selected

independently of the data. However, there is an arbitrary choice of cut points.

In the PCH model, the time t is partitioned into J disjoint intervals with $J - 1$ cut points given as $0 = \tau_0 < \tau_1 < \tau_2 < \dots < \tau_{J-1} < \tau_J = \infty$ (Ibrahim *et al.*, 2001). The j^{th} interval is defined as $[\tau_{j-1}, \tau_j)$ for $j = 1, 2, \dots, J$ with $\tau_0 = 0$ and $\tau_J = \infty$. The hazard is constant within each interval but is allowed to vary from one interval to another. Thus, the hazard function for individual i , $h_i(t)$ for each interval could be written as

$$h_i(t) = \begin{cases} h_{i,1}(t) & \text{for } 0 \leq t < \tau_1 \\ h_{i,2}(t) & \text{for } \tau_1 \leq t < \tau_2 \\ \vdots & \\ h_{i,J-1}(t) & \text{for } \tau_{J-2} \leq t < \tau_{J-1} \\ h_{i,J}(t) & \text{for } t \geq \tau_{J-1} \end{cases}$$

3.5.2 Time varying covariate effects

The proportional hazard model assumption has become widely used but it might not be true as the effects of covariates might vary over time. The proportional hazard model is unable to describe time varying covariate effects. If we ignore the time varying covariate effects and assume constant effects, the results for our model could be incorrect and hence, we have wrong conclusions. For instance, in the piecewise constant hazard models, we allow the baseline hazard to change at points but the coefficients of the covariates do not change. In this case, it is still a proportional hazard model. However, non-proportional hazards could arise if the covariate effects change over time. In some applications, especially cases with large population sizes, the effect of the covariates could be weaker if the follow up times were longer and the proportional hazard assumption might be violated.

3.5.3 Piecewise constant hazard model with time varying covariate effects

We can allow the effects of the covariates to vary over time. The piecewise constant hazard model with time varying covariates effects can be a very useful way to investigate the problem of non-proportionality present in the data. Gamerman (1991) allowed dependence on the covariates to change at points which makes the hazards non-proportional and allow time varying covariate effects. Martinussen & Scheike (2006) also developed dynamic survival models in which the coefficients varied over time.

If we suppose a piecewise constant hazard model with time varying covariate effects,

the hazard function $h_{i,j}(t)$ for individual i in the j^{th} interval for $j = 1, 2, \dots, J$ is given by

$$h_{i,j}(t) = h_{0,j}(t) \exp \left\{ \sum_{s=1}^S \beta_{j,s} x_{i,s} \right\}$$

where $x_{i,s}$ denotes the value of the covariate s for the i^{th} individual and $\beta_{j,s}$ is the regression parameter for the covariate s in the j^{th} interval.

The cumulative hazard function for a patient i , $H_i(t)$ is given by

$$\begin{aligned} H_i(t) &= \int_0^t h_i(t) dt \\ &= \sum_{j=1}^{k-1} \int_{\tau_{j-1}}^{\tau_j} h_{i,j}(t) dt + \int_{\tau_{k-1}}^t h_{i,k}(t) dt \\ &= \sum_{j=1}^{k-1} h_{i,j}(t)(\tau_j - \tau_{j-1}) + h_{i,k}(t)(t - \tau_{k-1}) \end{aligned}$$

The survival function for a patient i , $S_i(t)$ can then be given by

$$\exp \left\{ - \sum_{j=1}^{k-1} h_{i,j}(t)(\tau_j - \tau_{j-1}) - h_{i,k}(t)(t - \tau_{k-1}) \right\}$$

where $\tau_{k-1} \leq t < \tau_k$.

3.6 Frailty

In survival analysis, the hazard or survival function for an individual may be dependent on a set of covariates or risk factors which were not all known or measured (Ibrahim *et al.*, 2001). These unknown and unobserved covariates or risk factors are called frailties (Vaupel *et al.*, 1979). Frailties are very often or almost always present in survival data. If frailties are present but ignored this will cause underestimation of covariate effects. Frailties might also be a way to allow for differences between individuals and measurement errors or missing information in observed survival times. For instance, in medical studies, some patients will survive or stay healthy relatively long despite adverse observable risk factors whereas some others will survive shorter than expected. Henderson & Oman (1999) showed the inclusion of frailty in the survival model as unexplained heterogeneity in the survival data. They concentrated on frailty models as used to describe association between subjects in univariate survival analysis with members of a cluster sharing a common observable frailty effect. Sahu *et al.* (1997) examined semi-parametric and fully parametric approaches to frailty models.

Frailties have an effect on univariate (i.e only one-event) survival data. It can cause a form of non-proportionality of hazards when modelling one time event for a patient. Frailty could also be relevant when each individual has more than one event time since frailties allow for correlations within individuals. Burzykowski *et al.* (2008) and Booth & Eisenhauer (2012) introduced frailties as random effects in progression free survival (PFS) where we model the correlation between more than one time event which introduces dependence in times for an individual. For instance, times of event of a patient could be when a patient is treated and the disease stops developing (*known as remission*), time when the disease starts to progress again (*known as progression or relapse*) and death time. The length of time between remission and the start of progression is known as the progression free survival (PFS). The time events are not independent because the two time observations are for the same patient and the dependence can be modelled using frailties.

Suppose we have a proportional hazard model and the hazard function for an individual i with covariates $\underline{X}_i = (1, x_{i,1}, x_{i,2}, \dots, x_{i,S})$ is given by

$$h_i(t|\underline{X}_i) = h_0(t) \exp\{\underline{X}_i^T \underline{\beta}\}$$

where $h_0(t)$ is the baseline hazard function. Let $\underline{X}_i^T \underline{\beta} = \eta_i$ and $\lambda_i = \exp\{\eta_i\}$.

If we introduce the frailty z_i into the proportional hazard model, the conditional hazard function of the time given the unobserved frailty z_i for the i^{th} individual is given as follows

$$h_i(t|z_i) = h_0(t)\lambda_i z_i$$

The unobserved random variable Z with realisation z_i is assumed to act multiplicatively on the hazard $h(t)$ for a survival variable T . The conditional cumulative hazards $H_i(t|z_i)$ is given by

$$H_i(t|z_i) = H_0(t)\lambda_i z_i$$

The conditional survival function $S_i(t|z_i)$ is given by

$$S_i(t|z_i) = [\exp\{-H_0(t)\}]^{\lambda_i z_i} = [S_0(t)]^{\lambda_i z_i}$$

where $S_0(t)$ is the baseline survival function.

We can get the marginal survival function by marginalising over z by taking the expectation with respect to Z since the conditional survival function is a probability. We have that

$$S_i(t) = E_Z \left\{ [S_0(t)]^{\lambda_i z_i} \right\}$$

To take the expectation, we take the integral

$$S_i(t) = \int_0^\infty f_z(z_i) [S_0(t)]^{\lambda_i z_i} dz_i$$

where $f_z(z_i)$ is the probability density function of z_i .

The marginal survival function loses the property of proportionality. We can get the marginal survival probability density function $f_i(t)$ as

$$\begin{aligned} f_i(t) &= -\frac{d}{dt} S_i(t) \\ &= \int_0^\infty f_z(z_i) \lambda_i z_i [S_0(t)]^{\lambda_i z_i - 1} f_0(t) dz_i \\ &= f_0(t) \mathbb{E}_Z \left\{ \lambda_i z_i [S_0(t)]^{\lambda_i z_i - 1} \right\} \end{aligned}$$

Hence, the hazard function is

$$\begin{aligned} h_i(t) &= \frac{f_i(t)}{S_i(t)} \\ &= f_0(t) \frac{\mathbb{E}_Z \left\{ \lambda_i z_i [S_0(t)]^{\lambda_i z_i - 1} \right\}}{\mathbb{E}_Z \left\{ [S_0(t)]^{\lambda_i z_i} \right\}} \\ &= h_0(t) \frac{\mathbb{E}_Z \left\{ \lambda_i z_i [S_0(t)]^{\lambda_i z_i} \right\}}{\mathbb{E}_Z \left\{ [S_0(t)]^{\lambda_i z_i} \right\}} \end{aligned} \tag{3.4}$$

Illustration: We will illustrate this with an example to show that including frailties when modelling can cause non-proportionality. We suppose that $S_0(t) = \exp\{-t\}$. From 3.4, we have that

$$h_i(t) = h_0(t) \frac{\mathbb{E}_Z \left\{ \lambda_i Z_i \exp(-\lambda_i Z_i t) \right\}}{\mathbb{E}_Z \left\{ \exp(-\lambda_i Z_i t) \right\}}$$

Suppose $Z_i \sim \text{Ga}(a, b)$ then,

$$\mathbb{E}_Z \left\{ \exp(-\lambda_i Z_i t) \right\} = \int_0^\infty f_Z(Z_i) [S_0(t)]^{\lambda_i Z_i} dZ_i$$

and

$$\mathbb{E}_Z \left\{ \exp(-\lambda_i Z_i t) \right\} = \frac{b^a}{(\lambda_i t + b)^a}$$

Similarly,

$$\mathbb{E}_Z \left\{ \lambda_i Z_i \exp(-\lambda_i Z_i t) \right\} = \frac{a \lambda_i b^a}{(\lambda_i t + b)^{a+1}}$$

So

$$h_i(t) = h_0(t) \frac{\lambda_i a}{\lambda_i t + b}$$

and

$$\frac{h_i(t)}{h_j(t)} = \frac{\lambda_i(\lambda_j t + b)}{\lambda_j(\lambda_i t + b)}.$$

Let $m = a/b$, $v = a/b^2$ so $b = m/v$. Then let $m = 1$.

Now,

$$\frac{h_i(t)}{h_j(t)} = \frac{\lambda_i(v\lambda_j t + 1)}{\lambda_j(v\lambda_i t + 1)}. \quad (3.5)$$

Equation 3.5 is not constant in t so it is not proportional.

3.7 Introduction to prognostic indices

Prognostic models are used for predicting the outcome in patients on the basis of the clinical information of the patient usually before treatment. The prognostic indices can be used to estimate the length of an individual's survival. It helps to take clinical decisions and helps doctors choose an appropriate treatment for the patients. The prognostic indices could help in creating clinical risk groups which stratify patients by the severity of the disease.

Most researchers use the proportional hazard model to analyse time to event and survival prediction by estimating hazard ratios which measures how much a covariate affects the hazard function for an event of interest. The prognostic index is the main product of a proportional hazard model. The obvious way to construct the prognostic index is to use the linear predictor from a proportional hazard model. The linear predictor is a weighted sum of the prognostic variables in the model, where the weights are the regression coefficients (see Equation 3.2). The prognostic index of an individual is the logarithm of the hazard multiplier. High values of the prognostic index indicate a worse prognosis or adverse outcome for the event of interest. Mallett *et al.* (2010) reviewed current practice in methods used to develop and evaluate the performance of prognostic indices and risk groups from the prognostic models. It was summarised that 94% of the studies used proportional hazard model where the effects of the covariates were assumed to have a linear functional form. The assumption of proportional hazard was not applied in the remaining 6% but instead the recursive partitioning analysis (RPA) and artificial neural network (ANN). The recursive partitioning analysis (RPA) can be used to classify patients by splitting them into sub-populations according to their variables. It allows varying misclassification in order to create a decision rule. One of the disadvantages of RPA is that it does not work well for continuous variables (Gordon *et al.*, 1984). Artificial neural networks (ANN) can be used to approximate or estimate non-linear functions of variables (Naguib & Sherbet, 2001). Chih-Lin *et al.* (2007) used ANN to solve the complex relationship among variables and predict the time of reoccurrence of a breast cancer.

3.8 Review of prognostic indices across lymphoma subtypes

Haematologists have developed a number of prognostic indices because of the large number of lymphoma patients. The most widely used lymphoma prognostic indices are the International Prognostic Index (IPI) for aggressive lymphoma and Follicular Lymphoma International Prognostic Index (FLIPI) for follicular lymphoma. Other indices will also be discussed.

3.8.1 International Prognostic Index (IPI)

The Ann Arbor stage was used to assess prognosis but because of its inadequacy in predicting survival outcomes, the International Prognostic Index (IPI) was developed in 1993 to help doctors determine the outlook for people with fast growing lymphoma. Shipp *et al.* (1993) developed the International Prognostic Index (IPI) for predicting the outcome in patients with aggressive non-Hodgkins lymphoma based on the universally recognised clinical covariates before treatment. The IPI identified four risk groups of patients: low, low intermediate, high intermediate and high. These groups depend on five binary variables which include age ≥ 60 , high serum LDH, stage 3 or 4, ECOG 2,3 or 4 and more than 1 extranodal site. The adverse indications or “risk factors” include age > 60 , stage III or IV, lymphoma is in more than 1 organ of the body outside of lymph nodes and the performance status is that the patient needs a lot of help with daily activities.

The IPI ranges from point 0-5 and with one point assigned for each of the five risk factors. The IPI divides patients into four risk groups which are summarised in Table 3.1 with the IPI point, risk group and 5-year survival probability based on an analysis performed on 2031 patients with aggressive non-Hodgkins lymphoma.

Table 3.1: Summary of the IPI point, risk group and 5-year survival

IPI point	Risk group	5-year survival(%)
0 -1	low	73%
2	low intermediate	51%
3	high intermediate	43%
4-5	high	26%

3.8.2 Age-adjusted International Prognostic index (AaIPI)

The Age-adjusted International Prognostic index (AaIPI) (Shipp *et al.*, 1993) differentiates the prognosis for younger patients from the older ones. A simplified index was used to compare patients within an age group (i.e 60 or younger or over 60) with one point assigned for each of the following three risk factors: stage, LDH and performance status (ECOG).

Shipp *et al.* (1993) used the Age-adjusted International Prognostic index as a simplified model for 1274 younger patients (age ≤ 60) by using the clinical covariates. Table 3.2 summarises the AaIPI point, risk group and 5-year survival.

Table 3.2: Summary of the AaIPI point, risk group and 5-year survival

AaIPI point	Risk group	5-year survival(%)
0 -1	low	83%
1	low intermediate	69%
2	high intermediate	46%
3	high	32%

3.8.3 Follicular lymphoma International Prognostic Index (FLIPI)

The IPI is useful for most lymphomas but it is not as helpful for the follicular lymphoma which is slower growing. Follicular lymphoma accounts for one third of non-Hodgkins lymphoma in adults and it is the second most frequent subtype of non-Hodgkins lymphoma (Solal-Celigny, 2004).

The FLIPI was developed based on 4167 diagnosed patients with five adverse prognostic factors selected as follows: age ($> 60; \leq 60$), stage(*III – IV; I – II*), haemoglobin level(HB)($< 120g/l; \geq 120g/l$), extranodal (> 4 nodal areas; ≤ 4 nodal areas) and LDH (abnormal ; normal). Solal-Celigny (2004) defined three risk groups; low risk, intermediate risk and high risk which are summarised in the Table 3.3:

Table 3.3: Summary of FLIPI score, risk group and 5-year survival

FLIPI score	Risk group	5-year survival(%)
0 -1	low	90.6%
2	intermediate	77.6%
≥ 3	high	52.5%

3.8.4 Revised International Prognostic Index (R-IPI)

The Revised International Prognostic Index (R-IPI) is a more recent IPI which is based on patients with lymphomas who have received a new modern treatment called rituximab. The RIPI depends on the same factors as IPI but divides patients into three distinct prognostic groups. These groups are summarised as very good (if no poor prognostic factors), good (if 1 or 2 poor prognostic factors) and poor (if 3 or more poor prognostic

factors). Sehn *et al.* (2007) published the R-IPI with an analysis of 365 patients with DLBCL non-Hodgkins Lymphoma using the IPI factors which is summarised in Table 3.4.

Table 3.4: Summary of R-IPI score, risk group and 4-year survival

R-IPI score	Risk group	4-year survival(%)
0	very good	94%
1,2	good	79%
3,4,5	poor	55%

3.8.5 Mantle cell International Prognostic Index (MIPI)

Hoster *et al.* (2008) developed the prognostic index based on 455 patients with advanced stage mantle cell lymphoma since the IPI and FLIPI were developed for the diffuse large B-cell and follicular lymphoma respectively. According to the Mantle cell International Prognostic Index (MIPI), Age, ECOG, LDH and WBC counts were the four independent prognostic factors used in classifying patients for the overall survival. Hoster *et al.* (2008) calculated the MIPI prognostic score as

$$s = 0.03535x_1 + 0.6978x_2 + 1.367 \log_{10} x_3 + 0.9393 \log_{10} x_4$$

where s is the MIPI score

x_1 is the age in years

x_2 is the ECOG value used if it is greater than 1

x_3 is the LDH

x_4 is the white blood cell count

According to Hoster *et al.* (2008) the patients were classified into three risk groups; low, intermediate and high risk groups on the basis of the four prognostic factors as summarised in Table 3.5. Table 3.5 summarises the risk groups, the percentage of the patients and the median overall survival.

Table 3.5: Summary of MIPI risk groups, the percentage of the patients and the median overall survival

Risk group	(%) of patients	median overall survival
low	44%	median overall survival time not reached
intermediate	35%	51 months
high	21%	29 months

3.8.6 National Comprehensive Cancer Network International Prognostic Index (NCCN-IPI)

In recent times, there has been a great effort to improve the prognostic value of the original IPI. Until the development of the National Comprehensive Cancer Network International Prognostic Index (NCCN-IPI), the original IPI was a powerful prognostic tool for determining the prognosis among patients with DLBCL for over 20 years. Clinical data were collected from 1650 adult patients diagnosed with DLBCL from 2000-2010 who were treated with rituximab from seven National Comprehensive Cancer Network (NCCN) centers (Zhou *et al.*, 2014). The NCCN-IPI used five variables which include age, LDH, sites of involvement (extranod), stage and ECOG with a maximum of 8 points assigned with each having a score of 1. Table 3.6 summarises the 8 scoring points for the NCCN-IPI.

Table 3.6: Summary of the 8 scoring points for the NCCN-IPI

NCCN-IPI variable	score
$40 < \text{Age} \leq 60$	1
$60 < \text{Age} \leq 75$	2
$75 < \text{Age}$	3
$1 < \text{LDH ratio} < 3$	1
≥ 3	2
Stage 3,4	1
Extranodal disease	1
$\text{ECOG} \geq 2$	1

The model was validated using a population-based registry cohort from the British Columbia Cancer Agency (BCCA). Table 3.7 summarises 4 distinct risk groups (low, low intermediate, high intermediate and high) with comparison with the IPI with 5-years overall survival (OS) and progression free survival (PFS) (see Section 3.6).

The BCCA cohort supported the generalisability of the NCCN-IPI. Both NCCN-IPI and the original IPI included the same set of clinical variables and recognised four risk groups. In NCCN-IPI, the variable age was categorised and LDH was normalised to capture the associated increased risk of mortality. Hence, in recent times the NCCN-IPI is used to stratify the prognostic index of DLBCL patients. The application of NCCN-IPI is easy and more powerful than the original IPI for predicting survival of patients.

3.9 Summary

In this chapter, we have discussed the basic aspects of survival analysis. We showed how the covariates are related to the survival distribution using the proportional hazards and

Table 3.7: Summary of 4 distinct risk groups with comparison with the IPI with 5-years overall survival (OS) and progression free survival (PFS) with NCCN and BCCA cohort

NCCN cohort			NCCN	IPI	NCCN	IPI
Risk Group	NCCN score	IPI score	5yr OS	5yr OS	5yr PFS	5yr PFS
Low	0-1(19%)	0-1(38%)	96%	90%	91%	85%
L.I	2-3(42%)	2(26%)	82%	77%	74%	66%
H. I	4-5(31%)	3(22%)	64%	62%	51%	52%
High	≥ 6 (8%)	4-5(14%)	33%	54%	30%	39%
BCCA cohort			NCCN	IPI	NCCN	IPI
Risk Group	NCCN score	IPI score	5yr OS	5yr OS	5yr PFS	5yr PFS
Low	0-1(12%)	0-1(33%)	96%	84%	94%	81%
L.I	2-3(37%)	2(22%)	77%	72%	72%	66%
H.I	4-5(37%)	3(22%)	56%	54%	54%	54%
High	≥ 6 (14%)	4-5(21%)	38%	43%	35%	41%

the accelerated life model using a parametric distribution like the Weibull distribution. We discussed a semi-parametric model like the piecewise constant hazard model where we could have time varying covariate effects if the covariate effects varied over time. We illustrated with an example how frailties could cause non-proportionality of hazards. We also gave some existing prognostic indices in lymphoma.

Chapter 4

Introduction to Bayesian inference

4.1 Introduction

This chapter introduces Bayesian inference for reference in later chapters. This is largely based on previously published work but we also introduce some new ideas on the construction of prior distributions for the effects of factors.

Bayesian inference requires the combination of prior experience (which is in the form of prior probability) and the observed data (which is in the form of a likelihood). Therefore, we will discuss the specification of prior information about the model parameters in Section 4.4.1. We will review prior elicitation as a process which helps an expert in the field to express his or her knowledge in a probabilistic form.

We will present a way of specifying the covariances between parameters by thinking in terms of the coefficient of determination and we get the correlation between the parameters. We incorporate our prior beliefs into the construction of the covariance matrix of the parameters. We will also present some structures of prior distributions which are used to construct priors depending on the type of variable. We discuss the construction of prior for the coefficient of a quantitative covariates in the context of a general linear model following an example in Farrow (2011). We will review the requirement of exchangeability and identifiability of the parameters of a model. We will consider some alternatives of reparameterisation of the quantities of the model by either having a zero-sum constraint or a corner constraint to avoid over-parameterisation.

We present a way of constructing priors for the coefficient of a 2-level factor following the construction of prior distribution of the coefficient of a quantitative covariate. We will also present new ideas on the construction of prior distributions for the effects of unordered categorical covariates by making the parameters identifiable by having a zero-sum constraint using some orthogonal contrasts scheme when the number of levels is a power of 2 and a general scheme for a case of any number of levels. We will also present

new ideas with illustrations of how we might construct the prior for a categorical factor where we want all levels of the factor to be exchangeable using the general scheme and how we can transform $m-1$ uncorrelated random variables into m random variables which are constrained to sum to zero for $i = 1, \dots, m$ levels of factors. We will show that the variances of the parameters are equal and covariances between pairs of parameters are also equal to each other.

There was the need to build a set of ordered random variable with a zero-sum constraint. We present new ideas to do this where the means for neighbouring categories are more positively correlated than those of categories further apart. We considered using a correlation matrix and modified them so that they have any required means and variances. We give some numerical illustrations for some numbers of levels on how we choose the elements of the correlation matrix depending on the distance and not violating some defined conditions.

Having discussed the structures of prior distributions, we recall that Bayesian inference often involves calculations which are analytically intractable. These are typically done using Markov chain Monte Carlo methods (MCMC). We will review these methods in Section 4.6. The methods include Metropolis and Metropolis Hastings algorithm, Gibbs sampler and Metropolis within Gibbs algorithm.

We will give a clear discussion on how sampling from a multivariate normal distribution can be done using the Cholesky decomposition of the covariance matrix as a fast and numerically stable way. Section 4.7 reviews two numerical methods for improving mixing in MCMC since the Markov chain may move around very slowly in the posterior distribution. These methods include sampling the parameters using the idea of the forward backward algorithm where the forward recursion accumulates information in the form of a conditional distribution and the backward recursion updates the distribution of the parameters after the information has been collected from the observed data. The other method is the Gaussian approximation of the proposal distribution of the parameters. This is done by expanding around the mean of the approximation to generate an improved approximation until it is equal to the mode of the full conditional distribution. We discuss using the Newton Raphson step and the procedure is iterated until the parameter values stabilise. We approximate a multivariate normal distribution as the proposal distribution using the mode as the mean and the value of the partial second derivative evaluated at the mode to obtain the precision matrix.

We review data missingness and types of missingness which include missing at random (MAR), missing completely at random (MCAR) and when the missing data mechanism is said to be ignorable.

4.2 Bayes' Theorem and Bayesian Inference

Bayes' theorem can be used to show the relationship between two conditional probabilities that are the reverse of each other. In general, Bayes' theorem combines the prior experience (in the form of a prior probability) with observed data (in the form of a likelihood) to interpret these data (in the form of a posterior distribution) in a process known as Bayesian inference. Bayes' theorem is the probabilistic result which plays a central role in Bayesian inference.

Suppose we have data $D = (x_1, x_2, \dots, x_n)$ which we model using the probability density function $f(D|\theta)$ for an unknown parameter θ . The likelihood of the parameter θ is given by

$$\begin{aligned} L(\theta|D) &= f(D|\theta) \\ &= \prod_{i=1}^n f_i(x_i|\theta, x_1, \dots, x_{i-1}) \end{aligned}$$

where $f_i(x_i|\theta, x_1, \dots, x_{i-1})$ is the conditional probability density function of x_i given x_1, \dots, x_{i-1} and θ .

If x_1, \dots, x_n are independent given θ then,

$$f_i(x_i|\theta, x_1, \dots, x_{i-1}) = f_i(x_i|\theta)$$

Hence

$$L(\theta|D) = \prod_{i=1}^n f_i(x_i|\theta)$$

The prior beliefs about θ , with no reference to the data, can be expressed in the form of the probability density function $\pi(\theta)$. The posterior distribution combines the likelihood and the prior and captures all we know about the parameters. Then, the posterior probability density function for θ , $\pi(\theta|D)$, which summarises our beliefs about θ after seeing the data, is derived from the joint density $\pi(\theta)L(\theta|D)$ according to Bayes' formula and is given by

$$\pi(\theta|D) = \frac{\pi(\theta)L(\theta|D)}{\int_{\Theta} \pi(\theta)L(\theta|D)d\theta} \quad (4.1)$$

The integral $\int_{\Theta} \pi(\theta)L(\theta|D)d\theta$ is not a function of θ and so the posterior distribution would be given as

$$\pi(\theta|D) \propto \pi(\theta) \times L(\theta|D)$$

which is often expressed as

$$\text{Posterior} \propto \text{Prior} \times \text{Likelihood}.$$

The integral $\int_{\Theta} \pi(\theta)L(\theta|D)d\theta$ is known as the normalising constant. It is also known as the marginal likelihood of the data (Berger, 1985). In many models and applications, the integral does not have an analytical closed form. The difficulty of not having a closed form has led to the use of sampling from the posterior distribution $\pi(\theta|D)$.

4.3 Monte Carlo integration

The posterior distribution can sometimes not be represented analytically in realistically complex problems because of the intractability of the normalising constant. This created an obstruction to the implementation of the Bayesian approach until the development of suitable numerical methods.

Suppose we want to evaluate an integral

$$\int_G \Phi(x)dx$$

for which there is no closed analytic solution. If we suppose that, for some density function f , $\Phi(x)$ has the form

$$\Phi(x) = \widehat{\Phi}(x)f(x),$$

then, the integral has the form

$$\int_G \Phi(x)dx = \int_G \widehat{\Phi}(x)f(x)dx = E[\widehat{\Phi}(X)]$$

where X is a random variable which has probability density function f . If we know how to simulate realisations x_1, \dots, x_n of X , then we will have an estimate for the integral which is given by

$$\int_G \Phi(x)dx = E[\widehat{\Phi}(X)] \approx \frac{1}{n} \sum_{i=1}^n \widehat{\Phi}(x_i) = \widehat{I}$$

This method of approximating integrals is known as Monte Carlo integration. An estimate should converge to the true value in expectation. We consider the variance of the estimate as

$$\begin{aligned} \text{Var}(\widehat{I}) &= \text{Var}\left[\frac{1}{n} \sum \widehat{\Phi}(x_i)\right] \\ &= \text{Var}[\widehat{\Phi}(X)] \\ &= \frac{1}{n} \int_G f(x) \left(\widehat{\Phi}(x) - E[\widehat{\Phi}(x)]\right)^2 dx \\ &= \frac{1}{n} \int_G f(x) \left(\widehat{\Phi}(x) - \int_G \widehat{\Phi}(y)f(y)dy\right)^2 dx \end{aligned}$$

We cannot evaluate the above integral so an easier way is to estimate the variance by

working out the standard error (s.e) of the estimator \hat{I} . So,

$$(s.e)^2 = \frac{1}{n-1} \sum_{i=1}^n (\hat{\Phi}(x_i) - \hat{I})^2$$

Monte Carlo integration draws samples from a distribution and then forms sample averages to approximate expectations. It can be used to calculate the normalising constant in Bayesian analysis.

4.4 Prior distributions

A Bayesian analysis requires the specification of prior information about the model parameters by expressing beliefs about the parameters in the form of a probability distribution before we look at the observations. The prior distribution should reflect information about the model parameters. The prior information is often an opinion or subjective belief of an “expert” within the field of investigation from whom information is being elicited. This is appropriate, for example, when the purpose of the analysis is to inform a decision which must be made. In other cases, the purpose may be simply to communicate the results of a scientific investigation. In such cases, one or more “reasonable” prior specification may be used.

4.4.1 Prior elicitation

Prior elicitation is a process whereby a person’s knowledge or belief about some uncertain quantities are expressed in the form of a probability distribution of the quantities before the data are taken into account. Hence, the person’s personal beliefs about the unknowns are given in the form of a probability distribution. The prior distribution is an important issue in Bayesian inference because it is used in the calculation of the posterior distribution which brings the analyst closer to what is being modelled. Elicitation helps expert in the field to express their current knowledge in a probabilistic form (Garthwaite *et al.*, 2005). We find ways of using the limited specification of beliefs from the experts when trying to fit a prior distribution for a parameter. In many cases, we will elicit the first and second order moments (i.e means, variances and covariances).

In the case of multivariate elicitation, we elicit beliefs about two or more unknown parameters. These parameters may or may not be independent. If the parameters are independent, information about some of the parameters would not affect the beliefs about the other parameter. The joint probability density for these parameters will be the product of their marginal densities. In cases where the parameters are dependent, the dependence can be expressed in terms of correlation by eliciting either the correlation coefficient or covariances. The correlation between parameters can be directly specified as between the

values -1 and $+1$ (Clemen *et al.*, 2000). More generally, we might assess the conditional distribution of one unknown given another. Revie *et al.* (2010) investigated different methods of eliciting prior correlations with applications to Bayes linear modelling of the reliability of defence systems.

The prior probability distributions are often classified into either informative or uninformative priors. A prior is informative when prior information about the parameter is available and included in the prior distribution. An expert's opinion is a form of an informative prior because the expert provides honest information of the present state of knowledge before updating with the new information (James *et al.*, 2010). In the absence of an expert's knowledge, prior information may be obtained from earlier studies. We could also have an informative prior when the posterior distribution of the previous model which is similar to the form of a present model is used as the prior distribution of the present. In this case, the present model is not starting from scratch but based not only on the present data but the cumulative effects of past and present data are taken into account.

Berger *et al.* (2009) categorised prior distributions into four categories according to information and the goal in the use of the prior. These categories were informative, weakly informative, least informative, and uninformative prior. Priors which have little effect on the posterior are sometimes called uninformative priors. Irony & Singpurwalla (1997) explained that truly uninformative priors do not really exist and that all priors are informative in some way. The weakly informative prior uses less prior information than is actually available. It provides some of the benefit of prior information while avoiding some of the risk from using information that is not generally agreed. Selecting a weakly informative prior is difficult but these are the most commonly used priors in practice. In some cases, we may choose to use a prior distribution which is non-prejudicial about some aspect of the unknowns. For example, in a clinical trial to compare the effects of two treatments, we might choose to give the effects of the two treatments the same marginal prior distributions to avoid favouring one over the other a priori. Gosling (2014) suggested that a criterion or standardized method for the judgement of an elicitation exercise in health technology assessment (HTA) can be produced through a joint discussion on the expectation of elicitation exercises between experts in the process of elicitation, health technology assessors and policy makers.

4.4.1.1 Construction of prior covariances between parameters

We might want to think of the degree of dependency between two variables. Eliciting covariances between parameters directly could be difficult as experts do not think in terms of covariances. Revie *et al.* (2010) discussed four different methods of determining the dependency between variables. These methods include direct calculation, direct elicitation

of correlation, adjusted expectation and adjusted uncertainty.

We will want to specify the covariances between pairs of parameters by incorporating our prior beliefs into the parameterisation and construction of the covariance matrix. We could make the covariance between parameters zero if we reasonably think that the parameters of a model are independent of each other in our beliefs. One way of thinking of how to get actual values for the prior covariance between any two parameters is to think of the correlation between the parameters. We could also think in terms of the coefficient of determination which is the square of the correlation between the parameter. The coefficient of determination would be the proportion of the variance of one of the parameter that we will lose if we knew the value of the other parameter. If knowing the value of one parameter does not affect our belief about another parameter, then we have a covariance of zero. This means that the parameter is unable to explain some of the uncertainty by knowing the other. For instance, we will suppose that if we knew one parameter then 50% of the proportion of the variance of the other parameter will be lost. Then, the coefficient of determination (r^2) is 0.50 and the correlation will be 0.7. We will have to decide on the sign of the correlation by thinking if the parameter is bigger in the same direction with the other parameter in which case we have a positive sign. The covariance between the parameters X and Y can then be calculated using

$$\text{Covar}\{X, Y\} = \sigma_X \sigma_Y r \tag{4.2}$$

where σ_X and σ_Y are the standard deviation of X and Y respectively and r is the correlation between the parameters.

4.4.2 Centering of covariates

Quantitative variables can sometimes be standardized or centered (i.e subtracting a central value such as the mean). Centering of covariates is important in prior elicitation. It is good to center the covariates because it makes it easier to choose a sensible prior. The constant (intercept) will correspond to a plausible case thereby making elicitation more realistic. Furthermore, it will then be more reasonable for the intercept and the coefficient of the covariate to be independent in the prior. For instance, we could center the covariate age, x_{age} by using $x_{\text{age}} = a - 60$ where a is the patient's age in years. This should be done throughout.

4.4.3 Structures for prior distributions

When we have many related parameters in a model, it is usually best to build a qualitative structure for their joint prior distribution first before quantifying prior means, variances et.c. (Farrow, 2003). We will construct priors depending on the type of variable. We

will frame our discussion in this chapter in the context of a general linear model. The construction of a prior distribution would be specified depending on the type of covariate.

4.4.3.1 Construction of prior distribution for the coefficient of a quantitative variable

We discuss the construction of prior for the coefficient of a quantitative variable by considering a regression example in Farrow (2011).

We will suppose a simple regression model

$$Y = \alpha + \beta x + \epsilon$$

for a pair of values (x_i, y_i) for a group of people, where y_i is the height in inches and x_i is the shoe size.

Suppose we wish to predict the height of someone if we know the person's shoe size. We could express this model by saying that the conditional mean height for someone with shoe size x is $\alpha + \beta x$. It will follow that someone with a shoe size of zero will have a mean height of α . It is difficult to think about our prior beliefs about α since there is no person with shoe size of zero. We refer to the reason for centering discussed in Section 4.4.2. So, we elicit a mean height of a realistic shoe size. Hence, we will think of a reference value x_{ref1} as the shoe size 11. We let

$$z = x - x_{ref1} = x - 11$$

The regression equation becomes

$$Y = \tilde{\alpha}_1 + \beta z + \epsilon$$

where $\tilde{\alpha}_1 = \alpha + \beta x_{ref1} = \alpha + 11\beta$.

We have that $\tilde{\alpha}_1$ represents the mean height of people who have size 11 shoes. As a guide to the likely values of $\tilde{\alpha}_1$, we will suppose that the height of the person who has shoe size 11 is 74 inches. So, $\tilde{\alpha}_1$ has a normal prior distribution with mean 74. We also have to think of the standard deviation for $\tilde{\alpha}_1$ since 74 might not exactly be the average height for people with shoe size 11. We will suppose that with probability about 0.95, 74 is not more than six inches from the conditional mean and so the standard deviation is 3.

We will need a prior distribution for β . We will choose another reference shoe size $x_{ref2} = 5$ with a 64 inches height for the person. So,

$$\tilde{\alpha}_2 = \alpha + \beta x_{ref2} = \alpha + 5\beta$$

We suppose that $\tilde{\alpha}_2$ has a normal prior distribution $\tilde{\alpha}_2 \sim N(64, 9)$

We will make $\tilde{\alpha}_1$ and $\tilde{\alpha}_2$ independent in our prior distribution. We let

$$\underline{\tilde{\beta}} = \begin{pmatrix} \tilde{\alpha}_1 \\ \tilde{\alpha}_2 \end{pmatrix}$$

We will give $\underline{\tilde{\beta}}$ a bivariate normal prior distribution with mean

$$\underline{\tilde{M}} = \begin{pmatrix} \mathbf{E}(\tilde{\alpha}_1) \\ \mathbf{E}(\tilde{\alpha}_2) \end{pmatrix} = \begin{pmatrix} 74 \\ 64 \end{pmatrix}$$

and variance

$$V_0 = \begin{pmatrix} \text{Var}(\tilde{\alpha}_1) & \text{Covar}(\tilde{\alpha}_1, \tilde{\alpha}_2) \\ \text{Covar}(\tilde{\alpha}_1, \tilde{\alpha}_2) & \text{Var}(\tilde{\alpha}_2) \end{pmatrix} = \begin{pmatrix} 9 & 0 \\ 0 & 9 \end{pmatrix}.$$

We will have that

$$\beta = \frac{\tilde{\alpha}_1 - \tilde{\alpha}_2}{x_{\text{ref1}} - x_{\text{ref2}}}$$

and

$$\alpha = \frac{\tilde{\alpha}_2 x_{\text{ref1}} - \tilde{\alpha}_1 x_{\text{ref2}}}{x_{\text{ref1}} - x_{\text{ref2}}}$$

We will have that

$$\underline{\beta} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \frac{1}{x_{\text{ref1}} - x_{\text{ref2}}} \begin{pmatrix} -x_{\text{ref2}} & x_{\text{ref1}} \\ 1 & -1 \end{pmatrix} \underline{\tilde{\beta}} = H \underline{\tilde{\beta}}.$$

Thus α, β have a bivariate normal prior distribution with mean

$$M_0 = H \underline{\tilde{M}} = \frac{1}{6} \begin{pmatrix} -5 & 11 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 74 \\ 64 \end{pmatrix} = \begin{pmatrix} 55.7 \\ 1.7 \end{pmatrix}$$

and variance

$$V_0 = H \underline{\tilde{V}}_0 H^T = \frac{1}{36} \begin{pmatrix} -5 & 11 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 9 & 0 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} -5 & 11 \\ 0 & 9 \end{pmatrix} = \begin{pmatrix} 36.5 & -4 \\ -4 & 0.5 \end{pmatrix}.$$

We would want to show what happens when we center the variable. Suppose we want to center the covariate at x_0 , then we have that

$$\begin{aligned} Y &= \alpha + \beta x_0 + \beta(x - x_0) \\ &= \alpha^* + \beta^*(x - x_0) \end{aligned}$$

If we suppose that $x_0 = 8$ is the central value then the variance of the modified parameters

is

$$\text{Var} \begin{pmatrix} \alpha^* \\ \beta^* \end{pmatrix} = \begin{pmatrix} 1 & x_0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 36.5 & -4 \\ -4 & 0.5 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ x_0 & 1 \end{pmatrix} = \begin{pmatrix} 4.5 & 0 \\ 0 & 0.5 \end{pmatrix}$$

4.4.3.2 Exchangeability and identifiability

Suppose that we have observations Y_1, \dots, Y_n on individuals $i = 1, \dots, n$. The observations are said to be exchangeable if the information that the quantities provide is independent of the order in which they are collected and hence it is a property of symmetry with respect to order.

Suppose now that observations made on individuals are in J groups. Observation $Y_{i,j}$ is the i^{th} individual in the j^{th} group for $i = 1, \dots, n_j$. Often, we would regard individuals within a group as exchangeable. We might model this by saying that the distribution of observation within a distribution within group j is such that, conditional on the value of an unknown parameter, $\theta_j, Y_{1,j}, \dots, Y_{n_j,j}$ are independent and identically distributed with a distribution which has parameter θ_j .

Now consider a collection of observation, one from each group. Without loss of generality, suppose that these are $Y_{1,1}, \dots, Y_{1,J}$. An expert may well also consider these to be exchangeable. We can elicit beliefs about the parameters $\theta_1, \dots, \theta_J$ by asking suitable questions about observations within the groups. For example, suppose that $Y_{i,j} | \theta_j \sim N(\theta_j, 1)$. Then, $E(\theta_j) = E(Y_{i,j})$ and $E(\theta_j^2) = E(Y_{i,j} Y_{k,j})$ where $i \neq k$. Hence, $\text{Var}(\theta_j) = E(Y_{i,j} Y_{k,j}) - [E(Y_{i,j})]^2$. Now, the requirement that $Y_{1,1}, \dots, Y_{1,J}$ are exchangeable leads to the requirement that beliefs about $\theta_1, \dots, \theta_J$ are such that $\theta_1, \dots, \theta_J$ are also exchangeable.

Now, we consider a case where we group individuals in more than one way. For simplicity, suppose that we group individuals in two ways. So, $Y_{i,j,k}$ is the observation on the i^{th} individual in group j, k . That is, level j of the first factor and level k of the second factor, where $j = 1, \dots, J$ and $k = 1, \dots, K$. In the same way as above, we might use parameters $\theta_{j,k}$. Now, we might well require $Y_{1,1,k}, \dots, Y_{1,J,k}$ to be exchangeable for any k and $Y_{1,j,1}, \dots, Y_{1,j,K}$ to be exchangeable for any j but not that $Y_{1,1,k}, \dots, Y_{1,J,k}$ are all exchangeable because the relationships between

- (i) two groups with the same j but different k
 - (ii) two groups with the same k but different j and
 - (iii) two groups between which both j and k differ
- are not the same.

In many cases, the expert's uncertainty about the values of $\theta_{1,1}, \dots, \theta_{J,k}$ may be usefully described in the form of an overall level and main effects for factors 1 and 2. We might write

$$\theta_{j,k} = \mu + \alpha_j + \beta_k \quad (4.3)$$

For example,

$$\begin{aligned} \theta_{1,k} &= (\mu + \beta_k) + \alpha_1 \\ \theta_{2,k} &= (\mu + \beta_k) + \alpha_2 \\ &\vdots \\ \theta_{J,k} &= (\mu + \beta_k) + \alpha_J \end{aligned}$$

In practice, we specify the mean vector and covariance matrix. We learn the values of some quantities and use this information to update our beliefs about others. It is difficult to specify the covariances directly or construct a covariance matrix for quantities that have the property of exchangeability. We make specifications relating the observables to uncertainty factors that have zero prior means but we need to specify their variances following Farrow (2003).

For programming in BUGS or JAGS specification language, it is convenient to be able to specify these because of the way we parameterize and construct the prior when we do MCMC. For computational reasons, it can be convenient to use Equation (4.3) to construct unknown quantities. Then, for identifiability, we need constraints.

The expectations of $(\mu + \beta_k), \alpha_1, \dots, \alpha_J$ are then not identifiable in terms of the J expectation $E(Y_{1,1,k}), \dots, E(Y_{1,J,k})$ unless we constrain $\alpha_1, \dots, \alpha_J$. That is, there are only $J - 1$ degrees of freedom for $\alpha_1, \dots, \alpha_J$. In order to maintain exchangeability, we treat those in the same way and impose the constraint $\sum_{j=1}^J \alpha_j = 0$. Imposing the so-called ‘‘corner constraint’’, $\alpha_1 = 0$ would violate exchangeability.

4.4.3.3 Construction of prior for the coefficient of a 2-level factor

Suppose that we have a 2-level categorical factor. If the model also contains an intercept then we would have only 1 degree of freedom and we could have only a single contrast with coefficients -1 and 1 . We might think of constructing the priors for the coefficient of a 2-level factor by representing the covariate values as $-1, 1$. Then, it is just the same as the construction of the prior for a quantitative covariate. Again, the center of the covariate values as $-1, 1$ is 0 which is exactly the case in Section 4.4.3.1.

4.4.3.4 Construction of prior distribution for the categorical (unordered) factors

Suppose that we have a categorical factor with m unordered categories. If the model also contains an intercept then we would have $m - 1$ degrees of freedom. We might want to

make the parameters identifiable by either constraining them to sum to zero or by fixing one of the parameters to be zero.

If we have a corner constraint where we fix one of the parameters to be zero, this suggests that we have a greater prior knowledge about this parameter. In some cases, this parameter could be the parameter for the baseline category and we would find the distribution of the remaining parameters. Suppose that we want to construct prior for four categorical factor. We might choose level 1 as the baseline (known) and consider the other levels of the categorical factor as in the scheme in Table 4.1.

Table 4.1: Scheme with level 1 as baseline

Level	covariate		
1	0	0	0
2	1	0	0
3	0	1	0
4	0	0	1

This gives a different prior variance to level 1 compared to the other three levels. In reality, we might not have greater prior knowledge about level 1. There are some possible ways to overcome this problem. We could constrain the parameters to sum to zero and make the parameters exchangeable. That is, the parameters could all have the same means, variances and each pair of parameters would also have the same covariance.

We illustrate constraining the parameters to sum to zero by supposing that we have a factor with four levels. The following scheme of orthogonal contrasts given in Table 4.2 could be used to constrain parameters to sum to zero.

Table 4.2: Scheme of orthogonal contrasts constrain to sum to zero

Level	contrast		
1	1	1	1
2	1	-1	-1
3	-1	1	-1
4	-1	-1	1

When the number of levels is not a power of 2 it is more complicated to construct a scheme of contrasts. The following scheme given in Table 4.3 could be used for a general case of m levels.

Table 4.4 shows a suitable scheme for $m = 5$.

As an illustration, we might want to construct prior for categorical factor where we want all levels of the factor to be exchangeable using the scheme in Table 4.3. We can

Table 4.3: A scheme when the number of levels is not a power of 2

Level	contrast				
1	-1	-1	-1	-1
2	1	-1	-1	-1
3	0	2	-1	-1
4	0	0	3	-1
⋮	⋮	⋮	⋮	
m	0	0	0	m - 1

Table 4.4: Example of a scheme number of levels is 5

Level	contrast			
1	-1	-1	-1	-1
2	1	-1	-1	-1
3	0	2	-1	-1
4	0	0	3	-1
5	0	0	0	4

transform $m - 1$ uncorrelated random variables into m random variables which are constrained to sum to zero. These random variables would be the parameters for a categorical covariate with m levels. For example, suppose we have $m - 1$ independent zero - mean random variables $\delta_1, \delta_2, \dots, \delta_{m-1}$ with $\underline{\delta} = (\delta_1, \delta_2, \dots, \delta_{m-1})^T$ and $\underline{\beta} = (\beta_1, \dots, \beta_m)^T$. We let $\text{Var}(\delta_j) = w_j$, $\underline{\beta} = M\underline{\delta}$ where M is a $m \times (m - 1)$ matrix given by the scheme in Table 4.3. We note that each column is a contrast so the total is fixed at zero.

We let $(m - 1)^2 w_{m-1} = v$ and so $\text{Var}(\beta_m) = v$. Hence,

$$w_{m-1} = \frac{v}{(m - 1)^2} \tag{4.4}$$

For $1 \leq i \leq m$, we require

$$\text{Var}(\beta_i) = v = (i - 1)^2 w_{i-1} + \sum_{k=i}^{m-1} w_k$$

$$\text{Var}(\beta_{i+1}) = v = i^2 w_i + \sum_{k=i+1}^{m-1} w_k$$

Hence

$$(i - 1)^2 w_{i-1} - i^2 w_i + \sum_{k=i}^{m-1} w_k - \sum_{k=i+1}^{m-1} w_k = 0$$

$$(i-1)^2 w_{i-1} - i^2 w_i + w_i = 0$$

and

$$w_{i-1} = \frac{(i^2 - 1)w_i}{(i-1)^2} = \frac{(i+1)}{(i-1)} w_i \quad (4.5)$$

We need to show that the covariances are also equal to each other. We would also show what the values are.

$$\text{Covar}(\beta_m, \beta_{m-1}) = -(m-1)w_{m-1} = -(m-1) \frac{v}{(m-1)^2} = \frac{-v}{m-1}$$

For $i = 2, \dots, m-1$

$$\text{Covar}(\beta_i, \beta_{i-1}) = -(i-1)w_{i-1} + \sum_{k=i}^{m-1} w_k$$

$$\text{Covar}(\beta_{i+1}, \beta_i) = -i w_i + \sum_{k=i+1}^{m-1} w_k$$

Hence

$$\begin{aligned} \text{Covar}(\beta_i, \beta_{i-1}) - \text{Covar}(\beta_{i+1}, \beta_i) &= i w_i - (i-1)w_{i-1} + w_i \\ &= (i+1)w_i - (i-1)w_{i-1} \\ &= (i+1)w_i - (i-1) \frac{(i+1)}{(i-1)} w_i \\ &= 0 \end{aligned}$$

and

$$\text{Covar}(\beta_1, \beta_2) = \text{Covar}(\beta_2, \beta_3) = \text{Covar}(\beta_{m-1}, \beta_m) = -\frac{v}{m-1}$$

Clearly, for $i = 3, \dots, m$ and $j = 1, \dots, i-2$

$$\text{Covar}(\beta_i, \beta_j) = \text{Covar}(\beta_i, \beta_{i-1}).$$

Hence for all $i \neq j$,

$$\text{Covar}(\beta_i, \beta_j) = -\frac{v}{m-1} \quad (4.6)$$

We could also want to relax the assumption of exchangeability by allowing the prior means of the parameters to be different by applying the exchangeability to the differences between the parameters and their prior means. We can construct a set of exchangeable parameters β 's by having $\beta_j - E(\beta_j)$ for $j = 1, \dots, m$. We can also allow the variances to be different while maintaining $m-1$ degrees of freedom. Let $\beta_1^*, \dots, \beta_m^*$ be a set of

exchangeable quantities with a zero-sum constraint. Then, we can let $\beta_j = m_j + s_j\beta_j^*$ for some choice of m_j, s_j . This will allow the prior means to be different. We will note that more general structures for the construction of priors for categorical variables are discussed in Farrow (2003).

4.4.3.5 Building a set of ordered random variables with a zero-sum constraint

We might also want to build a set of ordered random variables with a zero-sum constraint. Suppose that we wish to construct a multivariate normal prior distribution for the means, β_1, \dots, β_m , a set of m ordered categories. We require the sum of the means to be constrained to be zero. We will not constrain the means to be ordered but we may wish the means for neighbouring categories to be more positively correlated than those of categories further apart. We might also wish to impose stationarity.

Let the correlation between β_i and β_j be $r_{i,j}$. We assume that we wish $r_{i,j}$ to depend only on $|i - j|$. Let $\underline{\beta} = (\beta_1, \dots, \beta_m)^T$. Let $\underline{1}$ be a m -vector, $\underline{1} = (1, 1, \dots, 1)^T$.

We will consider the form of the correlation matrix C . Of course, given a set of variables with the required correlation matrix, it is easy to modify them so that they have any required means and variances. For simplicity in what follows, we will assume that β_1, \dots, β_m all have variance 1 so that C is the covariance matrix.

Since $r_{i,j}$ depends only on $|i - j|$, we write $r_{i,j} = c_k$ where $k = |i - j|$.

The zero-sum constraint implies that the variance of $\underline{1}^T \underline{\beta} = \sum_{i=1}^m \beta_i$ is zero. Thus, $\underline{1}^T C \underline{1} = 0$. That is the sum of the elements of C is zero. Hence,

$$m + 2 \sum_{j=1}^{m-1} (m - j)c_j = 0. \tag{4.7}$$

The zero-sum constraint will also imply that the correlation matrix is singular so

$$|C| = 0. \tag{4.8}$$

The requirements that correlations are greater between neighbours and that $-1 \leq r_{i,j} \leq 1$ lead to

$$-1 \leq c_{m-1} \leq c_{m-2} \leq \dots \leq c_2 \leq c_1 \leq 1. \tag{4.9}$$

We will later suggest how to choose c_1, \dots, c_{m-1} in practice. The exact choice will depend on how quickly we think that the correlation should decrease as $|i - j|$ increases. We will avoid violating any of the condition 4.7, 4.8 and 4.9. As an example of what happens if we violate (4.8), consider

$$C^* = \begin{pmatrix} 1 & -\frac{1}{4} & -1 \\ -\frac{1}{4} & 1 & -\frac{1}{4} \\ -1 & -\frac{1}{4} & 1 \end{pmatrix}$$

which violates (4.8). In fact $|C^*| = -1/4 < 0$ so this is not a valid correlation matrix. For example, consider $\text{Var}(\beta_1 + k\beta_2 + \beta_3)$ for some constant k . We can easily calculate $\text{Var}(\beta_1 + k\beta_2 + \beta_3) = (1, k, 1)C^*(1, k, 1)^T = k^2 - k$ but $k^2 - k < 0$ for $0 < k < 1$.

If there is no ordering, or we consider the order to be unimportant, we can set

$$c_1 = \dots = c_{m-1} = -(m-1)^{-1}.$$

In this case (4.8) is always satisfied since each row of C sums to zero. As an example, consider the case $m = 3$. Here, we have

$$C = \begin{pmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & 1 & -\frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2} & 1 \end{pmatrix}.$$

Consider $\text{Var}(k_1\beta_1 + k_2\beta_2 + \beta_3)$. (That is $k_1\beta_1 + k_2\beta_2 + k_3\beta_3$ where we set $k_3 = 1$ since we could rescale the total). It is easy to show that $\text{Var}(k_1\beta_1 + k_2\beta_2 + \beta_3) = (k_1 - k_2)^2/2$ and $(k_1 - k_2)^2 \geq 0$ for all k_1 and k_2 .

More generally, we can write

$$c_j = \delta_j - (m-1)^{-1}.$$

Then constraints (4.7) and (4.9) become

$$\sum_{j=1}^{m-1} (m-j)\delta_j = 0$$

and

$$\delta_1 \geq \delta_2 \geq \dots \geq \delta_{m-1},$$

$$\delta_1 - (m-1)^{-1} \leq 1,$$

$$\delta_{m-1} - (m-1)^{-1} \geq -1.$$

Clearly, for $m > 2$, we will have $\delta_1 \geq 0$ and $\delta_{m-1} \leq 0$. If $\delta_j = 0$ for $j = 1, \dots, m-1$, then we have the unordered case.

We will have numerical example of valid solutions for some values of m .

$m = 2$

The only valid solution is

$$C = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$$

$m = 3$

Condition (4.8) reduces the degrees of freedom to zero so it appears that the *only* value which provides a valid correlation matrix with the zero sum constraint is $c_1 = c_2 = -1/2$. That is $\delta_1 = \delta_2 = 0$. Hence, the only valid solution is

$$C = \begin{pmatrix} 1 & -0.5 & -0.5 \\ -0.5 & 1 & -0.5 \\ -0.5 & -0.5 & 1 \end{pmatrix}.$$

We can use Cholesky decomposition to sample this as follows, where ε_1 and ε_2 are independent random variables with $\varepsilon_i \sim N(0, 1)$.

$$\begin{aligned} \beta_1 &= \varepsilon_1 \\ \beta_2 &= -1/2\varepsilon_1 + \sqrt{3/4} \varepsilon_2 \\ \beta_3 &= -\beta_1 - \beta_2 \end{aligned}$$

4.4.3.6 Construction of prior for parameters representing a set of proportion

The Dirichlet distribution could be a choice of expressing our beliefs about a set of proportions because of the convenience of being a conjugate family to the multinomial likelihood in Bayesian analysis (see Elfadaly & Garthwaite (2013)). Suppose we wish to elicit expert beliefs about a set of uncertain proportion $\vartheta = (\vartheta_1, \dots, \vartheta_k)$ of k categories where $\vartheta_i \geq 0$ for $i = 1, 2, \dots, k$ and $\sum_{i=1}^k \vartheta_i = 1$. The Dirichlet distribution has a probability density function discussed in Appendix A.2.3 and parameters a_1, \dots, a_k where $\sum_{i=1}^k a_i = n$ and $\vartheta_i = a_i/n$. We can elicit the ϑ_i 's as the probability that we think that a single observation is in category i . Zapata-Vazquez *et al.* (2014) suggested thinking in terms of a judgement of the expected value of each ϑ_i together with judgement concerned with uncertainty to identify n . Some recent work on elicitation of parameters of Dirichlet distribution are given by Elfadaly & Garthwaite (2013).

4.5 Prior and posterior predictive distribution

Prediction is a major and important aspect of Bayesian inference. Suppose that before taking our sample, the uncertainty in the parameters θ is given by a prior distribution $\pi(\theta)$.

Then, the prior predictive distribution $\pi(x_{new})$ for some new data value x_{new} averaging over $\pi(\theta)$ is given as follows

$$\pi(x_{new}) = \int_{\Theta} \pi(x_{new}, \theta) d\theta = \int_{\Theta} \pi(x_{new}|\theta)\pi(\theta) d\theta$$

where Θ is the space of the parameter θ .

Suppose that after taking our sample, the posterior distribution is $\pi(\theta|D)$. The posterior predictive distribution $\pi(x_{new}|D)$ for some new data value x_{new} averaging over $\pi(\theta|D)$ is given as follows

$$\pi(x_{new}|D) = \int_{\Theta} \pi(x_{new}|\theta, D)\pi(\theta|D) d\theta = \int_{\Theta} \pi(x_{new}|\theta)\pi(\theta|D) d\theta.$$

4.6 Markov chain Monte Carlo (MCMC) techniques

4.6.1 Introduction

Often, it is not feasible to draw independent samples from the posterior distribution since the posterior might not be in a standard form. Hence, sampling is done through a Markov chain which has the required distribution as its stationary distribution by using Markov chain Monte Carlo (MCMC) simulation (Gilks *et al.*, 1996). Markov chain Monte Carlo is a generalised and flexible way of simulating a sample from the joint posterior distribution of the unknown parameters. Each sample may depend on the previous one and the sequence of samples follows a Markov chain and so, the past states provide no information about the future state if the present is known.

The Bayesian Inference Using Gibbs Sampler (BUGS) (Spiegelhalter *et al.*, 1995), WinBUGs (Lunn *et al.*, 2000), OpenBUGS (Spiegelhalter *et al.*, 2003) and Just Another Gibbs Sampler (JAGS) (Plummer, 2013) software packages can be used to implement MCMC algorithms without having to write new programs each time. These softwares can be used to analyse highly complex problems. With MCMC, we are able to generate the whole distribution numerically from which we can make any inferences of interest. In this section, MCMC algorithms like Metropolis-Hastings (MH) algorithms, Gibbs sampling algorithm and Metropolis within Gibbs will be discussed.

4.6.2 Metropolis and Metropolis-Hastings algorithm

The Metropolis algorithm was developed by Metropolis *et al.* (1953). The Metropolis algorithm is one of the MCMC algorithms used to simulate a Markov chain with the posterior $\pi(\theta|D)$ as its stationary distribution. Suppose we have some (arbitrary) conditional distribution for a value θ_{new}^t of θ , given another value θ^t . This is called a proposal distribution. Suppose that the density is $q(\theta_{new}^t|\theta^t)$ and that this is symmetric so,

$q(\theta_{new}^t|\theta^t) = q(\theta^t|\theta_{new}^t)$. The steps of the Metropolis algorithm (Metropolis *et al.*, 1953) would be given as follows:

Step 1: We set $t=1$ and choose an initial value θ^0 .

Step 2: We generate a new sample θ_{new}^t from the proposal distribution with density $q(\theta_{new}^t|\theta^t)$.

Step 3: We calculate the acceptance probability A which is

$$A = \min \left\{ \frac{\pi(\theta_{new}^t|D)}{\pi(\theta^t|D)}, 1 \right\}$$

Step 4: We sample u from the uniform distribution $U(0,1)$.

Step 5: We set $\theta^{t+1} = \theta_{new}^t$ if $u < A$ otherwise we set $\theta^{t+1} = \theta^t$.

Step 6: We set $t = t + 1$ if t is less than the number of desired samples and return to step 2. Otherwise we stop.

In this case, it is not necessary to know the normalising constant in Equation 4.1.

Hastings (1970) gave a more general form of the Metropolis algorithm and it was called the Metropolis-Hastings (MH) algorithm. The MH algorithm can be used with asymmetric proposal distribution where $q(\theta_{new}^t|\theta^t) \neq q(\theta^t|\theta_{new}^t)$. The Metropolis-Hastings (MH) algorithm takes all the steps of the Metropolis algorithm except that the acceptance probability A would be given by

$$A = \min \left\{ \frac{\pi(\theta_{new}^t|D)q(\theta^t|\theta_{new}^t)}{\pi(\theta^t|D)q(\theta_{new}^t|\theta^t)}, 1 \right\}$$

At each stage, a new value is generated from the proposal distribution. Again, the chain moves if this value is accepted (that is if $u < A$). Otherwise it stays where it is if the value is rejected.

4.6.3 Choice of proposal distribution

The MH algorithm allows using any distribution with suitable support as the proposal since the algorithm will still converge to the target distribution. However, the performance of the MH algorithm depends on the choice of the relationship between the density of interest and the proposal distribution. Choosing a suitable proposal distribution is very important to make sampling from the non-standard posterior distribution easy when using the MH algorithm. The choice of variance will affect the overall proportion of accepted moves. If the spread or standard deviation is too large, some points will have low acceptance probability and the chain will move less frequently because of many rejections. On the other hand, if the spread is too small, the chain will take a longer time to cross the support of the target density but more proposed values are accepted. Thus, in order to make our chain more efficient (better mixing), the proposal distribution should be carefully

chosen. Some commonly used classes of proposal distributions are the symmetric chain (as discussed above), the random walk and the independence chain.

The MH algorithm can be extended to higher-dimensions where the parameter $\theta = (\theta_1, \theta_2, \dots, \theta_d)^T$ could be a vector. Initial values should be selected for each element and a multivariate proposal distribution $q(\theta_{new}^t | \theta^t)$ should be used to select a d-dimensional new parameter. In particular, we might have to sample a vector from a multivariate normal distribution. We will discuss sampling from a multivariate normal distribution using the Cholesky decomposition in Section 4.6.6.

4.6.3.1 Random walk proposal

Suppose we define the proposed move at counter t as $\theta_{new}^t = \theta^t + z_t$ where z_1, z_2, \dots is a sequence of independent and identically distributed random variables and z_t has a distribution with density $g(\cdot)$ which is easily simulated. Since the proposed value is equal to the current value θ^t plus a random step, the algorithm is called random walk Metropolis-Hastings (See Gamerman (1997) for further information). The proposal density is given by $q(\theta_{new}^t | \theta^t) = g(\theta_{new}^t - \theta^t)$.

4.6.3.2 The independence proposal

In the independence sampler, the proposal distribution in the algorithm does not depend on the current value (Gamerman, 1997). Rather it is like the target distribution. The MH algorithm takes all the steps of Metropolis algorithm except that the acceptance probability A will be given as

$$A = \min \left\{ \frac{\pi(\theta_{new}^t | D) q(\theta^t | \theta_{new}^t)}{\pi(\theta^t | D) q(\theta_{new}^t | \theta^t)}, 1 \right\}$$

4.6.4 Gibbs Sampler

The Gibbs sampler was developed by Geman & Geman (1984). Gibbs sampling is one of the best known MCMC sampling algorithms. It allows sampling and updating parameter by parameter from the posterior distribution without knowing the normalising constant. It samples from a multivariate distribution by simulating from the conditional distributions. We can obtain a full conditional distribution for any parameter by ignoring all terms that are constant with respect to that parameter. When updating the j^{th} component, all other components are kept as constant and the proposal distribution for the new value will be the full conditional distribution. Some full conditional distribution are well known distributions which are easy to sample from such as normal or gamma distributions.

Suppose $\pi(\theta | D)$ is the posterior distribution given data D where $\theta = (\theta_1, \theta_2, \dots, \theta_d)^T$

is a d -dimensional vector of parameters and the full conditional densities are denoted by

$$\pi(\theta_j|\theta_1, \dots, \theta_{j-1}, \theta_{j+1}, \dots, \theta_d, D) \quad \text{for } j = 1, 2, \dots, d.$$

The Gibbs sampling algorithm follows the following steps:

Step 1: Set the iteration counter to $t=1$ and the chain at

$$\theta^0 = (\theta_1^0, \theta_2^0, \dots, \theta_d^0)^T$$

Step 2: Obtain a new value

$$\theta^{t+1} = (\theta_1^{t+1}, \theta_2^{t+1}, \dots, \theta_d^{t+1})^T.$$

from θ^t by successive simulation from the full conditional distributions as follows:

$$\begin{aligned} \theta_1^{t+1} &\sim \pi(\theta_1|\theta_2^t, \theta_3^t, \dots, \theta_d^t, D) \\ \theta_2^{t+1} &\sim \pi(\theta_2|\theta_1^{t+1}, \theta_3^t, \dots, \theta_d^t, D) \\ &\vdots \\ \theta_{d-1}^{t+1} &\sim \pi(\theta_{d-1}|\theta_1^{t+1}, \theta_2^{t+1}, \dots, \theta_{d-2}^t, \dots, \theta_d^t, D) \\ \theta_d^{t+1} &\sim \pi(\theta_d|\theta_1^{t+1}, \theta_2^{t+1}, \dots, \theta_{d-1}^{t+1}, D) \end{aligned}$$

Step 3: We move from counter t to $t+1$ and return to step 2.

Each of $\theta_1, \theta_2, \dots, \theta_d$ could be a vector. Hence, we can sample the unknowns in groups, known as blocks. Blocking of variables can greatly improve mixing and convergence of the MCMC scheme for carrying out inferences (Roberts & Sahu, 1997).

4.6.5 Metropolis within Gibbs

Metropolis within Gibbs or “component-wise MH” is used when the distribution of interest has a complicated form which prevents sampling directly from it. Some or all of its full conditional distribution have forms which make direct sampling difficult and so MH samples are drawn instead in these cases. The Metropolis within Gibbs algorithm goes through each unknown and samples directly from the corresponding full conditional distribution or carries a Metropolis-Hastings update as necessary (Gamerman, 1997) where we sample from a suitable proposal distribution and either accept or reject the proposal value according to a Metropolis-Hastings acceptance rule.

4.6.6 Sampling from a multivariate normal distribution using the Cholesky decomposition

Sampling using MCMC might involve sampling an “n” dimensional vector with multivariate normal distribution given by $N_n(\underline{\mu}, V)$ where $\underline{\mu}$ is the mean vector and V is the covariance matrix. We will discuss sampling from $N_n(\underline{\mu}, V)$ using the Cholesky decomposition.

Suppose we wish to sample a vector of parameters $\underline{\omega}$ from the multivariate normal distribution. We can subtract its mean $E(\underline{\omega})$ from it as follows:

$$\tilde{\omega} = \underline{\omega} - E(\underline{\omega})$$

We will decide on how to sample $\tilde{\omega}$. Suppose we have

$$\tilde{\omega} = L\underline{\varepsilon}$$

where we have $\underline{\varepsilon} = \varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ such that $\varepsilon_i \stackrel{iid}{\sim} N(0, 1)$. Then, the covariance matrix of the vector of ε 's, $\underline{\varepsilon}$ is given by

$$\text{Covar}(\underline{\varepsilon}) = I$$

where I is an identity matrix.

For convenience, L will be a lower triangular matrix. Then $U = L^T$ is an upper triangular matrix. The covariance matrix of $L\underline{\varepsilon}$ is given by

$$\text{Covar}(L\underline{\varepsilon}) = LIU = LU = V \quad (4.10)$$

which implies that

$$\tilde{\omega} = L\underline{\varepsilon}$$

where V is the covariance matrix of $\tilde{\omega}$.

We can sample values for $\tilde{\omega}$ and then add the mean $E(\underline{\omega})$ to the sampled values as

$$\underline{\omega} = \tilde{\omega} + E(\underline{\omega})$$

The Cholesky decomposition of V can be used to find the matrix L . There are standard R functions to calculate the Cholesky decomposition. The Cholesky decomposition is fast and numerically stable.

4.6.7 Initial values

We must first initialise the Markov chain by providing starting values for each unknown parameter in the statistical model before the MCMC method is started. In some cases,

like WinBUGs, OpenBUGS and JAGS, the initial values can be chosen automatically by sampling from the prior distribution of each parameter (Plummer, 2012). However, this may cause inappropriate starting values and may cause the program to crash if the prior distribution has very large variances. In BUGS and JAGS, the initial values can also be supplied as a separate file. We can also run several chains by supplying different starting values for each chain in order to monitor convergence.

4.6.8 Burn-in and convergence

There is need to be able to detect when the simulated values will be close to stationarity. Burn-in is the process of discarding the initial non-stationary portion of the Markov chain which minimises the effect of the initial values on the posterior inference (Lunn *et al.*, 2013); (Gamerman, 1997). At convergence, the chains will have mixed, so that the distribution of the simulations between and within chains will be identical. We can sometimes detect convergence in a straight forward way by eye. Visual inspection of plots is the most obvious and commonly used method of determining burn-in. This can be done by a history or trace plot of successive realisations of a particular variable against the iteration number of the sampler. However, this may not always be reliable. Therefore, using two or more parallel chains can also help to determine when the convergence has occurred. There are also various diagnostics available. A review of convergence diagnostics is given by Cowles & Carlin (1996). As well as the question of convergence to the stationary distribution, there is the question of how quickly the sampler moves around the parameter space. If the sampler moves quickly around the parameter space, and there is little dependence between successive samples, the sampler is said to mix well. Poor mixing often shows as strong autocorrelation of the samples.

4.6.9 Thinning

Thinning is done so that we do not have to store so many samples in the memory. When the samples are strongly positively auto correlated, the thinned samples contain almost as much information as the whole set but require less space to store. In thinning, we discard all but every n^{th} sampled value. For example, because of dependence, thinning 10000 samples with strong correlation by using only every 10^{th} sample will reduce the storage to 1000 samples while losing little information. It will not give a better representation of the joint posterior distribution. It simply saves space. Thinning does reduce the precision of the summary of the Markov chain (Link & Eaton, 2011) but, when there is strong positive autocorrelation, this reduction may be slight. We could just collect all samples and have more efficient summary than using the thinned chain if there is little or no dependence between the samples.

4.6.10 Efficiency and accuracy

In practice, we judge how long it takes for the sampler to converge. We need to be able to detect when the marginal behaviour of our MCMC is sufficiently close to stationarity. We run the simulation well beyond the point of convergence whereas detecting stationarity requires a sample size which depends very much on the efficiency of the Markov chain. The accuracy of our inference depends on the efficiency of the posterior sample. We will have to think of the choice of the sample size that will achieve the specific level of accuracy that is required. We refer to Spiegelhalter *et al.* (1995) for the batch means method of calculating Monte Carlo standard error.

Suppose we have H posterior samples for a parameter θ_i and we want to estimate the posterior expectation of some function $f(\theta_i)$. The Monte Carlo estimate using the samples is given by

$$\bar{f}_H = \frac{1}{H} \sum_{h=1}^H f(\theta_i^{(h)})$$

We again refer to Spiegelhalter *et al.* (1995) for the estimation of the Monte Carlo standard error (MCSE) where the samples $\theta_i^1, \dots, \theta_i^H$ are independent and Jones (2004) for dependent samples. We split the samples into B batches of length l . We assume that the central limit theorem holds for each batch and hence $\bar{f}_{b,l} = \frac{1}{l} \sum_{h=(b-1)l+1}^{ba} f(\theta_i^{(h)}) \sim_{approx} N(E[f(\theta_i)], \rho/H)$ for $b = 1, \dots, B$ and some positive ρ .

We can estimate ρ using

$$\hat{\rho} = \frac{l}{B-1} \sum_{b=1}^B (\bar{f}_{b,l} - \bar{f}_H)^2.$$

The MCSE of \bar{f}_H will tell us how accurately the Monte Carlo samples estimate the true posterior expectation. This MCSE of \bar{f}_H is approximately $\sqrt{\hat{\rho}/H}$ where the value of l is given by $\lceil \sqrt{(H)} \rceil$ (Abrams *et al.*, 1988). We will be interested in comparing the MCSE to the posterior standard deviation s which estimates the effective sample size (ESS). This will tell us whether the inaccuracy about estimating the posterior mean of $f(\theta)$ is large compared to the overall uncertainty about it. The ESS is given by

$$ESS = \left(\frac{s}{MCSE} \right)^2.$$

We would want to know the effect of sample size that we get per second. We can get this by dividing the average of the values of the effective sample size of all parameters by the time taken to collect H samples.

4.7 Numerical methods for improving mixing in MCMC

Mixing is the movement of the sampler around the posterior distribution. One at a time updating of parameters can be disadvantageous if the parameters are highly dependent in the posterior distribution and thus mixing can be poor. This is because the Markov chain may move around very slowly in the posterior distribution. Mixing of a Gibbs sampler can sometimes be improved by “blocking” groups of correlated unknowns. MCMC mixing and convergence can be improved by centering the variables because it reduces correlations between the parameters in the posterior. Thus, it is likely to make the coefficients of the parameters closer to being independent in the posterior which makes a Gibbs sampler work more efficiently. We introduce blocking of the parameters via the forward backward algorithm and Gaussian approximation of the proposal distribution of the parameters as numerical approaches to improving mixing.

4.7.1 Sampling the parameters using the idea of forward backward recursion

The forward backward algorithm developed by Baum *et al.* (1970) is a Markov chain Monte Carlo (MCMC) sampling technique used to simulate hidden Markov model parameters from the posterior distribution given observed data (Scott, 2002). This is done by calculating the conditional distributions of the hidden states given observed data and the model parameters. The forward backward recursion is usual when the data follow Markov processes (Scott, 1999) and is typically used with state space models (West & Harrison, 1997) which are assumed to follow a Gaussian process. The forward recursion accumulates information in the form of a conditional distribution about the process as it moves down the hidden Markov chain while the backward recursion updates the distribution of the parameter calculated in the forward recursion after information has been collected from all observed data.

Suppose we have a simple Gaussian Markov process $\theta_1, \dots, \theta_n$, where $\theta_j = \theta_{j-1} + W_j$ and $w_j \sim N(0, V_w)$, that we wish to sample given some observed data η_1, \dots, η_n using the forward backward algorithm. In the forward recursion of a simple Markov process, we change the distribution of θ_1 by conditioning on η_1 , $f(\theta_1|\eta_1)$. This will in turn change the distribution of θ_2 since $\theta_2 = \theta_1 + W_2$ for $W_2 \sim N(0, V_w)$. We then condition on η_2 and get the distribution of θ_2 , $f(\theta_2|\eta_1, \eta_2)$. We keep conditioning until we have conditioned on all the η 's and we get the distribution of θ_n , $f(\theta_n|\eta_1, \dots, \eta_n)$. We store all conditional distributions in the forward recursion. We then start the backward recursion by sampling θ_n from $f(\theta_n|\eta_1, \dots, \eta_n)$. We will go back to the distribution of θ_{n-1} , $f(\theta_{n-1}|\eta_1, \dots, \eta_{n-1})$. We condition on θ_n since we have now observed θ_n , $f(\theta_{n-1}|\eta_1, \dots, \eta_n, \theta_n)$. Then, we sample θ_{n-1} from $f(\theta_{n-1}|\eta_1, \dots, \eta_n, \theta_n)$. We will keep conditioning on the previous known values

of θ 's until we have the distribution of θ_1 given all other θ 's, $f(\theta_1|\eta_1, \dots, \eta_n, \theta_2, \dots, \theta_n)$ and we sample θ_1 from this distribution. We then have a sample from the joint distribution of $(\theta_1, \dots, \theta_n)$ given (η_1, \dots, η_n) .

4.7.2 Gaussian approximation of the proposal distribution of the parameters

Sampling a vector of dependent unknowns η as a block can improve mixing but when the joint full conditional distribution is not Gaussian and does not take some other convenient form, the options of a forward backward algorithm or sampling the joint distribution via a Cholesky decomposition are not available. Therefore, we use a Metropolis-Hastings step and we wish to choose an appropriate multivariate proposal distribution. One way to do this is to use a Gaussian approximation to the joint full conditional distribution. Suppose that the full conditional distribution of the vector of parameters $\underline{\eta}$ we want to sample from is $\pi(\underline{\eta}|D)$. A natural choice could be expanding around the mean of the approximation to generate an improved approximation until it is equal to the mode of the full conditional distribution (Rue & Held, 2005). This might require applying iterative quadratic approximation to the posterior density until convergence in order to obtain a normal proposal with mean equal to the posterior mode.

We wish to approximate the function

$$g = \log(\pi(\underline{\eta}|D)) .$$

We find the posterior mode using a Newton Raphson algorithm. Thus, we use an iterative method which is based on a quadratic Taylor series approximation of the logarithm of the posterior density g . We start by determining the vectors of the first derivatives $g'(\underline{\eta})$ and the matrix of second derivatives $g''(\underline{\eta})$ of the logarithm of the posterior density. We choose a starting or initial set of parameter values (probably the maximum likelihood estimate) say $\underline{\eta}_0$. For $p = 1, 2, \dots$, we compute $g'(\underline{\eta}_{p-1})$ and $g''(\underline{\eta}_{p-1})$. The Newton Raphson step at $p = 1$ is based on the quadratic approximation to g at $\underline{\eta}_{p-1}$. We set the new iterate or value $\underline{\eta}_p$ to maximise the quadratic approximation. We have

$$\underline{\eta}_p = \underline{\eta}_{p-1} - [g''(\underline{\eta}_{p-1})]^{-1} g'(\underline{\eta}_{p-1})$$

This procedure is iterated until the parameter values stabilize. We suppose that G''_{mode} is the value of the partial second derivative when evaluated at the mode. Then, the precision matrix of the proposal distribution is approximately

$$Q_1 = -G''_{\text{mode}}$$

Hence, we get the covariance matrix.

Once we have found the mode, we fit a multivariate normal distribution based on the mode. We use the approximated multivariate normal distribution as the proposal distribution using the Cholesky decomposition discussed in Section 4.6.6. The advantage of Newton Raphson algorithm is that the parameter values stabilizes very fast. Depending on the form of the full conditional distribution, it may be that the Newton Raphson algorithm is not guaranteed to converge to the model from all starting values so choice of initial values may be important.

4.8 Missing data

Data missingness is a frequent characteristic of many clinical data sets. Missing data may occur because of failure to report some evaluations, refusal to answer questionnaires, lost data et.c. It is a common technique to exclude or discard individuals with partially or totally missing data and use only the cases with complete data (Congdon, 2001). Little & Rubin (1987) discussed the disadvantages of case deletion. Discarding the observation where there are missing data is not advisable because this is inefficient and may lead to biased inferences. Longmore & Longmore (2007) in the development of the National Comprehensive Cancer Network International Prognostic Index (NCCN-IPI) (Section 3.8.6) used only cases with completely observed clinical information while 15% of the patients were deleted from the data analysis due to missing values in one or more clinical variables which is likely to cause biased inference in the analysis. The Scotland and Newcastle Lymphoma Group (SNLG) data set, which is one of the data sets used in this thesis, has 1391 patients and only 636 patients had complete data. Thus, almost half of cases have missing data which can cause the posterior summaries using the complete data set to be different from those using the full data set (including cases with missing data) and might give misleading results. There is therefore the need to develop methods which would deal with missing data (Zhao, 2010).

In Bayesian inference, the missing data become extra unknown quantities and the model can be extended to include the variables that may be missing. If a covariate is sometimes missing, we can use the cases where it is not missing to learn about its relationship with other covariates. Hence, we can obtain a joint posterior distribution for the model parameters by including the extra unknown quantities.

4.8.1 Types of Missingness

Suppose that $Y = (y_1, \dots, y_n)^T$ represents a response variable where some values of $x_{i,j}$ might be missing and therefore not observed. We introduce an indicator $I_i = (I_{i,1}, I_{i,2} \dots I_{i,J})^T$ which is a vector indicating which observations are missing for the i^{th} individual, where

$I_{i,j} = 1$ if $x_{i,j}$ is observed and $I_{i,j} = 0$ if $x_{i,j}$ is missing.

Suppose we use the parameters β to denote the model parameters and ϕ to denote the parameters of the missing data mechanism. The joint distribution of (T, X, I) given (β, ϕ) can be written as

$$f(T, X, I|\beta, \phi) = f(T, X|\beta)P_I(I|T, X, \phi)$$

where $P_I(I|T, X, \phi)$ is the conditional probability of the indicator I given the survival time T , the covariate matrix X and the missing data mechanism ϕ .

We partition the values of the covariate X into the observed X values (X_{obs}) and the missing X values (X_{miss}). Thus, $X = (X_{obs}, X_{miss})$. The distribution of the observed data is obtained by integrating over the distribution of X_{miss} :

$$f(T, X_{obs}, I|\beta, \phi) = \int f(T, X_{obs}, X_{miss}|\beta)P_I(I|T, X_{obs}, X_{miss}, \phi)dX_{miss}$$

Little & Rubin (1987) defined three different types of missingness. When the missingness on the covariate is conditionally independent of the missing values given observed values then the missing data are missing at random (MAR). That is

$$P_I(I|T, X_{obs}, X_{miss}, \phi) = P_I(I|T, X_{obs}, \phi).$$

The joint distribution under missing at random can be written as

$$f(T, X_{obs}, I|\beta, \phi) = f(T, X_{obs}|\beta)P_I(I|T, X_{obs}, \phi)$$

When the distribution of I is not related to either missing or observed values then the missingness is said to be completely at random (MCAR). In MCAR, we have that

$$P_I(I|T, X_{obs}, X_{miss}, \phi) = P_I(I|\phi)$$

The joint distribution under missing completely at random can be written as

$$f(T, X_{obs}, I|\beta, \phi) = f(T, X_{obs}|\beta)P_I(I|\phi)$$

The missing data mechanism is said to be ignorable when in addition to the missing data being missing at random, the model parameter β and the missing data parameter ϕ are independent in their prior. The joint prior density is given by

$$f_{\beta, \phi}(\beta, \phi) = f_{\beta}(\beta)f_{\phi}(\phi).$$

The joint posterior density is proportional to

$$f_{\beta}(\beta)f_{y,obs}(y_{obs}|\beta) \times f_{\phi}(\phi)p(I|y_{obs}, \phi)$$

In this thesis, the missing data are assumed missing at random (MAR). We can justify this assumption of missing at random because a lot of covariates were observed and the missingness indicator and the missing observations only have to be conditionally independent given the observed covariates. Hence, the missingness of the covariates may depend on the observed covariates X_{obs} and not on the missing values X_{miss} (Little & Rubin, 1987). The basic idea is that the more covariates we can condition on, the more likely it is that the missing values are conditionally independent of the missingness given the observed values. In the case of survival, there are a lot of covariates we are conditioning on which convey enough information to make the missingness conditionally independent of the missing observations.

4.9 Summary

In this chapter, we introduced Bayesian inference which is our intended approach of inference. We have also discussed various Markov chain Monte Carlo (MCMC) techniques. The elicitation and structures of prior distribution were discussed with illustrations. We have also discussed some numerical methods such as the forward backward algorithm and Gaussian approximation of the proposal distribution which will be suggested as approaches to improving poor mixing in later chapters. We also discussed missing data and different types of missingness.

Chapter 5

Introduction to Bayesian inference in survival analysis

5.1 Introduction

In this chapter we review the application of Bayesian inference to the analysis of survival data, based on previously published work. In addition we present new work on two topics. One is the construction of prior distributions for the parameters of survival models and the elicitation of hyperparameters for these prior distributions. The second is the construction of prognostic indices using the piecewise constant hazard model.

In Bayesian survival analysis, our beliefs about the unknown parameters of the survival model are expressed in terms of a probability distribution in the form of the prior which is updated to the posterior distribution after seeing the data in the form of the likelihood function. In cases where the posterior distribution does not have a standard form, it is usual to sample from the posterior distribution using Markov chain Monte Carlo (MCMC) methods.

We discuss in Section 3.2.1 the fact that censoring is very common in most survival data. We will discuss how censored observations are handled in survival analysis. Section 5.3 presents (with illustration) the construction of prior distributions of the parameters with application to survival analysis following the proportional hazard framework. We apply our discussion on the specification of prior distribution of the parameters in the model to survival analysis. We present constructing the prior distribution of the baseline parameters of a survival model by supposing an exponential lifetime distribution and assuming a normal prior distribution for the log-hazard. We will also present the construction of the prior distribution for the coefficient of a quantitative covariate and a 2-level factor with examples in survival models. We give some examples on the construction of the prior distribution for unordered factors with application to survival model. We present

the prior distributions for linear predictors by assuming a normal distribution and we plot the probability density function of the survival function with different means and standard deviation for the linear predictors.

We review a Bayesian approach to survival analysis using a Weibull lifetime distribution. We present a new approach to constructing the prior distribution for a Weibull shape parameter by thinking in terms of the median survival time, upper quartile survival time and the lower quartile survival time. We apply Bayesian survival modelling to the SNLG data set using our methods of construction of the prior distributions of all parameters of the model. We give posterior summaries to the parameters. We show with illustrations how the prognostic indices for a Weibull survival model is calculated using the SNLG data set.

We present Bayesian survival modelling using a piecewise constant hazard model. We also present with some illustrations a way of choosing the cutpoints of the time intervals for the piecewise constant hazard model. We review and construct the likelihood of a piecewise constant hazard model. We will construct the prior distribution such that the parameters are correlated over time by assuming a stationary distribution for the parameters using a first order auto-regressive process. We present a new approach to constructing the prior distribution for a frailty variance by supposing a normal distribution for the log-frailty and a gamma distribution for the precision of the log-frailty. We apply Bayesian survival modelling using a piecewise constant hazard model using the two example data sets as illustrations. We present time dependent effects by fitting two versions of the model. One without and the other with log normal frailties using the example data sets. We will also present ways of constructing the prognostic indices of a piecewise constant hazard model using both the survival probability at some time or the predictive median survival time, with illustrations.

5.2 Censoring and data augmentation

Some models have complicated likelihood functions which, if handled directly, would lead to difficult calculations. We may introduce extra variables known as auxiliary variables which are not observed but, if observed, would make the likelihood simpler. These auxiliary variables are then treated as if they were missing data. This is known as data augmentation. Tanner & Wong (1987) defined data augmentation as a concept of augmenting the observed data so as to make it easier to analyse. In Bayesian analysis for survival data, censored observations can be handled as “missing data” and the actual death times for censored patients are auxiliary variables. Alternatively, we evaluate the likelihood directly where a right censored observation contributes a term equal to the survival function as discussed in Section 3.4.2.

5.3 Proportional hazards and prior distribution for the coefficients of covariates

In this section, we follow the classical proportional hazard framework discussed in Section 3.3.1 in modelling the hazard rate of the survival time in terms of the covariates. We have an example of a baseline hazard using the exponential distribution which has constant hazard given by

$$h(t) = \lambda.$$

The survivor function is

$$S(t) = \exp\{-\lambda t\}$$

and the density is

$$f(t) = \lambda \exp\{-\lambda t\}.$$

The proportional hazard assumption means that every individual has a hazard function that is a constant multiple of the underlying baseline hazard. The constant of proportionality depends on the covariates. This is done following the linear approach or linear predictor given in Equation (3.2).

5.3.1 Construction of prior distribution for the coefficients of a linear predictor

In this section, we will construct the joint prior distribution for the coefficients in a linear predictor using a multivariate normal distribution. We will suppose for now that we have an exponential lifetime distribution so that there are no additional parameters. We will consider the case of a Weibull distribution, where there is, in addition, a shape parameter, in Section 5.5.

The construction of the prior distribution will often involve elicitation of prior beliefs from a person other than the analyst. This person will often have expert knowledge of the subject matter of the study. We will refer to this person as the “expert”. The expert may not have great expertise in probability and statistics so the design of suitable questions for elicitation is important. Translating the beliefs of the expert into a statistical form to suit the analysis might be challenging. The expert may not be able to provide probability distributions for the parameters of interest directly but rather the researcher is required to find appropriate questions to ask the expert in order to determine the probability distribution of the parameters. It is generally considered that the expert should be asked questions about observable quantities. Hiance *et al.* (2009) used a structured questionnaire to elicit the opinion of a group of experts.

The details of the elicitation method may vary depending on the type of covariates

involved. We will give examples of the construction of priors with covariates of different types.

Ren & Oakley (2013) considered elicitation of prior distributions for the parameters of lifetime distributions. In the case of a single lifetime distribution, they referred to Sections 6.3 and 6.4 of O’Hagan *et al.* (2006). Ren & Oakley (2013) then considered the case of two groups. They recommended eliciting judgements about the survival probability at some specified time and we follow this approach. However, in our application, we have more than two groups and need to consider beliefs about the effects of several covariates. Therefore, we adapt the approach by adopting the following strategy.

We first identify a suitable “baseline” case. This is a hypothetical patient with “typical” or “central” covariate values. We will then consider the effects of changing covariate values away from this baseline case by varying one covariate at a time from the baseline values. By choosing such a central case as a baseline, it becomes reasonable to suppose that the coefficients of covariates are independent of the baseline log-hazard in the prior distribution and may be elicited separately. It does not follow that the coefficients of covariates are mutually independent but this may be a reasonable assumption in many cases in practice. It would be possible to extend the method to allow dependence between coefficients by considering the effects of varying more than one covariate.

So, our first question is the following.

Q1: *“Please identify a set of covariate values for a “baseline” patient. This should be a “typical” case so that it is relatively easy to use your experience to express judgements about this case. The baseline case should be the case about which you have the least uncertainty. Making this case “central” or “average” enables us to ask about judgements for this case separately and then to ask other questions about the effects of changing covariate values from this case.”*

Because we will ask about the survival probability at a specified time, it is necessary to identify a suitable choice of time to specify. Our second question therefore asks for the median survival time for a baseline patient.

Q2 *“Please identify a survival time t such that you feel that it is equally likely that a baseline patient would die before or after this time.”*

5.3.1.1 Construction of prior distribution for the baseline parameter of a survival model

We start by eliciting a prior distribution for the log-hazard of a baseline patient. Let the reference time elicited in response to Question Q2, or some rounded value close to it, be

t_0 , although, in subsequent questions, the actual value would be given rather than the symbol “ t_0 ”.

Given a sample of baseline patients, the number who survive beyond t_0 has a binomial distribution. We wish to elicit beliefs about the parameter of this binomial distribution. Winkler (1967) discusses a number of methods for eliciting prior distribution for binomial parameters. These methods may be used but we might also use a more direct approach with questions such as the following.

- Q3** “Please think about the proportion of baseline patients who would survive beyond t_0 . Please give a value p such that you think that it is equally likely that the proportion is less than or greater than p .” Let the given value be p_2 .
- Q4** “Still thinking about the proportion of baseline patients who would survive beyond t_0 , please give a value p such that you think that it is equally likely that the proportion is between p_2 and p and that it is greater than p .” Let the given value be p_3 .
- Q5** “Still thinking about the proportion of baseline patients who would survive beyond t_0 , please give a value p such that you think that it is equally likely that the proportion is between p and p_2 and that it is less than p .” Let the given value be p_1 .

We thus elicit the three quartiles, p_1, p_2, p_3 , of the prior distribution for the survival probability at time t_0 . These may be transformed to give the quartiles of the prior distribution for the baseline log-hazard β_0 . Since the survival probability is $\exp(-\lambda t_0)$ where $\lambda = \exp(\beta_0)$, the three quartiles for β_0 become

$$b_{0,h} = \ln \left\{ -\frac{\ln p_{4-h}}{t_0} \right\} \quad (5.1)$$

for $h = 1, 2, 3$.

Using a normal prior distribution, we can determine the mean as $(b_{0,1} + b_{0,3})/2$ and the standard deviation as $(b_{0,3} - b_{0,1})/1.349$. The additional value, $b_{0,2}$ can be used as a check of normality, or, at least, symmetry, since we should have $b_{0,2} \approx (b_{0,1} + b_{0,3})/2$.

We will now consider the elicitation of prior distributions for other parameters in the log-hazard for non-baseline patients.

5.3.1.2 Example: Construction of prior distribution for the coefficient of a quantitative variable

To construct the prior distribution for the effect of a quantitative covariate, x_j , we introduce a second hypothetical patient. The covariate values for this patient are the same

as those for the baseline patient except in the case of x_j where the baseline value $x_{j,0}$ is replaced by $x_{j,0} + \delta_j$. A suitable value for δ_j should be determined in consultation with the expert.

Q6 *“We are going to consider the effect of a change in the value of (covariate x_j). Please suggest a value, different to the baseline value but where you feel able to give judgements based on your experience which may be used for this purpose. The value should not be too close to the baseline value but should be within the range of your experience.”*

Using the same method as in Section 5.3.1.1 we elicit a normal distribution for the log-hazard in this case, which is $\beta_0 + \beta_j\delta_j$.

Given the underlying linear model and assuming that the value of β_j is not known with certainty, the variance of this distribution should be greater than that of β_0 . If it is found that it is not then the elicitation process may be repeated to try to overcome this problem. Alternatively, if the variance given in this second case is less than that for the baseline, the elicitation might be repeated with this case as the baseline. If the two variances are equal then increasing the value of $|\delta_j|$ might solve the problem. Note that problems of this sort are a consequence of the linear parametric model and will not apply when we consider nonparametric models in Chapters 6 and 7.

Assuming that the value given for $\text{Var}(\beta_0 + \beta_j\delta_j)$ is greater than that for $\text{Var}(\beta_0)$ and that β_0 and β_1 may be considered to be independent, we have

$$\text{E}(\beta_j) = \frac{1}{\delta_j} \{ \text{E}(\beta_0 + \beta_j\delta_j) - \text{E}(\beta_0) \} \quad (5.2)$$

and

$$\text{Var}(\beta_j) = \frac{1}{\delta_j^2} \{ \text{Var}(\beta_0 + \beta_j\delta_j) - \text{Var}(\beta_0) \}. \quad (5.3)$$

5.3.1.3 Example: Construction of prior distribution for the coefficient of a 2-level factor

Suppose that we have a 2-level factor, that is a binary covariate. We might choose to use the so-called “corner constraint” and set the effect of the baseline level of the factor to zero. Then a prior distribution for the effect of the other level of the factor can be obtained in the same way as for a quantitative covariate, with $\delta_j = 1$. However, we might prefer to use the “zero-sum” constraint, particularly if we regard the effects of the two levels of the factor as exchangeable. In this case, the model for the log-hazards for the

two levels, ignoring other covariates, would take the form $\beta_0 - \beta_j$, $\beta_0 + \beta_j$. In this case, when eliciting prior distributions for the effects of other covariates, we can regard one of the levels as the baseline and, as long as there are no interaction effects and the effects are regarded as independent, this has no effect on the cases of the other covariates. The only difference is that the difference in mean between the two levels of the factor is now $2\beta_j$ so, in effect, $\delta_j = 2$.

5.3.1.4 Example: Construction of prior distribution for unordered factors

Now we consider the case of unordered factors with $K > 2$ levels. We might think of constructing the prior distribution for the effects of the levels by choosing a baseline or reference level. If we fix one of the parameters to be zero as discussed in Section 4.4.3.4 then this parameter is the baseline and we compare its level with the others (see Scheme in Table 4.1). However, again, we might prefer to use a zero-sum constraint. In this case, we adopt the following procedure.

If we have more than one factor then, because of the need to have a common baseline, we consider these factors together. Let us leave aside other covariates for now and suppose that we have J factors and that factor j has K_j levels. Then we represent the mean of the log-hazard, when the levels of the factors are $\underline{k} = (k(1), \dots, k(J))$, as

$$\beta^*(\underline{k}) = \beta_0 + \sum_{j=1}^J \beta_{j,k(j)} \quad (5.4)$$

where, for $j = 1, \dots, J$,

$$\sum_{k=1}^{K_j} \beta_{j,k} = 0.$$

We start by selecting a single reference profile of factor levels. This should preferably be a profile which would be likely in practice but, without loss of generality, we can label this profile $\underline{k}_0 = (1, 1, \dots, 1)$. Using the method described in Section 5.3.1.1, we can elicit a mean m_0 and variance v_0 for the log-hazard at this profile, that is for $\beta^*(\underline{k}_0)$.

Next we consider other profiles. For an additive model, we can hold the levels of all but one factor at their reference levels and vary just the level of the factor under consideration. That is, for $j = 1, \dots, J$ and $k = 2, \dots, K(j)$ we consider the profile $\underline{k}_{j,k} = (1, 1, \dots, 1, k, 1, \dots, 1)$ where the value k occurs for factor j . This leads to a mean

$m_{j,k}$ and a variance $v_{j,k}$ for the log-hazard. The prior mean for $\beta_{j,k}$ is then

$$E(\beta_{j,k}) = m_{j,k} - \sum_{k=1}^{K_j} m_{j,k} / K_j,$$

where, for $j = 1, \dots, J$, $m_{j,1} = m_0$. We can then find the prior mean for β_0 as

$$E(\beta_0) = m_0 - \sum_{j=1}^J E(\beta_{j,1}) = \sum_{j=1}^J \sum_{k=1}^{K_j} \frac{m_{j,k}}{K_j} - (J-1)m_0.$$

The method can be adapted if interaction effects are included.

The effects $\beta_{j,1}, \dots, \beta_{j,K_j}$ can now be given a correlation matrix which imposes the zero-sum constraint. For example, we recall the correlation matrix in Section 4.4.3.5 for a 3-level factor given by

$$C = \begin{pmatrix} 1 & -0.5 & -0.5 \\ -0.5 & 1 & -0.5 \\ -0.5 & -0.5 & 1 \end{pmatrix}.$$

However, to find the variances of $\beta_0, \beta_{1,1}, \dots, \beta_{J,K_j}$, we need to assess covariances between levels. We will often be willing to simplify matters by assuming that *a priori* \underline{B}_j is independent of \underline{B}_h and of β_0 , where $h \neq j$ and $\underline{B} = (\beta_{j,1}, \dots, \beta_{j,K_j})$, and the elements of \underline{B}_j are exchangeable for each j . Then, from (5.4), we see that

$$\text{Var}[\beta^*(\underline{k}_{j,k})] - \text{Covar}[\beta^*(\underline{k}_{j,k}), \beta^*(\underline{k}_0)] = \text{Var}(\beta_{j,k}) - \text{Covar}(\beta_{j,k}, \beta_{j,1}).$$

Using the exchangeability assumption, this simplifies to

$$\text{Var}[\beta^*(\underline{k}_{j,k})] - \text{Covar}[\beta^*(\underline{k}_{j,k}), \beta^*(\underline{k}_0)] = \frac{K_j}{K_j - 1} \text{Var}(\beta_{j,k}).$$

Hence, given assessments of the covariances, we can find $\text{Var}(\beta_{j,k})$ for all j, k and, in turn, using the independence assumption, this allows us to calculate

$$\text{Var}(\beta_0) = \text{Var}[\beta^*(\underline{k}_0)] - \sum_{j=1}^J \text{Var}[\beta_{j,1}].$$

Relaxing the independence and exchangeability assumptions would lead to the need to assess a greater number of covariances.

To find the required covariances we can adapt the hypothetical future sample method of Winkler (1967). For example, we might ask a question such as the following. The actual numbers used in the question would be chosen to be reasonably realistic.

Q7 “Still thinking of patients where all other covariates are at their baseline values, consider how your judgements might change if we observed some data. Suppose that we observed a sample of 30 patients with all factors at Level 1 and that 20 of these survived beyond time (t_0). What would your assessed mean now be for the proportion of patients, with all factors at Level 1 except for Factor (j) which is at Level 2, who would survive beyond time (t_0)?”

Similarly we elicit revised quartiles and hence a revised mean and variance for $\beta^*(\underline{k}_{j,2})$.

In order to compute the effect of the observed patients at Profile k_0 on the distribution, we approximate the prior distribution for π_0 , the survival probability at Profile k_0 and time t_0 , by a Beta($a_{0,0}, b_{0,0}$) distribution where $a_{0,0}$ and $b_{0,0}$ are chosen to match the specified lower and upper quartiles. The hypothetical observed data then lead to a Beta($a_{0,1}, b_{0,1}$) distribution as an approximation posterior distribution. With the example values given above, $a_{0,1} = a_{0,0} + 20$ and $b_{0,1} = b_{0,0} + 10$. This allows new quartiles to be computed for π_0 . Then we apply (5.1) to obtain new quartiles for β_0^* and hence a revised mean and variance for β_0^* .

If we observed the value of β_0^* , then the revised mean for $\beta^*(\underline{k}_{j,2})$ would be

$$E_1[\beta^*(\underline{k}_{j,2}) \mid \beta_0^*] = E_0[\beta^*(\underline{k}_{j,2})] + C_j \text{Var}_0[\beta_0^*]^{-1} \{\beta_0^* - E_0[\beta_0^*]\} \quad (5.5)$$

where $E_0[\beta_0^*]$ and $E_0[\beta^*(\underline{k}_{j,2})]$ are the prior means of β_0^* and $\beta^*(\underline{k}_{j,2})$ respectively, $\text{Var}_0[\beta_0^*]$ is the prior variance of β_0^* and C_j is the prior covariance.

Applying the idea of Bayes linear kinematics, as discussed by Goldstein & Shaw (2004), we suppose that the relationship (5.5) still holds when we do not observe β_0^* but observe data D which cause us to revise our mean and variance for β_0^* . Then we replace (5.5) with

$$E_1[\beta^*(\underline{k}_{j,2}) \mid D] = E_0[\beta^*(\underline{k}_{j,2})] + C_j \text{Var}_0[\beta_0^*]^{-1} \{E_1[\beta_0^* \mid D] - E_0[\beta_0^*]\}$$

where $E_1[\beta_0^* \mid D]$ is the revised mean for β_0^* after observing the data. Hence we assess C_j using

$$C_j = \text{Var}_0[\beta_0^*] \frac{E_1[\beta^*(\underline{k}_{j,2}) \mid D] - E_0[\beta^*(\underline{k}_{j,2})]}{E_1[\beta_0^* \mid D] - E_0[\beta_0^*]}.$$

5.4 Prior distributions for linear predictors

Prior judgements should be related to observable quantities (Zapata-Vazquez *et al.*, 2014). In assessing prior distributions for the parameters of a survival model, we should consider their effects on the prior predictive distributions of observable quantities such as survival proportions or, at least, on the prior distribution of summaries of distributions of observable quantities. To elicit the prior distribution of the logarithm of the hazard of a

patient, we might think of the proportion of a large group of patients that will survive for a particular, typical, number of years (e.g 3 years in the case of the SNLG data set). For example, suppose that we think that about one tenth to nine tenth of the patients will survive 3 years. Let us suppose that we have an exponential lifetime distribution with parameter λ . This corresponds to having no changes in parameters between time intervals in the piecewise constant hazard model or $\alpha = 1$ in the Weibull model. We will obtain

$$0.1 \leq \exp^{-3\lambda} \leq 0.9$$

$$\log 0.1 \leq -3\lambda \leq \log 0.9$$

$$-3.36 \leq \log \lambda \leq -0.265$$

This gives a range for the linear predictor (logarithm of the hazard). A normal prior distribution for logarithm of the hazard will give us a mean of $\mu = -1.81$ and standard deviation $\sigma = 0.77$. So,

$$\eta \sim N(\mu, \sigma^2)$$

where

$$\eta = \log \lambda$$

and λ is the hazard. We would like to find the probability of surviving say τ years. This will be done by the transformation of the random variables. The survival probability (p) is

$$p = \exp \{-\tau\lambda\}.$$

Given a value for p

$$\lambda = \frac{-\log p}{\tau}$$

is implied. Then $\eta = \log \lambda = \log \frac{-\log p}{\tau}$

This is called a “complementary log-log transformation”. We will let the probability density functions of p and η be $f(p)$ and $f(\eta)$ respectively. Then

$$f(p) = f(\eta)/|J|$$

where

$$|J| = \left| \frac{dp}{d\eta} \right| = \tau \exp \{\eta - \tau \exp^\eta\} = \lambda \tau \exp \{-\lambda \tau\}$$

Hence we have

$$f(p) = \frac{1}{\tau} (2\pi\sigma^2)^{1/2} \exp \left\{ -\frac{1}{2\sigma^2} (\eta - \mu)^2 - \eta + \tau e^\eta \right\} \quad (5.6)$$

where $\eta = \log \frac{-\log p}{3}$.

Figure 5.1 shows various plots of the probability density function of the survival function with different means and standard deviations for η .

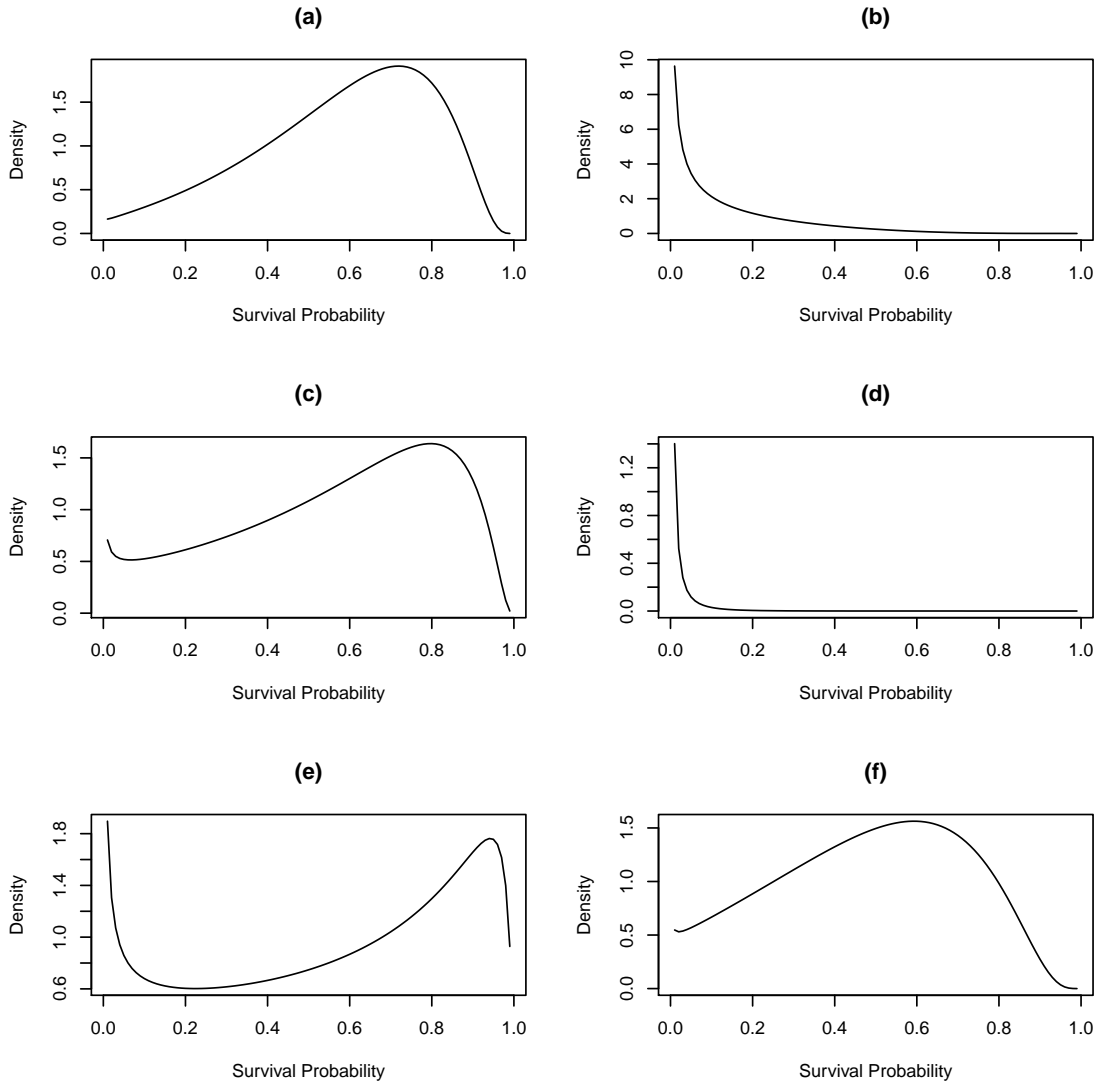


Figure 5.1: Density plots of the 3 years survival probability with (a) mean -1.81 and standard deviation 0.77 (b) mean 0 and standard deviation 0.77 (c) mean -1.81 and standard deviation 1 (d) mean 2 and standard deviation 0.77 (e) mean -1.81 and standard deviation 1.5 (f) mean -1.465 and standard deviation 0.77

The plot of the 3 years survival probability with mean = -1.81 and standard deviation = 0.77 does not look symmetric hence there will be the need to investigate further on how the plots will look when the elicited prior means standard deviations are either increased or decreased. When the means and standard deviations of the 3 years survival probability were increased or reduced the plots no longer looked unimodal.

We might want to think of the value of the linear predictor η when the survival probability $p = 0.5$. The value of the linear predictor at $p = 0.5$ given by

$$\exp\{-3 \exp^\eta\} = 0.5$$

$$\eta = \log \frac{\log 2}{3} = -1.465$$

We will try various values of standard deviation that makes the plot look symmetric. The standard deviation value of 0.3 makes it look symmetric. Since η has a normal distribution, we will work out the corresponding 95% interval values of the survival function as between 0.282 and 0.683. Figure 5.1 also shows various plot of the probability density function of the survival function with mean = -1.465 and standard deviation = 0.3 which looks close to being symmetric. We then have to check for the conditions for symmetry and unimodality in the function.

We attempt to check for the conditions of symmetry by replacing η with a negative η in the function and we try to make the equation back to its original form. We will work in terms of the logarithm of the density in Equation (5.6) since it is easier. The log of the density is

$$\log(f(p)) = -\frac{1}{2\sigma^2}(\eta - \mu)^2 - \eta + \tau e^\eta$$

We will not end up with the exact function that we started with at the beginning. This is not symmetric.

We might think of the conditions for unimodality. The partial first derivative of the log of the pdf with respect to η is

$$\frac{d[\log(f(p))]}{d(\eta)} = -\sigma^2(\eta - \mu) - 1 + \tau \exp\{\eta\} = 0$$

The second partial derivative is also given as

$$\frac{d^2(\log(f(p)))}{d^2(\eta)} = -\sigma^2 + \tau \exp\{\eta\}$$

This does not lead to tractable conditions for unimodality.

We might think that it will be more tractable in the case of the median survival time. The median survival time q_2 is the time at which the the survival probability is half. Thus

$$\frac{1}{2} = \exp\{-q_2\lambda\}$$

$$q_2 = \frac{\log(2)}{\lambda}$$

We have that

$$\frac{1}{\lambda} = \exp\{-\eta\}$$

and since $\eta \sim N(\mu, \sigma^2)$, q_2 has a log-normal distribution. We work out the mean and variance of this log-normal distribution. The mean of the log-normal is $\exp\left\{\mu + \frac{\sigma^2}{2}\right\}$ and the variance is $(\exp\{2\mu + \sigma^2\}) (\exp\{\sigma^2\} - 1)$. This turns out to be tractable. Some plots of the log-normal density at different means and variances of η are given in Figure 5.2. The log-normal distribution is not symmetric but it is always unimodal.

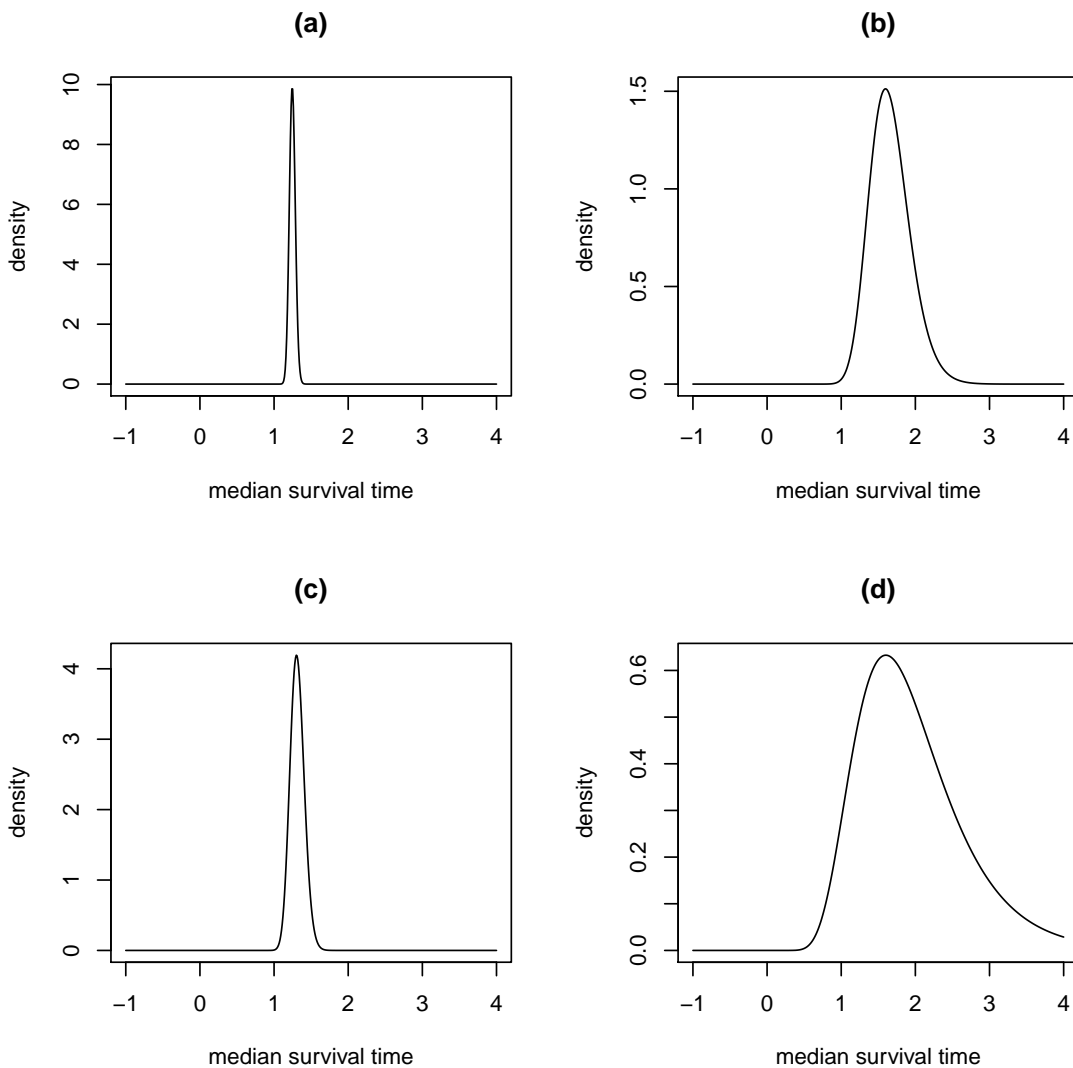


Figure 5.2: Density plots of the median survival time using different means and variances for η (a) mean -1.81 and standard deviation 0.77 (b) mean -1 and standard deviation 0.77 (c) mean -1.81 and standard deviation 1 (d) mean -1 and standard deviation 1

5.5 Bayesian survival analysis with Weibull lifetimes

5.5.1 Model, prior and likelihood

We refer to the Weibull distribution discussed in Section 3.4.1. It will be used because the shape parameter accounts for additional possible hazard shapes and the scale parameter will be used to incorporate the covariates of the data. The logarithm of the likelihood is given by

$$\ell = \log \{L(\underline{\beta}, \alpha|D)\} = \sum_{i \in E} \{\log \{\lambda_i\} + (\alpha - 1) \log \{t_i\}\} + n_D \log \{\alpha\} - \sum_{i=1}^n \lambda_i t_i^\alpha$$

The vector of regression coefficients is

$$\underline{\beta} = (\beta_0, \beta_1, \dots, \beta_S)^T$$

where β_0 is the intercept and β_s is the regression coefficient for the s^{th} covariate. We recall that $\lambda_i = \exp\{\eta_i\}$ and $\eta_i = \beta_0 + \sum_{s=1}^S \beta_s x_{i,s}$ (see Equation (3.2)). The prior density for the vector of regression coefficients $\underline{\beta} = (\beta_0, \beta_1, \dots, \beta_S)^T$ could be a multivariate normal distribution $N_{S+1}(\underline{\mu}, V)$ where $\underline{\mu}$ is the vector of prior means given as $\underline{\mu} = (\mu_0, \dots, \mu_S)^T$ and V is a $(S+1)$ by $(S+1)$ covariance matrix. The prior density for the vector of regression coefficient is then given by

$$(2\pi)^{-S/2} |V|^{-1/2} \exp \left\{ -\frac{1}{2} [(\underline{\beta} - \underline{\mu})^T V^{-1} (\underline{\beta} - \underline{\mu})] \right\}$$

For illustration, we may choose to give α and $\underline{\beta}$ independent prior on the grounds that beliefs about $\underline{\beta}$ are beliefs about the effects of covariates on the hazard function. In general, α and $\underline{\beta}$ need not be independent. A special case where the regression coefficients are independent simplifies the logarithm of the multivariate normal prior density to

$$-\frac{S}{2} \log\{2\pi\} - \frac{1}{2} \sum_{s=0}^S \log |V_s| - \frac{1}{2} \sum_{s=0}^S \frac{(\beta_s - \mu_s)^2}{V_s}$$

given that $V = \text{diag}(V_0, V_1, \dots, V_S)$

where V_0 is the variance of the intercept regression coefficient and V_s is the variance of the coefficient of the s^{th} covariate since the coefficients are assumed independent.

The prior distribution for α could be a gamma distribution $\alpha \sim \text{Ga}(a, b)$ with density

given by

$$\begin{aligned}\pi(\alpha|a, b) &= \frac{b^a}{\Gamma(a)} \alpha^{a-1} \exp\{-b\alpha\} \\ &\propto \alpha^{a-1} \exp\{-b\alpha\}\end{aligned}$$

Suppose that we denote the joint prior density of the parameters by $\pi(\underline{\beta}, \alpha)$. The posterior density $\pi(\underline{\beta}, \alpha|D)$ is then given by

$$\begin{aligned}\pi(\underline{\beta}, \alpha|D) &\propto \text{prior} \times \text{likelihood} \\ &= \kappa \pi(\underline{\beta}, \alpha) L(\underline{\beta}, \alpha|D)\end{aligned}$$

where κ is the constant of proportionality. The logarithm of the posterior density is

$$\begin{aligned}\log\{\pi(\underline{\beta}, \alpha|D)\} &= \log\{\kappa\} + \log\{\pi(\underline{\beta}, \alpha)\} + \log\{L(\underline{\beta}, \alpha|D)\} \\ &= \log \kappa + (a-1) \log \alpha - b\alpha - \frac{S}{2} \log\{2\pi\} - \frac{1}{2} \sum_{s=0}^S \log\{V_s\} - \frac{1}{2} \sum_{s=0}^S \frac{(\beta_s - \mu_s)^2}{V_s} \\ &\quad + \sum_{i \in E} \{\log(\lambda_i) + (\alpha-1) \log(t_i)\} + n_D \log(\alpha) - \sum_{i=1}^n \lambda_i t_i^\alpha\end{aligned}$$

The joint posterior density would not have a closed form but we can simulate from it using MCMC techniques. Metropolis within Gibbs sampling (discussed in Section 4.6.5) can be used.

5.5.1.1 Construction of prior distribution for Weibull shape parameter

One way to think of the prior of the shape parameter (α) of a Weibull distribution is by thinking in terms of the median survival time t_m , upper quartile survival time $t_{q(3)}$ and lower quartile survival time $t_{q(1)}$. We have that the median survival time t_m has a probability $\Pr(T < t_m) = 0.5$. The survival function at the median survival time is given by

$$S(t_m) = 1 - F(t_m) = 1/2$$

so,

$$\exp\{-\lambda t_m^\alpha\} = 1/2$$

and the median survival time t_m can be expressed as

$$t_m = \left(\frac{\log 2}{\lambda}\right)^{1/\alpha}$$

Let the lower and upper quartiles of the survival time be $t_{q(1)}$ and $t_{q(3)}$. We have

$$\begin{aligned} S(t_{q(3)}) &= 1 - F(t_{q(3)}) = 1/4 \\ \exp\{-\lambda t_{q(3)}^\alpha\} &= 1/4 \end{aligned}$$

The upper quartile survival time can then be expressed as

$$t_{q(3)} = \left(\frac{2 \log 2}{\lambda}\right)^{1/\alpha}$$

We have

$$\begin{aligned} S(t_{q(1)}) &= 1 - F(t_{q(1)}) = 3/4 \\ \exp\{-\lambda t_{q(1)}^\alpha\} &= 3/4 \\ -\lambda t_{q(1)}^\alpha &= \log(3/4) \end{aligned}$$

The lower quartile survival time can then be expressed as

$$t_{q(1)} = \left(\frac{-\log 3 + 2 \log 2}{\lambda}\right)^{1/\alpha}$$

We may decide to take the ratio between the upper and lower quartile survival times which gives:

$$t_{q(3)}/t_{q(1)} = \left(\frac{2 \log 2}{-\log 3 + 2 \log 2}\right)^{1/\alpha} = k^{1/\alpha} \quad (5.7)$$

We have eliminated λ from the ratio and found a quantity, given in Equation (5.7), which only depends on α . We have that

$$t_{q(3)}/t_{q(1)} = k^{1/\alpha}$$

If we take the logarithms of both sides we get

$$\alpha = \frac{\log(k)}{\log(t_{q(3)}/t_{q(1)})} \quad (5.8)$$

We can elicit values for $t_{q(1)}$ and $t_{q(3)}$ for a baseline patient by adapting Question **Q2**.

Q8 *“In response to Question **Q2** you identified t_0 as the time such that you felt that it was equally likely that a patient would die before or after this time. Please now consider just those baseline patients who die before time t_0 and identify a survival time t such that you feel that it is equally likely that such a patient would die before or after this time.”*

Q9 “Please now consider just those baseline patients who survive after time t_0 and identify a survival time t such that you feel that it is equally likely that such a patient would die before or after this time.”

The responses to Questions **Q8** and **Q9** give point assessments for $t_{q(1)}$ and $t_{q(3)}$ and we can calculate a point value for $r = t_{q(3)}/t_{q(1)}$. However we need to find a prior distribution for α and therefore we need an assessment of uncertainty in the ratio $r = t_{q(3)}/t_{q(1)}$. The quantities $t_{q(1)}$ and $t_{q(3)}$ are unlikely to be independent in the expert’s beliefs. Therefore, to assess uncertainty in the ratio we ask directly about it.

Q10 “Your responses to Questions **Q8** and **Q9** suggest that the time when the survival probability for a baseline patient reaches 75% is $t_{q(1)}$ and that the time when the survival probability for a baseline patient reaches 25% is $t_{q(3)}$. This suggests that, if the time when the survival probability reaches 25% is R times the time when it reaches 75%, then R is about r . Please think about the value of R and give a value r_2 such that you think that R is equally likely to be smaller or larger than r_2 .”

Q11 “Now please give a value r_1 which is smaller than r_2 such that R is equally likely to be less than r_1 or between r_1 and r_2 .”

Q12 “Now please give a value r_3 which is larger than r_2 such that R is equally likely to be greater than r_3 or between r_2 and r_3 .”

The three elicited values r_1, r_2, r_3 can be converted into three quartiles $Q_1(\alpha), Q_2(\alpha), Q_3(\alpha)$ for α using (5.8). We assume a gamma prior distribution, $\alpha \sim \text{Gamma}(a_\alpha, b_\alpha)$ for α . The parameters a_α and b_α are chosen to match the lower and upper quartiles α_1, α_3 and we can use the median α_2 as a check of the gamma distribution assumption.

The prior density for α is

$$\pi(\alpha|a_\alpha, b_\alpha) = \frac{b_\alpha^{a_\alpha} \alpha^{a_\alpha-1} \exp\{-b_\alpha \alpha\}}{\Gamma(a_\alpha)}$$

The elicited lower and upper quartiles for α are $Q_1(\alpha), Q_3(\alpha)$. Since b_α is a scale parameter, $Q_3(\alpha)/Q_1(\alpha)$ depends only on a_α , allowing a_α to be determined by numerical iteration. Once a_α is found, b_α is easily found using $b_\alpha = q_{a_\alpha}(0.75)/Q_3(\alpha)$ where $q_a(0.75)$ is the upper quartile of a gamma($a, 1$) distribution.

Ren & Oakley (2013) commented that a value for the Weibull shape parameter could be obtained by eliciting two survival rates say $S(t)$ and $S(t')$ at two times t and t' , where

$t' > t$. The Weibull shape parameter was then given as

$$\alpha = \frac{\log\left(\frac{\log(S(t'))}{\log(S(t))}\right)}{\log\left(\frac{t'}{t}\right)}. \quad (5.9)$$

However, as $S(t)$ and $S(t')$ might be dependent in the expert's beliefs, they suggested constructing a joint prior for α and λ by eliciting beliefs about $S(t)$ and the difference $S(t) - S(t')$.

5.5.1.2 Construction of prior distribution for the coefficients in a linear predictor in the Weibull case

To develop the prior distribution for the coefficients in the linear predictor in the Weibull case we can follow a procedure similar to that given in Section 5.3.1. However some modification is necessary.

Suppose that we use Questions **Q3**, **Q4** and **Q5** to determine the distribution for the baseline parameter. In this case, the survival probability is $p = \exp(-\lambda t_0^\alpha)$. Hence

$$\ln(-\ln p) = \beta_0 + \alpha \ln t_0.$$

Let

$$c_{0,h} = \ln(-\ln p_{4-h})$$

for $h = 1, 2, 3$ where p_2, p_3, p_1 are the responses to questions **Q3**, **Q4** and **Q5** respectively.

We approximate the prior mean for β_0 using $(c_{0,1} + c_{0,3})/2 - E(\alpha) \ln t_0$ and approximate the variance of β_0 using $(c_{0,3} - c_{0,1})/1.349 - (\ln t_0)^2 \text{Var}(\alpha)$.

Similarly, with these adjustments, we can find prior distributions for the linear predictors at other covariate values and hence use the methods described in Section 5.3.1 to find prior distributions for the coefficients of covariates.

5.5.2 Example: Metropolis-Hastings sampling with random walk proposal

The joint posterior density discussed in Section 5.5 did not have a closed form so MCMC methods will be used to sample from the posterior density. The performance of the algorithm will strongly depend on the choice of the proposal distribution. We could use a Metropolis-Hastings algorithm with a random walk proposal.

A multivariate normal distribution will be used as the proposal distribution for the

vector parameter $\underline{\beta}^i$ at the i^{th} iteration. This is given by

$$\underline{\beta}^i \sim N(\underline{\beta}^{i-1}, V)$$

where V is known and the same for all iterations. Hence, the proposal density for a proposed value β^{i*} given the old value β^i , $q(\beta^{i*}|\beta^i)$ and the proposal density of the old value given the proposed value $q(\beta^i|\beta^{i*})$ are the same.

The Weibull shape parameter α will have a gamma proposal distribution. Suppose $\alpha^{i*}|\alpha^i \sim Ga(a, b)$. The mean of α^{i*} is the previous value α^i . So,

$$\frac{a}{b} = \alpha^i$$

and

$$b = \frac{a}{\alpha^i}$$

We propose a new value $\alpha^{i*} \sim Ga(a, b)$. The proposal density of the new value of α , α^* given the current value of α denoted by $q(\alpha^{i*}|\alpha^i)$ has a $Ga(a, a/\alpha^i)$ distribution. The logarithm of the ratio

$$\frac{q(\alpha^i|\alpha^{i*})}{q(\alpha^{i*}|\alpha^i)}$$

is

$$(2a - 1)(\log \alpha - \log \alpha^*) - a(\alpha/\alpha^* - \alpha^*/\alpha)$$

The acceptance ratio which is denoted by A is given by

$$A(\beta^{i*}, \alpha^{i*}|\beta^i, \alpha^i) = \min \left\{ 1, \frac{\pi(\beta^*, \alpha^*|D)}{\pi(\beta, \alpha|D)} \times \frac{q(\beta^i|\beta^{i*})}{q(\beta^{i*}|\beta^i)} \times \frac{q(\alpha^i|\alpha^{i*})}{q(\alpha^{i*}|\alpha^i)} \right\}$$

Hence, the proposal density for a proposed value β^{i*} given the old value β^i , $q(\beta^{i*}|\beta^i)$ and the proposal density of the old value given the proposed value $q(\beta^i|\beta^{i*})$ are the same and therefore cancel from the acceptance ratio. Since the proposal distributions of $\underline{\beta}$, $q(\beta^{i*}|\beta^i) = q(\beta^i|\beta^{i*})$ is symmetric with a multivariate normal distribution, it cancels out from the acceptance probability and is discarded from equation. The proposal distribution of α is not symmetric and hence does not cancel from the acceptance ratio. The acceptance probability is simplified to

$$A(\beta^{i*}, \alpha^{i*}|\beta^i, \alpha^i) = \min \left\{ 1, \frac{\pi(\beta^*, \alpha^*|D)}{\pi(\beta, \alpha|D)} \times \frac{q(\alpha^i|\alpha^{i*})}{q(\alpha^{i*}|\alpha^i)} \right\} \quad (5.10)$$

The acceptance ratio is computed from equation 5.10 and a random sample of a value u is drawn from a uniform distribution on $(0,1)$. We set $(\beta^{i*}, \alpha^{1*}) = (\beta^i, \alpha^i)$ if the value $u < A$ at the end of the iteration. The distribution will tend to the posterior distribution

as i increases. The summary of the posterior can be worked out from the sampled values of the (β, α) after discarding burn in. We note that we could also do Metropolis within Gibbs with separate sampling of α and β .

5.6 Application: Bayesian survival modelling to SNLG data set

We will apply the procedure discussed in Section 5.5 to the SNLG data set described in Chapter 2. The covariates and their notations are given in Table 5.1.

Table 5.1: Covariates and notations

Covariates	Notation	Covariates	Notation
Age	x_1	Extranod	x_8
HB	x_2	Urea	x_9
WBC	x_3	Marrow	x_{10}
Sex	x_4	Bsy	x_{11}
Albumin	x_5	Stage	x_{12}
Ap	x_6	ECOG	x_{13}
Urea	x_7	LDH	x_{14}

For illustration, we suppose that the expert identifies the values for a “baseline” patient following Question Q1 and gives the choice of central values for the continuous covariates x_1, x_2 and x_3 as 60, 110 and 2 respectively. The covariates x_4 to x_{11} are the 2-level categorical covariates. The ordinal covariate x_{12} has values 1, 2, 3, 4, x_{13} has values 1, 2, 3, 4, 5 and x_{14} has values 1, 2, 3.

We will suppose that the overall survival (OS) lifetime $T_i \sim \text{Weibull}(\lambda_i, \alpha)$, where the hazard multiplier λ_i depends on the linear predictor η_i which is given by

$$\eta_i = \beta_0 + \beta_1 x_{i,1} \dots + \beta_{11} x_{i,11} + \sum_{k=1}^4 \beta_{12,k} \delta_{i,12,k} + \sum_{k=1}^5 \beta_{13,k} \delta_{i,13,k} + \sum_{k=1}^3 \beta_{14,k} \delta_{i,14,k} \quad (5.11)$$

where $\delta_{i,j,k} = 1$ if $x_{i,j} = k$ and $\delta_{i,j,k} = 0$ otherwise for $j = 12, 13, 14$.

We construct the design matrix $X = x_1, \dots, x_{S'}$ where S' is the number of columns of the design matrix and it depends on the number of covariates and the number of categories of the categorical covariates. We refer to our discussion in Section 4.4.3 to include categorical covariates (also known as “factors”) by considering a factor with m levels (*ie* possible values). The model also includes the overall constant β_0 . This means that using m parameters would be over-parameterising and the parameters would not be identifiable. Therefore a factor with m levels contributes $m - 1$ columns to the design

matrix X . Consider a two-level factor such as Sex. One way to put this in the model is to introduce a column in X which has a 1 for each male patient and a -1 for each female patient. There is a single parameter associated with this column. Another method would be to use 0 and 1 instead of 1 and -1 . This has the disadvantage that the prior variance for η_i for a male patient would be less than that for a female patient who has the same covariate values and otherwise. Other 2-level factors, such as those with values “Normal” and “Abnormal” can be treated in the same way. Each gets one column in X and one parameter. When $m > 2$ we get $m - 1$ columns in X and $m - 1$ parameters. In some ways, the easiest thing to do is to choose a baseline level for the factor and choose this as the one which does not get a parameter (like the second method for Sex above). We could revisit the example of a 4-level factor discussed in Section 4.4.3.

The number of regression parameters is 24. We also have the Weibull shape parameter α .

5.6.1 Prior distributions

We construct the prior distribution for the parameters of the linear predictor. We let the reference time elicited in response to Question Q2 be $t_0 = 3$. We will also let the response to Question Q3, Q4 and Q5 be 0.5, 0.4 and 0.6 respectively. Thus, we elicit the three quartiles p_1 , p_2 and p_3 at time t_0 as 0.6, 0.5 and 0.4 and we follow Equation (5.1) to get the quartiles of β_0 : $\beta_{0,1} = -1.77$, $\beta_{0,2} = -1.5$ and $\beta_{0,3} = -1.23$. Using a normal prior distribution, the mean and standard deviation of β_0 are given as $\mu_0 = -1.5$ and standard deviation $\sigma_0 = 0.4$.

For illustration, we construct the prior distribution for the coefficient of a quantitative variable using the covariate “age”. We let the response to Question Q6 be 70 years and thus $\delta_j = 10$. We follow Section 5.3.1.1 and we elicit a normal distribution for the log-hazard of $\beta_0 + \beta_j \delta_j$. Thus, we have $E(\beta_0 + \beta_j \delta_j) = -1.1$ and $\text{Var}(\beta_0 + \beta_j \delta_j) = 0.238$. We follow Equation 5.2 and 5.3 to get the prior distribution for β_{age} as mean $m = 0.04$ and standard deviation $\sigma = 0.028$. We perform a similar process for all other coefficients of the continuous covariates.

We will construct the prior distribution for the parameter of sex β_{sex} by thinking that it is a quantitative covariate with $\delta_j = 1$ following Section 5.3.1.3. We get the prior distribution for β_{sex} as mean $m = 0.046$ and standard deviation $\sigma = 0.15$. Other 2-level factors in the example data set are covariates with values such as “Normal” and “Abnormal”. These can be treated in the same way as sex.

We construct the prior distribution for the coefficient of the categorical covariates using the method of 5.3.1.4 with modification given in 5.5.1.2. We used judgements about survival at three years so $t_0 = 3$. As a baseline, we set all categorical variables to level 1. The assessed values of p_1 and p_3 were 0.232 and 0.735. These give $c_{0,1} = -1.179$ and

$c_{0,3} = 0.379$, leading to $E(\beta_0^*) = -1.5$ and $\text{Var}(\beta_0^*) = 0.25$. Following 5.3.1.4, we assess the variances of β_0 and the coefficients of the covariates LDH, ECOG and Stage.

We would want to construct the prior distribution of the Weibull shape parameter α as discussed in Section 5.5.1.1. For illustration, we let our responses to Questions Q8 and Q9 be 2 and 4 years and hence $t_{q(1)} = 2$ and $t_{q(3)} = 4$. In response to Questions Q10, Q11 and Q12, we have that the three quartiles of α are 0.6, 0.9 and 1.2 respectively.

We would want the prior distribution of the Weibull shape parameter in the form of a gamma distribution given by $\alpha \sim \text{Gamma}(a, b)$. We elicit the lower and upper quartiles of α as 0.6 and 1.2. The ratio of the upper quartile to the lower quartile which is 2 only depends on the value of the shape parameter a . We determine the value of a by numerical iteration and b as discussed in Section 5.5.1.1. The values of a and b are given as 4 and 4 respectively.

The prior means and standard deviations of the coefficients of linear predictors are given in Table 5.2. We refer to Section 4.4.3.4 for discussions on the δ quantities.

Table 5.2: The prior means and standard deviations for the unknown parameters

Parameter	prior mean	prior variance
β_0 (baseline parameter)	-1.500	0.400
β_1 (Age)	0.040	0.030
β_2 (HB)	0.020	0.014
β_3 (WBC)	0.080	0.060
β_4 (Sex)	0.050	0.150
β_5 (Albumin)	0.000	0.173
β_6 (Ap)	0.000	0.173
β_7 (Urea)	0.000	0.173
β_8 (Extranod)	0.000	0.173
β_9 (Bulk)	0.000	0.173
β_{10} (Marrow)	0.000	0.173
β_{11} (Bsy)	0.000	0.173
$\delta_{12,1}$	0.000	0.14
$\delta_{12,2}$	0.000	0.077
$\delta_{12,3}$	0.000	0.055
$\delta_{13,1}$	0.000	0.14
$\delta_{13,2}$	0.000	0.077
$\delta_{13,3}$	0.000	0.055
$\delta_{13,4}$	0.000	0.45
$\delta_{14,1}$	0.000	0.077
$\delta_{14,2}$	0.000	0.45
α	1	0.5

Table 5.3: The posterior means and standard deviations (std) for the unknown parameters

Parameter	prior mean	prior std	posterior mean	posterior std
β_0 (baseline parameter)	-1.500	0.400	0.352	0.194
β_1 (Age)	0.040	0.030	0.009	0.005
β_2 (HB)	0.020	0.014	0.002	0.003
β_3 (WBC)	0.080	0.060	0.050	0.018
β_4 (Sex)	0.050	0.150	-0.034	0.059
β_5 (Albumin)	0.000	0.173	-0.020	0.074
β_6 (Ap)	0.000	0.173	0.048	0.070
β_7 (Urea)	0.000	0.173	-0.097	0.067
β_8 (Extranod)	0.000	0.173	0.073	0.066
β_9 (Bulk)	0.000	0.173	-0.154	0.060
β_{10} (Marrow)	0.000	0.173	-0.171	0.088
β_{11} (Bsy)	0.000	0.173	-0.044	0.065
$\beta_{12,1}$ (Stage 1)	0.000	0.272	-0.048	0.087
$\beta_{12,2}$ (Stage 2)	0.000	0.272	-0.048	0.073
$\beta_{12,3}$ (Stage 3)	0.000	0.363	0.037	0.091
$\beta_{12,4}$ (Stage 4)	0.000	0.495	0.055	0.126
$\beta_{13,1}$ (ECOG 1)	0.000	0.722	-0.134	0.136
$\beta_{13,2}$ (ECOG 2)	0.000	0.722	-0.143	0.134
$\beta_{13,3}$ (ECOG 3)	0.000	0.813	0.148	0.150
$\beta_{13,4}$ (ECOG 4)	0.000	0.945	0.169	0.182
$\beta_{13,5}$ (ECOG 5)	0.000	7.2	-0.041	0.468
$\beta_{14,1}$ (LDH 1)	0.000	0.527	-0.113	0.059
$\beta_{14,2}$ (LDH 2)	0.000	0.527	-0.034	0.055
$\beta_{14,3}$ (LDH 3)	0.000	1.8	0.147	0.086
α	1	0.5	1.03	0.043

5.6.2 Posterior distributions

The joint posterior distribution have been given in Section 5.5.1. The Metropolis-Hastings within Gibbs algorithm was applied using R JAGS software (Plummer, 2012) (see R code in Appendix A.4.1 and an explanation of how R JAGS works for survival data with right censoring in Appendix A.1.2). Following a burn-in of 5000 iterations of the sampler, 100000 iterations were taken. Convergence was checked using two chains starting from very different values. Visual inspection of the trace plots of the covariate parameters showed that the mixing appeared very satisfactory. The posterior numerical summaries of the parameters are also given in Table 5.3.

5.6.3 Calculation of the prognostic index for the Weibull survival modelling to SNLG data set

We recall the formula of the linear predictor in Equation 3.2 as follows

$$\log \lambda_i = \eta_i = \beta_0 + \sum_{s=1}^S \beta_s x_{i,s} \quad (5.12)$$

The linear predictor has a linear structure with the posterior values of covariate effects $\underline{\beta} = \beta_0, \beta_1, \dots, \beta_S$ which are given in Table 5.3. We can obtain the expectation of the linear predictor by substituting the posterior means into the formula. We have that the expectation of the linear predictor for a new individual i' is

$$\log \lambda_{i'} = \eta_{i'} = \beta_{0'} + \sum_{s=1}^S \beta_{s'} x_{i',s} \quad (5.13)$$

where $\beta_{0'}$ is the posterior expectation of β_0 and $\beta_{s'}$ is the corresponding posterior expectation of β_s . We might prefer to index the linear predictor in a range (0, 100). We do this by finding $100\Phi^{-1}\left(\frac{\eta_{i'} - m}{s}\right)$ where $\Phi(\cdot)$ is the standard normal distribution function and m and s are the sample mean and sample standard deviation of the values of η for all patients in the SNLG data set using the posterior means of β_0, \dots, β_S . The function to do this is given in Appendix A.4.2. For instance, a patient with the covariate vector $\underline{x} = (72, 1, 4, 4, 2, 123, 1.9, 2, 2, 2, 1, 1, 2, 2)^T$ for the covariates age, sex, stage, ECOG, LDH, HB, WBC, albumin, AP, urea, extranod, bulk, marrow and bsy respectively will have an index of 83. This will mean that the patient has an index of 83 on a scale from 0 to 100 and this is a great risk value.

5.7 Bayesian survival modelling using the piecewise constant hazard model

We may wish to apply Bayesian inference to our discussions on the piecewise constant hazard model with time varying covariate effects as discussed in Section 3.5. Much work has been done on piecewise constant hazard models where the hazard parameters are independent between intervals. See, for example Genest & Kalbfleisch (1988); Ibrahim *et al.* (2001). Kim *et al.* (2007) proposed a class of semi-parametric models for survival data in which they constructed a dynamic model for piecewise constant hazard functions over a finite partition of the time axis. Gamerman (1991) introduced the dynamic generalised linear model into survival analysis where he applied it to the piecewise constant hazard model.

We will need to choose the cut points for the piecewise constant hazard model.

5.7.1 Choosing the cut points of time intervals for the piecewise constant hazard model

The piecewise constant hazard model requires the discretisation of the time axis into intervals and thereby choosing the number and locations of the cut points. Some authors have suggested defining the intervals as beginning and ending at the observed failure times while Kalbfleisch & Prentice (1973) suggested selecting intervals independently of the data. West (1992), in an attempt to define intervals, suggested shorter intervals over the first few years since deaths in cancer data are common in early stages and longer intervals in the later years since there are fewer deaths. One possibility is to choose our cut points deliberately so that we expect to have the same number of deaths in each interval. A researcher would think of a certain number of cut points and expect equal portions of death to be in each time interval using prior judgements.

Suppose we choose to have ten time intervals, we might then think ordinarily that we should have approximately 10% of the deaths in each interval. We might also suppose that the event times will have approximately an exponential distribution. The survival function of an exponential distribution with parameter λ is given by $\exp\{-\lambda t\}$.

Given the cut points $\tau_1 < \tau_2 < \dots < \tau_{j-1}$, the probability of surviving until τ_j will be

$$\exp\{-\lambda\tau_j\} = 1 - 0.1j$$

so

$$-\lambda\tau_j = \log(1 - 0.1j)$$

and

$$\tau_j = -\frac{1}{\lambda} \log(1 - 0.1j)$$

We let the mean of the exponential distribution ($\frac{1}{\lambda}$) be ν and $\zeta = \frac{1}{j}$. So,

$$\tau_j = -\nu \log(1 - j\zeta) \tag{5.14}$$

5.7.2 Construction of the likelihood for piecewise constant hazard model

We will suppose that associated with every patient is a time which could either be a death or censoring time. Every interval is associated with three different groups of patients, patients who died during the interval, patients who were censored during the interval and patients who survived the interval. The contribution of the likelihood from the patients in every interval will depend on the three different groups of patients. The likelihood

contribution L of the patients is given as

$$L = \prod_{i=1}^n \prod_{j=1}^J L_{i,j}$$

where

$$L_{i,j} = \begin{cases} 1 & \text{if } t_i < \tau_{j-1} \\ \lambda_{i,j}^{\delta_{i,j}} \exp \{-\lambda(t_i - \tau_{j-1})\} & \text{if } \tau_{j-1} \leq t_i < \tau_j \\ \exp \{-\lambda(\tau_j - \tau_{j-1})\} & \text{if } t_i \geq \tau_j \end{cases}$$

where n is the number of patients

J is the number of time intervals

t_i is the event or censoring time of the i^{th} patient.

$\lambda_{i,j}$ is the hazard of the i^{th} patient in the j^{th} interval.

$\delta_{i,j}$ is the indicator of death or censoring of the i^{th} patient in the j^{th} interval.

Within each interval the conditional survival distribution given that the patient is alive and uncensored at time τ_{j-1} , is exponential since the hazard is constant.

5.7.3 Construction of prior distribution for the parameters of piecewise constant hazard model

In a Bayesian context, we have the advantage of constructing a prior that makes the hazard parameters in neighbouring intervals to be correlated. In practice, we could use a piecewise constant hazard function which has prior distribution in which the parameters are correlated over time. Genest & Kalbfleisch (1988) and Ibrahim *et al.* (2001) made the prior distribution of the parameters independent between time intervals. However, it would be reasonable to think that the hazards in the intervals which are closely together are likely to be similar. McKeague & Tighiouart (2000) gave a dependent prior using a Markov random field.

We will assume that the prior used in the illustration of the piecewise constant hazard model will take the form of a realisation of a stochastic process which could either be stationary or non stationary. If the prior is made stationary then each parameter gets the same variance in each time period. For example, we might use an autoregressive process with autoregressive parameter which governs how strong the correlation between time periods will be. We choose the autoregressive parameter $\rho > 0$ to give positive autocorrelation. We will choose a first order autoregressive process such that for any given value of autoregressive parameter ρ , the process is given by

$$\beta_s - \mu = \rho(\beta_{s-1} - \mu) + \epsilon_s$$

where ϵ_s is normally distributed with mean zero and variance σ^2 (Chatfield, 2004). The variance of the process is

$$\text{Var}(\beta_s) = \sigma^2 \sum_{p=0}^{\infty} \rho^{2p} = \frac{\sigma^2}{1 - \rho^2}$$

Making the process stationary makes the prior variance the same in all intervals. We could also choose the value of ρ following Revie *et al.* (2010) on the construction of correlations.

5.7.4 Construction of prior distribution for the frailty variance

Supposing that the logarithm of the frailty has a normal distribution with zero mean and a variance of σ_f^2 . We can construct the prior distribution for the frailty variance σ_f^2 . We might suppose that the precision $\tau_f = \sigma_f^{-2}$ has a gamma prior distribution.

If the parameters remained constant over all time intervals then an individual's lifetime would have an exponential distribution since the hazard λ of the individual is constant. The individual's mean lifetime will be $1/\lambda$. Doubling λ would half the mean lifetime and correspond to a log-frailty of $\log 2$. Suppose, for example, we had $\sigma_f = \log 2$ then this corresponds to $\tau_f = \frac{1}{(\log 2)^2}$. Thus, we might set the prior mean for τ_f to be $\frac{1}{(\log 2)^2}$. So, if $\tau_f \sim \text{Ga}(a_f, b_f)$, then

$$\frac{a_f}{b_f} = \frac{1}{(\log 2)^2} \quad (5.15)$$

The prior variance of τ_f is $\frac{a_f}{b_f^2}$ and the coefficient of variation is $\frac{1}{\sqrt{a_f}}$. We choose a_f to reflect our prior uncertainty in τ_f . Since we have little prior information on τ_f we choose a small value for a_f . However, so that the prior density of τ_f is at τ_f , we choose $a_f > 1$, say 1.1. Combining this with Equation 5.15 gives $b_f = 0.53$.

5.8 Application: Bayesian survival modelling using the piecewise constant hazard model

We will apply the Bayesian survival modelling using the piecewise constant hazard model to the two example data sets described in Chapter 2. We follow Section 5.7.2 and give further explanation in Appendix A.1.3 for the construction of the likelihood contribution. We will apply Equation (5.14) to the two example data sets to choose the cut points. We may make a prior assessment of the mean survival life time of a patient, ν as 3 years in the SNLG data set, and we get the cut points as

0.105, 0.223, 0.357, 0.511, 0.693, 0.916, 1.204, 1.609, 2.303.

We may also assess the mean survival life time of a patient, ν as 500 days in the leukaemia

data set and we get the cut points as

52.68, 111.57, 178.34, 255.41, 346.57, 458.15, 601.99, 804.72, 1151.29.

We will use the same prior means and variances of the parameters of the SNLG data set given in Table 5.2. The prior means and standard deviations of the leukaemia data set will also be constructed in a similar way and given in Table 5.4. We would want to think of the value of the autoregressive parameter as discussed in Section 5.7.3 by considering how much the variance of the parameter will reduce in the next time periods if we know the value in the first time period. We suppose that 90% of the variance in the next time period is explained. This corresponds to a coefficient of determination $r^2 = 0.9$ and correlation $r = 0.95$.

Table 5.4: The prior means and standard deviations (std) for the coefficients of the covariates

Parameter	prior mean	prior std
β_0	-6.90	0.12
β_{age}	0.040	0.030
β_{sex}	0.050	0.150
β_{Wbc}	0.080	0.173
β_{Depsc}	0.120	0.110

The time dependent effects of covariates were illustrated using both the leukaemia and the SNLG data sets using the piecewise constant hazard model. Two versions of the model were fitted, one without and one with log normal frailties. We would refer to Section 5.7.4 for the prior distribution of the precision of the logarithm of the frailty. Posterior distributions for the covariate effects were evaluated using the RJAGS software and the functions for both SNLG and leukaemia data set are given in Appendix A.4.3 and Appendix A.4.4 respectively. We have plotted the effects of the covariates over time. Figure 5.3 shows time plots for the posterior means and ± 2 standard deviation intervals for the coefficients of age, sex and baseline with frailty and without frailty in the model using the SNLG data set. The posterior means and standard deviations for some of the parameters for each interval for the SNLG data set are given in Table 5.5.

The age effect decreases and later increases from one interval to another while the sex effect increases over time and later decreases over time (see Table 5.5). Figure 5.4 shows time plots for the posterior means and ± 2 standard deviation intervals for the coefficients of age, sex and baseline with frailty and without frailty in the model using the leukaemia data set. The posterior means and standard deviations for some of the parameters for each interval for the leukaemia data set are given in Table 5.6. The age effect decreases from one interval to another while the sex effects increases from one interval to another

Table 5.5: Posterior numerical summaries of some selected parameters in each interval in the SNLG model

j	τ_j	β_{age}	β_{sex}	β_0
1	0.105	0.031(0.007)	0.038(0.074)	-0.434(0.189)
2	0.223	0.026(0.007)	0.052(0.069)	-0.289(0.176)
3	0.357	0.019(0.007)	0.058(0.068)	-0.294(0.179)
4	0.511	0.022(0.007)	0.059(0.068)	-0.390(0.181)
5	0.693	0.018(0.008)	0.054(0.071)	-0.624(0.191)
6	0.916	0.026(0.009)	0.044(0.075)	-0.738(0.201)
7	1.204	0.030(0.009)	0.047(0.080)	-0.894(0.213)
8	1.609	0.035(0.010)	0.047(0.082)	-0.992(0.223)
9	2.303	0.040(0.011)	0.045(0.085)	-1.040(0.238)
10	∞	0.039(0.013)	0.042(0.09)	-1.051(0.247)

Table 5.6: Posteriors numerical summaries of some selected parameters in each interval in the leukaemia model

j	τ_j	β_{age}	β_{sex}	β_0
1	52.6	0.069(0.005)	-0.055(0.049)	-6.04(0.062)
2	111.6	0.053(0.005)	-0.064(0.049)	-6.10(0.055)
3	178.3	0.032(0.005)	-0.056(0.052)	-6.20(0.053)
4	255.4	0.027(0.005)	-0.071(0.053)	-6.30(0.054)
5	346.6	0.023(0.005)	-0.069(0.057)	-6.39(0.054)
6	458.1	0.010(0.006)	-0.068(0.060)	-6.48(0.057)
7	601.99	0.015(0.006)	-0.060(0.060)	-6.59(0.058)
8	804.7	0.015(0.005)	-0.077(0.060)	-6.71(0.059)
9	1151.3	0.023(0.006)	-0.089(0.061)	-6.82(0.060)
10	∞	0.036(0.005)	-0.10(0.065)	-6.92(0.062)

in some cases and decreases in other cases over time (see Table 5.6). The dependence on time was not obvious in some cases when frailty was included in the model.

The posterior means of the effects of age and sex in the SNLG data set are positive in all time intervals in both the cases with and without including frailty. The posterior means of the baseline parameter are negative in all time intervals. The effect of age decreases initially before increasing again over time. The effect of sex appears very similar over time in both the cases with and without including frailty (see Figure 5.3).

The posterior means of the effects of sex and baseline parameter in the leukaemia data set are negative in all time intervals in both the case with or without including frailty. The posterior means of the effect of age is positive in all time intervals. The effect of age decreases initially before increasing again over time. The effect of sex appears very similar

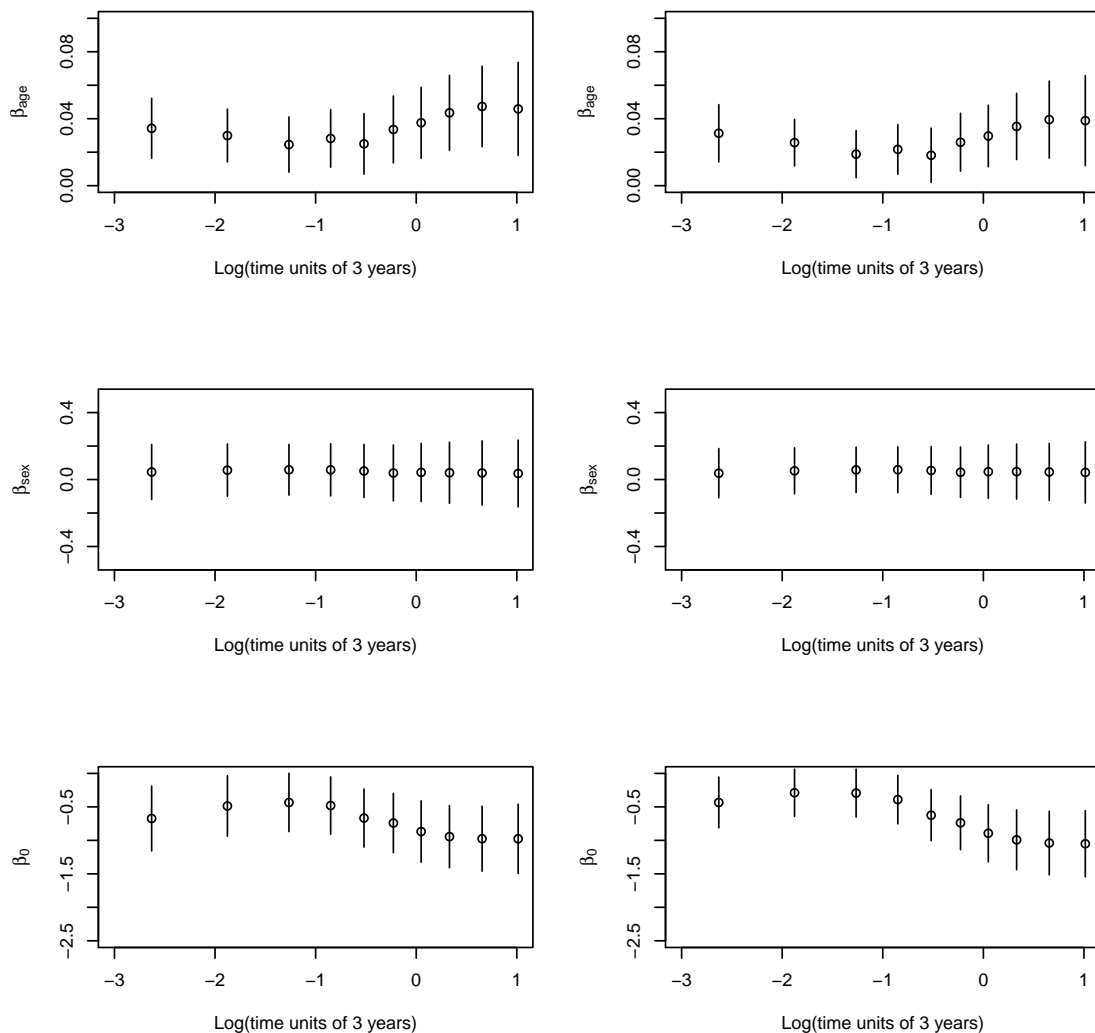


Figure 5.3: Time plots of the posterior means and ± 2 standard deviation intervals for the coefficients of age, sex and baseline with including frailty (left column) and without frailty (right column) using the SNLG data set

over time in both the case with and without including frailty (see Figure 5.4).

Introducing frailty sometimes reduces the apparent dependent covariate effect. Including frailty might be a less flexible way to allow covariate effects to depend on time than the piecewise constant hazard. Henderson & Oman (1999) showed that ignoring frailty when present can lead to inappropriate estimated values of the parameters and hence inaccurate survival plots.

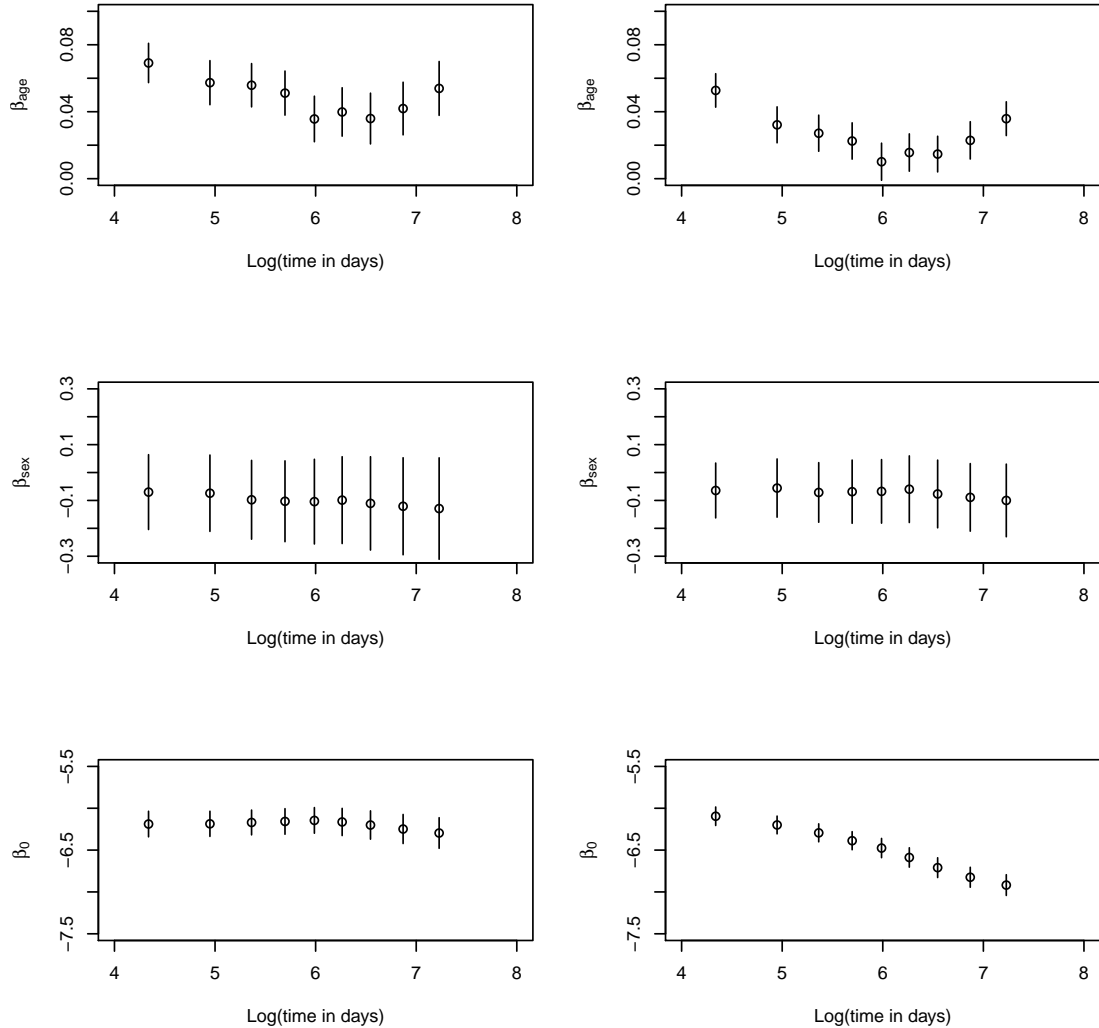


Figure 5.4: Time plots of the posterior means and ± 2 standard deviation intervals for the coefficient of age, sex and baseline with including frailty (left column) and without frailty (right column) in the model using the leukaemia data set

5.9 A Prognostic index based on the piecewise constant hazard model

5.9.1 Introduction

The prognostic index in a standard proportional hazard model is the linear predictor or some function of it, that is the logarithm of the hazard multiplier (see Section 3.3.1) while for the accelerated life model it is the logarithm of the acceleration factor or time multiplier (see Section 3.3.2). A prognostic index in the case of the piecewise constant hazard model

can not be constructed in the same way as the linear predictor changes from one time interval to another. It is not obvious what the prognostic index should be in the piecewise constant hazard model. We could use the survival probability at a particular time or the predictive median survival time as the prognostic index, for example.

We will need to compute the predictive survival probability in order to work out the predictive median survival time. The joint density of the lifetime T and the model parameter $\underline{\beta}$ is given by

$$f(t, \underline{\beta} | \underline{x}) = f(t | \underline{\beta}, \underline{x}) f(\underline{\beta})$$

Given values for the model parameters $\underline{\beta}$ and the covariate values \underline{x} , the survival function is then given by the probability

$$\Pr(T > t | \underline{\beta}, \underline{x})$$

If all the covariates are observed, the predictive survival function is given by

$$\int_{\underline{\beta}} \Pr(T > t | \underline{\beta}, \underline{x}) f(\underline{\beta}) d\underline{\beta} \tag{5.16}$$

5.9.2 Computation of survival probability at a fixed time

We wish to evaluate the integral in Equation (5.16). We propose that, in routine use with new patients, we represent the posterior distribution of the parameters by storing a set of sampled values from the MCMC computation. In order that we do not store so many samples in memory, MCMC samples are thinned with the reason that samples from the MCMC might be autocorrelated as discussed in Section 4.6.9. The following steps can be used to find the predictive survival probability at any fixed time t^* .

- We would calculate the survival probability at each cut point τ_j where $\tau_j < t^*$ for each vector of sampled parameter values. Suppose we have cut points $\tau_0 = 0, \tau_1 = \dots, \tau_J \rightarrow \infty$ with J intervals and hazards $h_{i1}, h_{i2}, \dots, h_{iJ}$ for a new case. We would want a predictive survival probability for a new case or a particular covariate profile. Then, the probability that the individual survives the first interval is

$$\Pr(T > \tau_1) = \exp \{-\tau_1 h_{i1}\}$$

Similarly, the probability that the individual survives the second interval is

$$\begin{aligned} \Pr(T > \tau_2) &= \exp \{-\tau_1 h_{i1}\} \exp \{-(\tau_2 - \tau_1) h_{i2}\} \\ &= \exp \{-\tau_1 h_{i1} - (\tau_2 - \tau_1) h_{i2}\} \end{aligned}$$

and the j^{th} interval is

$$\Pr(T > \tau_j) = \exp \left\{ - \sum_{m=1}^j h_{im}(\tau_m - \tau_{m-1}) \right\}$$

Finally

$$\Pr(T > t^*) = \exp \left\{ - \sum_{m=1}^{j^*-1} h_{im}(\tau_m - \tau_{m-1}) + h_{ij^*}(t^* - \tau_{j^*-1}) \right\}$$

where j^* is the smallest j such that $t^* < \tau_j$.

- The stored survival probabilities are then averaged over all sampled parameters sets. The averaged survival probability is the predictive survival probability of the patient at time t^* .

5.9.3 Practical computation of predictive median survival time

In order to find the predictive median survival time, we will first of all find the predictive survival probability, $\Pr(T > \tau_j)$ for $\tau_1, \tau_2 \dots$ until $S(\tau_j) < 0.5$. An iterative method can be used to find the predictive median survival time (t_m) such that $\Pr(T > t_m) = 0.5$ since we know the interval in which t_m falls. We could, for example, find where in the interval the predictive median survival time t_m is by using interval halving. We do interval halving by moving half way between the lower and the higher time limits of the interval until it eventually converges. In cases, where the predictive median survival time falls in the last interval, the upper limit would be infinity and the interval halving will not be direct. We could transform the life times to a function x where $x \in [0, 1)$. However, we propose a more efficient algorithm as follows.

5.9.4 An iterative algorithm for finding the predictive median survival time

We could avoid the problem of looking for the upper or lower limit and using interval halving as discussed in Section 5.9.3 by using the following algorithm.

Let $S(t) = \Pr(T > t)$ be the predictive survival probability at time t (i.e the function evaluated at time t). Suppose that we have evaluated $S(\tau_{j-1})$ and $S(\tau_j)$ and $S(\tau_{j-1}) > 0.5 > S(\tau_j)$ so $\tau_{j-1} < t_m < \tau_j$ where t_m is the predictive median survival time.

Let $S(t|\lambda) = \Pr(T > t|\lambda)$ be the survival function for a given value of λ . Now,

$$S(t|\lambda) = S(\tau_{j-1}) \exp \{-\lambda(t - \tau_{j-1})\} \quad \text{for } (\tau_{j-1} \leq t < \tau_j)$$

So,

$$\log[S(t|\lambda)] = \log[S(\tau_{j-1})] - \lambda(t - \tau_{j-1})$$

which is linear in t . We use a locally linear approximation to $g(t) = \log[S(t)]$ within $\tau_{j-1} \leq t < \tau_j$. We require t_m such that $g(t_m) = \log 0.5 = -\log 2$.

Algorithm

Let $t_1 = \tau_{j-1}$. Calculate $g_1 = g(t_1)$.

Let $t_2 = \tau_j$, if $j < J$ or $t_2 = 2\tau_{J-1}$ if $j = J$.

Calculate $g_2 = g(t_2)$.

For $k = 3, 4, \dots, k_{max}$

{ If $(t_{k-1} - t_{k-2})^2 > \delta$ then { Calculate

$$b = \frac{g_{k-1} - g_{k-2}}{t_{k-1} - t_{k-2}}$$

Calculate

$$t_k = t_{k-2} - \frac{1}{b}(g_{k-2} + \log 2)$$

}

Else Stop

}

5.9.5 Application: Finding the predictive median survival time and survival probability at some fixed time

The MCMC samples from the example in Section 5.8 were thinned and 1000 samples were retained. The procedure discussed in Section 5.9.3 and Section 5.9.2 was used to find the predictive median survival time and the survival probability at 3 years. The algorithm discussed in Section 5.9.4 was written using a R function (see Appendix A.4.5). The plot of the predictive median survival time against the 3 years survival probability of the SNLG data set is shown in Figure 5.5. The plot of the predictive median survival time against the 500 days survival probability of the leukaemia data set is shown in Figure 5.6.

Both plots revealed that the predictive median survival time and the fixed time survival probability have a close relationship over the region where most values occur. We would also have a closer look at both plots by restricting the plots to a smaller range (see Figure 5.7 and 5.8). Figure 5.8 reveals some kind of patterns or clustering of points on the plot. We identified the corresponding patients by ordering the covariates by the values of either the predictive median survival time or fixed time survival probability. Further investigation revealed that the clustering was caused by discrete or categorical covariates of groups which correspond to different values of that covariate.

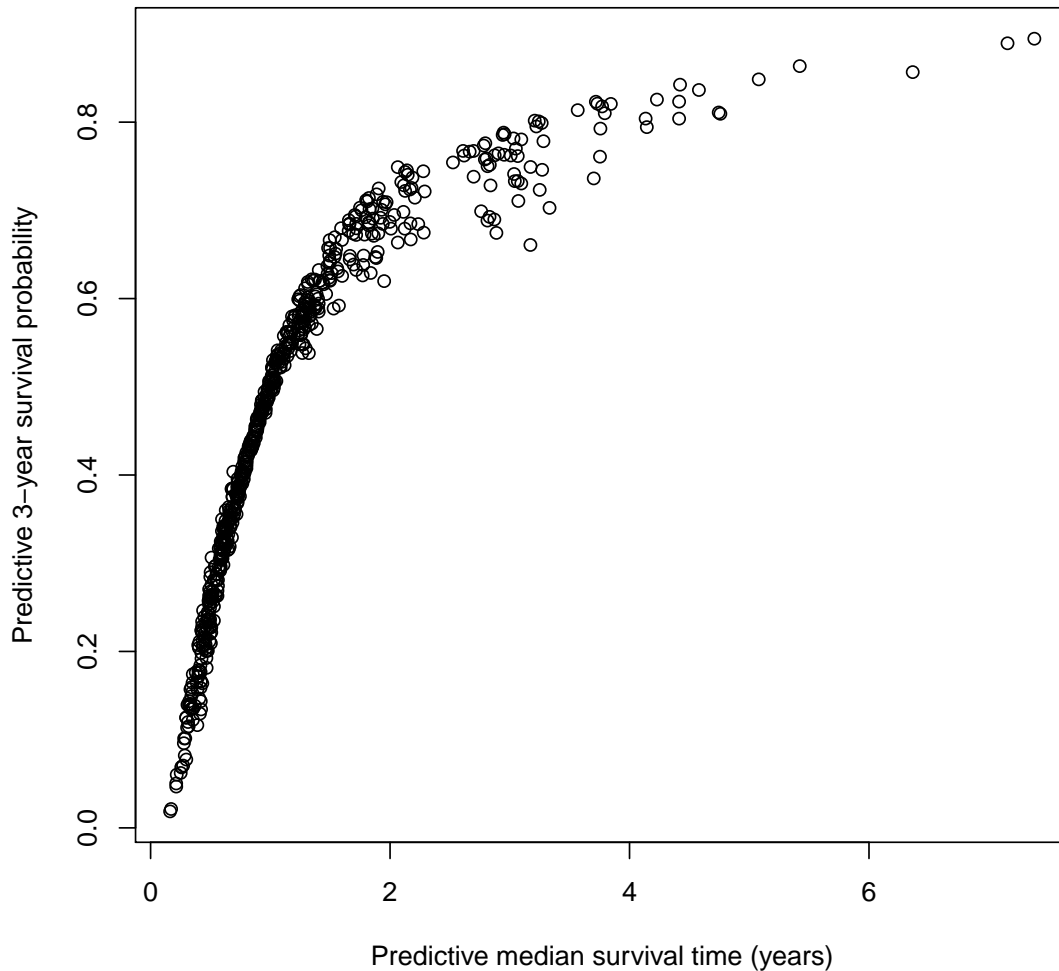


Figure 5.5: Plot of the predictive median survival time against the 3 years survival probability using a piecewise constant hazard model for the SNLG data set

In this application, there may be little to be gained by calculating the predictive median survival time rather than the simpler fixed time survival probability. The fixed time survival probability appears to be just as good as an index as the predictive median survival time does. The computation of the predictive median survival time is harder than that of the fixed time survival probability. We would have the problem of the choice of time when working out the fixed time survival probability. Another problem is that the time that is most informative for one patient might be different for another patient.

Another possibility of calculating the prognostic index of a piecewise constant hazard model would be to use the survival probabilities at more than 1 time. By calculating the

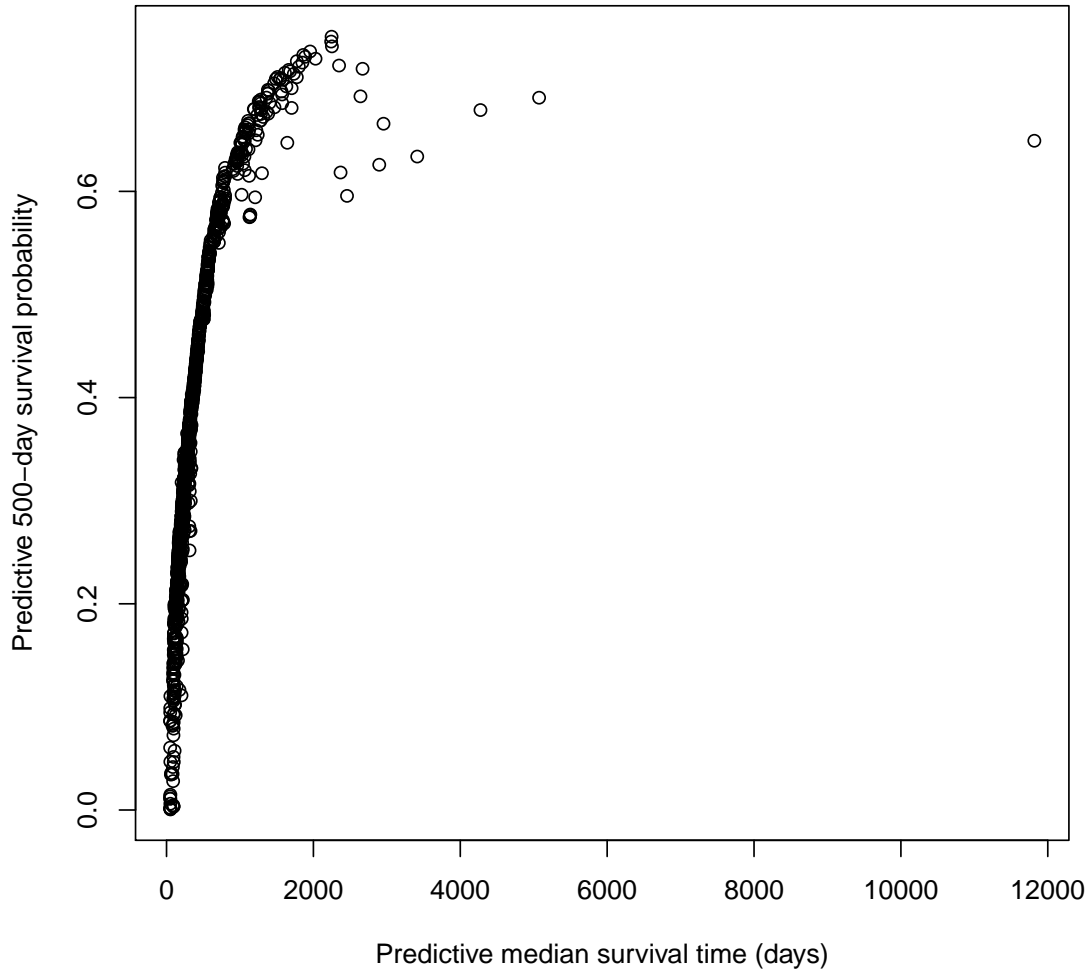


Figure 5.6: Plot of the predictive median survival time against the 500 days survival probability using a piecewise constant hazard model for the leukaemia data set

predictive survival probabilities at a suitable range of times, such as the cut points, we could use simple interpolation to give an approximate predictive median survival time.

5.10 Summary

The Weibull survival model is simple to handle and is widely used in survival analysis since the hazard function is easy (Collett, 1994). This model assumes there is proportionality in the hazards and hence the covariate effects stay the same at all time points. Here, we specify the form of the baseline hazard using a Weibull distribution. This was

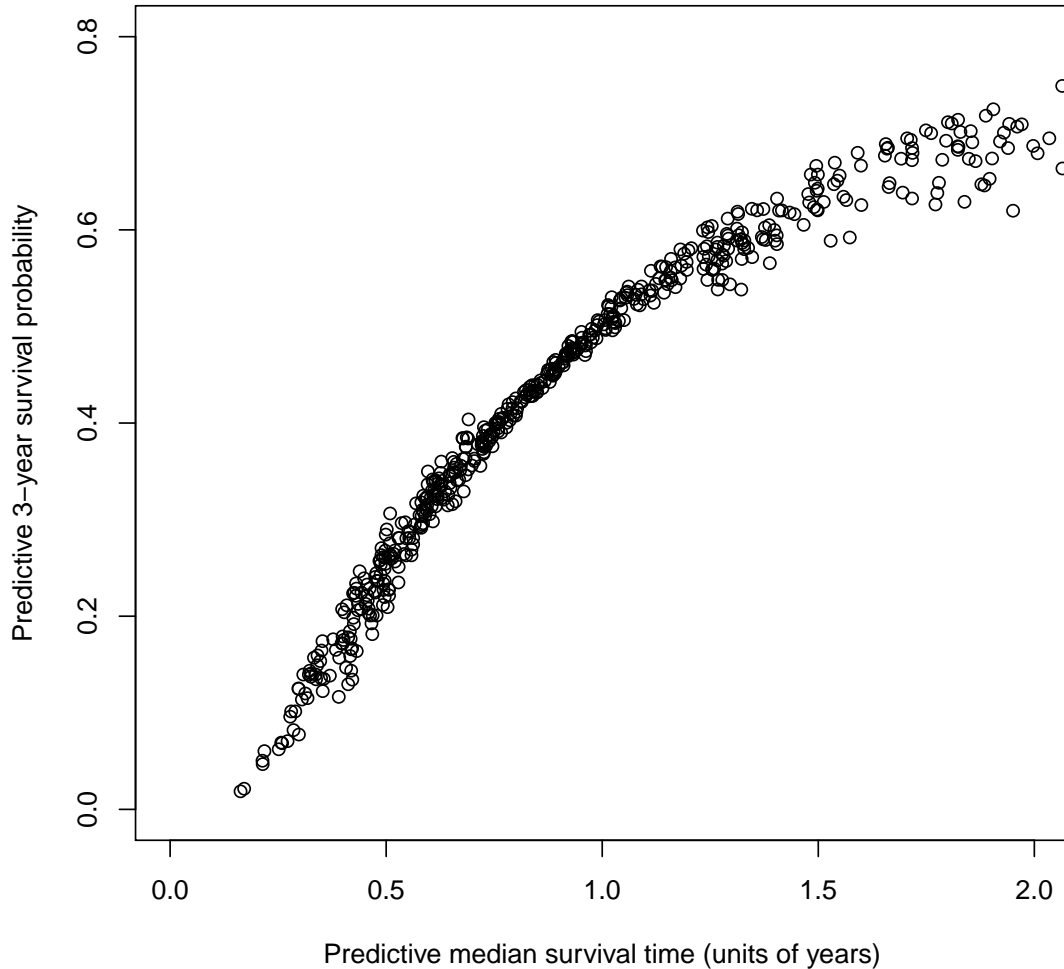


Figure 5.7: Plot of the predictive median survival time against the 3 years survival probability using a piecewise constant hazard model for the SNLG data set on a restricted axis

demonstrated with the SNLG data set. The JAGS software was used to implement this and we provided posterior summaries for the coefficient of the covariates.

We discussed the piecewise constant hazard model where the form of the dependence of the hazard function on the covariates is not specified. This has the advantage of not imposing the overall shape of the hazard function. We relaxed the proportional hazard assumption with a non-proportional hazard using the piecewise constant hazard model. We showed time dependent covariate effects using both the SNLG and leukaemia data sets. The predictive median survival time and the fixed time survival probability were discussed as ways of calculating the prognostic index of the piecewise constant hazard

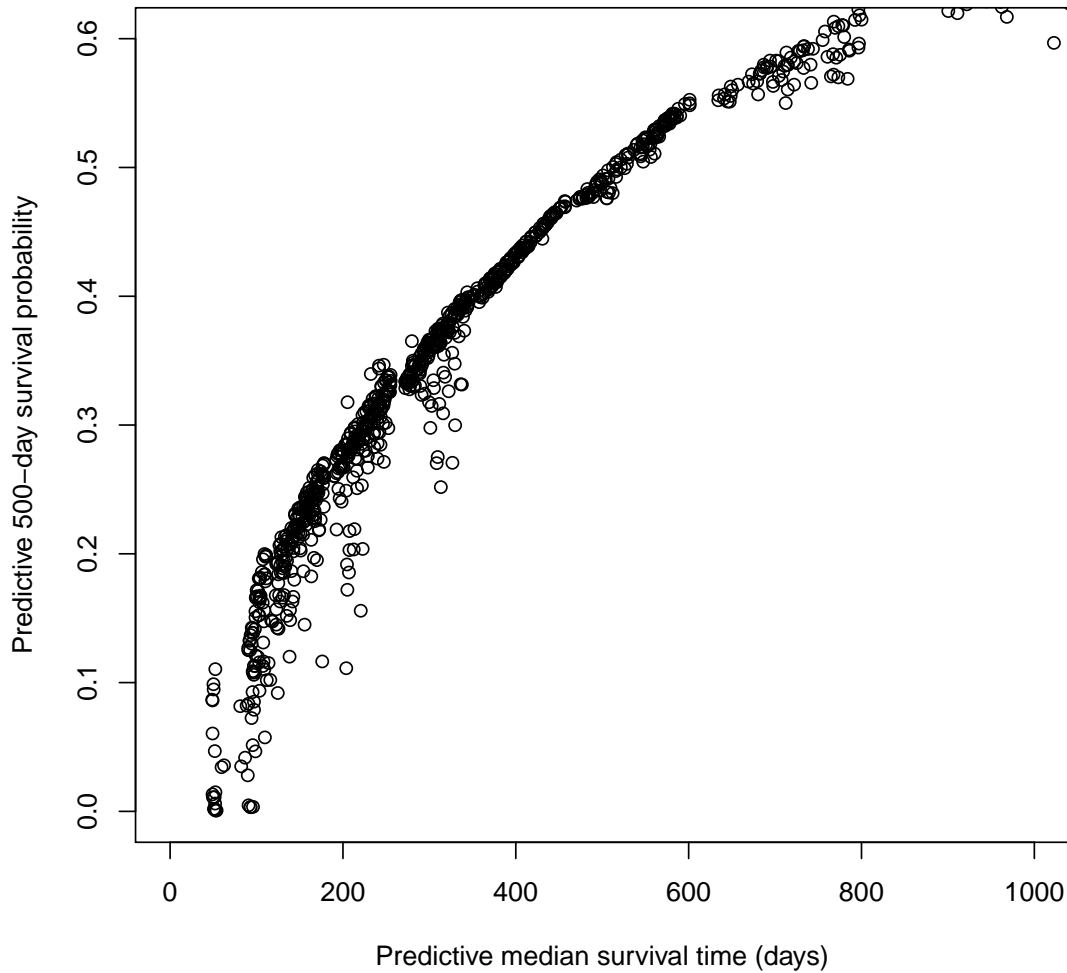


Figure 5.8: Plot of the predictive median survival time against the 500 days survival probability using a piecewise constant hazard model for the leukaemia data set on a restricted axis

model.

5.11 Flexible models for survival data

This research is aimed at creating flexibility in the form of the dependence of the hazard on covariates. Approaches to creating flexibility in the form of dependence of the hazard on covariates are linked with to the following

- the form of the baseline hazard

- non-proportionality of hazards and
- the relationship between covariate values and the hazard i.e. relaxing the assumption that

$$\eta_i = \log \lambda_i = \beta_0 + \sum_{s=1}^S \beta_s x_{i,s}. \quad (5.17)$$

Flexible modelling in survival avoids proportional hazards assumption of linearity of covariate effects. Flexible survival models avoid the restrictive parametric assumptions of proportionality of hazards or log-linearity of the relationships among covariates. We introduce flexibility to the covariate effects by using a generalised piecewise constant hazard model in which we categorise the quantitative covariates and make them ordinal and thus we divide the covariate space into cells. We will have two forms of prior distributions. One will be based on a parametric model and the other uses a Gaussian Markov random field. The second way of introducing flexibility to the covariate effects is by using the continuous parameter space model and supposing a model where the logarithms of the hazards are given a Gaussian process prior. It is important to think about priors when introducing flexibility to covariate effects as we will have to compensate for lack of structure by having a prior which makes neighbouring covariates profiles correlated in their prior. This will allow “borrowing of strength” (Morris & Normand, 1992).

Chapter 6

Generalised piecewise constant hazard model

6.1 Introduction

In this chapter, we introduce flexibility into the form of the dependence of the hazard function on the covariate effects by introducing a new model, the generalised piecewise constant hazard (GPCH) model in which quantitative covariates are categorised or made ordinal. The generalised piecewise constant hazard model will be discussed using two forms of prior distribution. One of the prior distributions will be based on a parametric model and the other uses a Gaussian Markov random field.

We recall that the proportional hazard model in Equation (3.2) restricts the hazard function to be linear in the covariates. In recent times, some methods like splines, basis function regression, Bayesian classification and regression trees (C&RT) have been used for evaluating the dependence of the survival time on the covariates. Classification and regression trees allow for recursive partitioning which reveals the structure of the data by examining the effect of the covariates on the survival distribution and whether the effect is true for all individuals. A simple prediction model is fitted for each partition. The main advantage of C&RT over other methods is the ability to analyse complex non-linear data sets with many variables by reducing the dimensionality of the data. See Denison *et al.* (2002) for further details on classification and regression trees.

The spline function is another approach to modelling the dependence of the survival time on the covariates. Spline functions can be used to investigate non-linear effects in covariates. The proportional hazard assumption is violated as the coefficient of the covariate can be represented by smooth non-linear functions and Equation (3.1) which is given as

$$h_i(t) = h_0(t) \exp \left\{ \sum_{s=1}^S \beta_s x_{i,s} \right\} \quad (6.1)$$

is replaced by

$$h_i(t) = h_0(t) \exp \{g(x)\} \quad (6.2)$$

where g is the log-hazard ratio function which is an unspecified smooth function of the covariates which can be approximated by a spline function. Molinari *et al.* (2001) developed a survival model in which the hazard ratios are not restricted to being linear in the covariates by using spline functions and the hazard ratio was estimated with piecewise linear splines. See Royston & Parmar (2002) for further details on spline functions.

Section 6.2 presents the generalised piecewise constant hazard model (GPCH) as a new approach to modelling where we categorise the covariates. We introduce a time variable and treat it like another ordinal covariate. We combine the time variable and the covariate profile as a cell. The new approach involves constructing the likelihood contribution from individuals. We introduced and presented two new alternatives to constructing the priors based on the parametric model. In the first alternative, the covariance matrix between the time periods tends to a zero matrix as the number of time steps tend to infinity while the second alternative has a non-zero limit.

We present a new approach of using a Gaussian Markov random field prior in the GPCH model using a distance based measure where we constructed a correlation matrix. We review the first order autoregressive (AR(1)) process as a Gaussian Markov random field (GMRF) in one dimension and illustrate this in a 2-dimensional case where the autoregressive parameters were isotropic. We give the provision for inclusion of frailty in the GPCH using GMRF priors. We will illustrate sampling the log-hazards using the two forms of priors. We present a new model for the GPCH model where we split the log-hazards of an individual into two parts. We will think of this approach as a way of improving the performance of sampling relative to updating each parameter independently. One part will carry the dependence between log-hazards and the other part is the logarithm of the frailty of the individual which does not carry dependence since every individual's frailty differs. We will also suggest using the forward backward algorithm in sampling the log-hazards. We present a new approach to improving mixing. These approaches include sampling the principal components of the log-hazards and Gaussian approximation of the proposal distribution.

6.2 Generalised piecewise constant hazard model (GPCH model)

We recall the piecewise constant hazard model discussed in Section 3.5.1 as a model that is flexible in the baseline hazard. The generalised piecewise constant hazard (GPCH) model is one form of the piecewise constant hazard model in which we will relax the parametric assumption of the relationships among the covariates. In the GPCH model, we categorise the quantitative covariates or make them ordinal and thus, we divide the covariate space into cells within each of which the hazard is constant. In the standard piecewise constant hazard model, we divide the time variable into intervals in which the hazard is constant. In the generalised piecewise constant hazard model, we have a time variable which indicates the period in which the individual had the event or was censored and we incorporate time intervals just as in the case of the standard piecewise constant hazard model. We define a “covariate profile” as a specific list of values for the categorised covariates. The time is treated as another ordinal covariate except that a patient can appear in several time intervals. We define a “cell” as a combination of a covariate profile and the time interval.

Suppose we have S covariates in the model. Let the number of levels of covariate s be p_s . Then, each covariate profile is associated with a unique node in a S -dimensional array in the covariate space containing $P = \prod_{s=1}^S p_s$ nodes. Combining this with J time intervals, we have a $S + 1$ dimensional array with $J \times P = c$ cells. A constant hazard is associated with each cell.

We define the collection of all log-hazards as

$$\underline{\eta} = (\underline{\eta}_1^T, \underline{\eta}_2^T, \dots, \underline{\eta}_P^T)^T$$

where $\underline{\eta}_p = (\eta_{p1}, \dots, \eta_{pJ})^T$ and η_{pj} is the log-hazard for covariate profile p in time interval j for $p = 1, \dots, P$ and $j = 1, \dots, J$. The vector $\underline{\eta}$ is ordered so that the time intervals are nested within the covariate profiles. We could also have a rearranged form of $\underline{\eta}$ so that we have covariate profiles nested within time interval as

$$\tilde{\underline{\eta}} = (\tilde{\underline{\eta}}_1^T, \dots, \tilde{\underline{\eta}}_J^T)^T$$

where $\tilde{\underline{\eta}}_j = (\eta_{j,1}, \dots, \eta_{j,P})^T$. We have that

$$\underline{\eta} = H\tilde{\underline{\eta}} \tag{6.3}$$

where H is a permutation matrix.

We think of the likelihood contribution of the individuals just like in the case of the piecewise constant hazard model discussed in Section 3.3.1. We denote the likelihood

contribution in the j^{th} time interval for the k^{th} individual with profile p as $L_{j,p,k}$. The overall likelihood is

$$L = \prod_{j=1}^J \prod_{p=1}^P \prod_{k=1}^{n_p} L_{j,p,k} = \prod_{j=1}^J \prod_{p=1}^P L_{j,p} \quad (6.4)$$

where the number of individuals with profile p is n_p .

The individuals are labelled $i = 1, \dots, n$, the covariate profiles are labelled $p = 1, \dots, P$, the time intervals as $j = 1, \dots, J$ and the individuals in the subset with covariate profile p are labelled $k = 1, \dots, n_p$. The number of individuals with profile p who die will be denoted by $n_{d,p}$. The number of individuals with profile p who die in time interval j will be denoted by $n_{d,j,p}$.

We define

$$\delta_{j,p,k} = \begin{cases} 1 & \text{if the } k^{th} \text{ individual in profile } p \text{ dies in interval } j \\ 0 & \text{otherwise} \end{cases}$$

Now,

$$\begin{aligned} L_{j,p} &= \left\{ \prod_{k=1}^{n_p} \lambda_{j,p}^{\delta_{j,p,k}} \right\} \exp \left\{ -\lambda_{j,p} \sum_{k=1}^{n_p} t_{j,p,k}^* \right\} \\ &= \lambda_{j,p}^{n_{d,j,p}} \exp \left\{ -\lambda_{j,p} \sum_{k=1}^{n_p} t_{j,p,k}^* \right\} \\ &= \lambda_{j,p}^{n_{d,j,p}} \exp \left\{ -\lambda_{j,p} T_{j,p}^* \right\} \end{aligned}$$

where $\lambda_{j,p} = \exp \{ \eta_{j,p} \}$, $T_{j,p}^* = \sum_{k=1}^{n_p} t_{j,p,k}^*$ and

$$t_{j,p,k}^* = \begin{cases} 0 & \text{if } t_{p,k} < \tau_{j-1} \\ t_{p,k} - \tau_{j-1} & \text{if } \tau_{j-1} < t_{p,k} \leq \tau_j \\ \tau_j - \tau_{j-1} & \text{if } \tau_j < t_{p,k} \end{cases}$$

The overall likelihood contribution can then be written as

$$\begin{aligned} L &= \prod_{j=1}^J \prod_{p=1}^P \lambda_{j,p}^{n_{d,j,p}} \exp \left\{ -\lambda_{j,p} T_{j,p}^* \right\} \\ &= \left\{ \prod_{j=1}^J \prod_{p=1}^P \lambda_{j,p}^{n_{d,j,p}} \right\} \left\{ \exp \left[-\sum_{j=1}^J \sum_{p=1}^P \lambda_{j,p} T_{j,p}^* \right] \right\} \quad (6.5) \end{aligned}$$

and the logarithm of the likelihood is

$$\ell = \log L = \sum_{j=1}^J \sum_{p=1}^P n_{d,j,p} \eta_{j,p} - \sum_{j=1}^J \sum_{p=1}^P \lambda_{j,p} T_{j,p}^* \quad (6.6)$$

We introduce flexibility to covariate effects using the generalised piecewise constant hazard model using the prior based on a parametric model and the Gaussian Markov random field prior. We will discuss the structures of these priors.

6.3 GPCH model with priors based on parametric model

In Section 5.7.3, we used a prior for the piecewise constant hazard model in which log-hazards in neighbouring time intervals were positively correlated. In the GPCH model, we also wish to have prior correlation between log-hazards for neighbouring covariate profiles. We will discuss the generalised piecewise constant hazard model using priors which are based on a parametric model. This structure of prior for the generalised piecewise constant hazard model will allow for the main and interaction effects. We will expect that neighbouring time intervals are correlated in their prior distribution.

Suppose we have S covariates for $s = 1, \dots, S$ with a vector of covariate effects $\underline{\beta} = (\beta_1, \dots, \beta_S)^T$. If we use just main effects and the P-way interaction, the linear predictor of the p^{th} covariate profile in the j^{th} time interval, $\eta_{p,j}$ with covariate vector $x_{p,1}, \dots, x_{p,S}$ is given by

$$\eta_{p,j} = \beta_{0,j} + \beta_{1,j}x_{p,1} + \dots + \beta_{S,j}x_{p,S} + \gamma_{x_{p,1}, \dots, x_{p,S}, j}$$

where $\beta_{0,j}$ is the baseline hazard in the j^{th} interval, $\gamma_{x_{p,1}, \dots, x_{p,S}, j}$ is the highest order interaction effect in the j^{th} interval, $\beta_{s,j}$ is the s^{th} covariate main effect in the j^{th} time interval. We could also include 2-way, 3-way etc. interactions.

We refer to the discussion in Section 5.6 on the construction of the design matrix.

6.3.1 Construction of prior distribution based on parametric model

We recall that the usual linear model for the logarithm of the hazards given in Equation 3.2 is

$$\eta_j = \beta_0 + \sum_{s=1}^S \beta_s x_{j,s} \quad (6.7)$$

where $x_{j,s}$ is the value of covariate s in covariate vector \underline{x}_j and hence, we have S covariates.

Let $\underline{\eta}^* = (\eta_1, \dots, \eta_P)^T$ be the log-hazards in the first time interval for the P covariate profiles. Let $X = x_1, \dots, x_{S'}$ be the *design matrix* of the data. We recall that the vector of

the parameters of the linear model $\underline{\beta}$ will also have S' parameter (see Section 5.6). Then

$$\underline{\eta}^* = X\underline{\beta}. \quad (6.8)$$

If the expectation of $\underline{\beta}$ is $E(\underline{\beta})$ then the expectation of $\underline{\eta}^*$ is

$$\underline{\mu}^* = XE(\underline{\beta}) \quad (6.9)$$

Let the covariance matrix of $\underline{\beta}$ be $V_{\underline{\beta}}$. Then, the covariance matrix of $\underline{\eta}^*$ is

$$\text{Var}(\underline{\eta}^*) = XV_{\underline{\beta}}X^T \quad (6.10)$$

We suppose that the prior mean and covariance matrix of the first time interval are given as $\underline{\mu}^* = E(\underline{\eta}^*)$ and $\text{Var}(\underline{\eta}^*)$ respectively. We now need to consider the dependence between parameters in different time intervals. One possible way of constructing the joint prior distribution of the c parameters of the model might be to give it either a stationary first order autoregressive process prior or a moving average process prior. In this thesis, we will assume a first order autoregressive model $AR(1)$ and so, we choose an autoregressive parameter ρ which gives different strength of relationship to the parameters. If the autoregressive parameter is positive then the collections of parameter that are closer to each other are more strongly correlated. Hence, the autoregressive parameter governs the degree of prior correlation between neighbouring time periods. We refer to Chatfield (2004) for further details on AR models. The autocovariance at lag j for $j = 1, \dots, J$ is given by $\rho^j \text{Var}(\underline{\eta}^*)$. Hence, the nested covariance matrix of the log-hazards with covariate profiles nested within the time intervals is given as

$$\text{Var}(\underline{\tilde{\eta}}) = \begin{pmatrix} V_0 & V_1 & V_2 & V_3 & \cdots \\ V_1 & V_0 & V_1 & V_2 & \cdots \\ V_2 & V_1 & V_0 & V_1 & \cdots \\ V_3 & V_2 & V_1 & V_0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ V_{J-1} & V_{J-2} & \cdots & V_1 & V_0 \end{pmatrix}$$

where V_0 is $\text{Var}(\underline{\eta}^*) + \text{Var}(\gamma)$ and $\text{Var}(\gamma)$ is a $P \times P$ diagonal matrix which is the covariance matrix of the interactions effects of the covariate profiles and V_k is $\rho^k \text{Var}(\underline{\eta}^*)$ for $k > 0$. The addition of $\text{Var}(\gamma)$ to $\text{Var}(\underline{\eta}^*)$ makes V_0 non-singular.

We might want to rearrange $\text{Var}(\underline{\tilde{\eta}})$ so that we have a covariance matrix for the vector of η in which the time intervals are nested within the covariate profiles. Hence, we have a re-ordered matrix

$$\text{Var}(\underline{\eta}) = H\text{Var}(\underline{\tilde{\eta}})H^T \quad (6.11)$$

Again, H is a permutation matrix.

The means of the model parameters might or might not depend on the time interval. We choose to make the means of the parameters not depend on the time interval so that they are the same at every time interval and the prior is stationary. Again, we might rearrange the means so that we have the time intervals within the covariate profiles. We get $\underline{\mu}$ from $\underline{\mu}^*$. The joint prior distribution of the parameters of the model now has a prior mean of $\underline{\mu}$ and covariance matrix $\text{Var}(\underline{\eta})$.

An alternative representation of the GPCH model using priors based on a parametric model is

$$\eta_{p,j} = \beta_0 + \beta_1 x_{p,1} + \dots + \beta_S x_{p,S} + \beta_{\text{time},j} + \gamma_{x_{p,1}, \dots, x_{p,S}, j}$$

where $\beta_{\text{time},j}$ is the time effect at the j^{th} time interval. In this model, we assume that the covariance matrix of the linear model without including the interaction effect is

$$\text{Var}(\underline{\eta}^*) = \text{Var}(\beta_0 + \beta_1 x_{p,1} + \dots + \beta_S x_{p,S})$$

where we have S covariates and P possible covariate profiles and $\text{Var}(\underline{\eta}^*)$ is a $P \times P$ covariance matrix. The variance of β_{time} , V_t is a scalar. Again, we have a nested covariance matrix of the covariate profiles within the time intervals where V_0 is $\text{Var}(\underline{\eta}^*) + V_t A + \text{Var}(\gamma)$ and V_k is $\text{Var}(\underline{\eta}^*) + \rho^k V_t A$ for $k > 0$.

Here, $\text{Var}(\gamma)$ is a diagonal matrix which is the variance of the interaction and A is a matrix with all entries as 1. Again, we might want to rearrange $\text{Var}(\underline{\eta})$ so that we have a covariance matrix for $\underline{\eta}$ where the time intervals are nested within the covariate profiles and hence, we do the transformation. We also rearrange the means so that we have the time intervals within the covariate profiles. Again, the joint prior distribution of the parameters of the model now has a prior mean and covariance matrix $\underline{\mu}$ and $\text{Var}(\underline{\eta})$ respectively.

6.3.2 MCMC sampling of the log-hazards (η) using priors based on parametric model

We recall the logarithm of the likelihood of the GPCH model given in Equation 6.6. A simple MCMC algorithm is the Gibbs sampler which involves sampling the log-hazards one at a time. The conditional prior distribution of each log-hazard given the others will be a normal distribution using the basic properties of the normal distribution (Rue & Held, 2005).

The conditional prior density for the p^{th} log-hazard representing a covariate profile at

a time interval, η_p , is then given by

$$\pi(\eta_p) \propto \exp \left\{ -\frac{1}{2} \frac{(\eta_p - \mu_{p|p'})^2}{V_{p|p'}} \right\}$$

where $\mu_{p|p'}$ and $V_{p|p'}$ are the conditional mean and variance of the log-hazard.

The full conditional distribution of the p^{th} log-hazard η_p is

$$\begin{aligned} \pi(\eta_p|D) &\propto \text{prior} \times \text{likelihood} \\ &= (\text{constant})\pi(\eta_p)L_{j,p} \end{aligned}$$

We propose a new value of η_p , η_p^* from a normal distribution. We compute the proposal density of η_p^* given η_p , $q(\eta_p^*|\eta_p)$ and the proposal density of η_p given η_p^* , $q(\eta_p|\eta_p^*)$. From the Metropolis-Hastings algorithm, the proposed log-hazard η_p^* is accepted with probability

$$A = \min \left\{ 1, \frac{\pi(\eta_p^*|D) q(\eta_p|\eta_p^*)}{\pi(\eta_p|D) q(\eta_p^*|\eta_p)} \right\}.$$

If $q(\eta_p|\eta_p^*) = q(\eta_p^*|\eta_p)$, that is, if we use a symmetric proposal such as a random walk proposal, the acceptance probability reduces to

$$A = \min \left\{ 1, \frac{\pi(\eta_p^*|D)}{\pi(\eta_p|D)} \right\}$$

since the proposal densities cancel out.

We could do block sampling using some other method: for example in the case of sampling where we approximate a Gaussian proposal distribution for the log-hazards (see Section 8.3.2).

6.4 Gaussian process prior

We define a Gaussian process as a set of random variables $Y(t)$ where every finite collection of random variables $Y(t_1), \dots, Y(t_k)$ has a joint Gaussian distribution. Just like a Gaussian distribution, a Gaussian process is specified by a mean function and a covariance function. The covariance function defines the properties and covariance between random variables in a space. As usual, the covariance function is symmetric and positive semi-definite.

We specify the covariance matrix of a Gaussian processes prior using a distance measure. Suppose that we compute prior means and marginal prior variances for the elements of $\underline{\eta}$ using a linear model as in 6.3.1 then

$$\underline{\eta} = X\underline{\beta}$$

where X is the *design matrix* and $\underline{\beta}$ is the S' linear model parameters.

The expectation of $\underline{\eta}$ is

$$B^* = E(\underline{\eta}) = XE(\underline{\beta}) \quad (6.12)$$

The implied covariance matrix of $\underline{\eta}$ would be

$$V_{\underline{\eta}}^* = XV_{\underline{\beta}}X^T \quad (6.13)$$

However, we do not use the implied covariance matrix since it will be singular. We construct a correlation matrix R based on “distances” between the c covariate vectors. The diagonal elements of $V_{\underline{\eta}}^*$ will be the marginal variances. We let the diagonal of $V_{\underline{\eta}}^*$ be $\underline{v}^* = (v_1^*, \dots, v_c^*)^T$ and let $\underline{s}^* = (s_1^*, \dots, s_c^*)^T$ be the vector of standard deviations where $s_j^* = \sqrt{v_j^*}$. Let $S^* = \text{diag}(\underline{s}^*)$, a diagonal matrix with diagonal \underline{s}^* . Then

$$\text{Var}(\underline{\eta}) = S^*RS^* \quad (6.14)$$

We will want to construct the correlation matrix R .

If we assume that there are no frailties, the correlation between a patient i in cell (p, j) i.e. profile p and time period j and another patient i' in cell (p', j') i.e. profile p' and time period j' $r_{(i,p,j),(i',p',j')}$ is given by

$$r_{(i,p,j),(i',p',j')} = a + (1 - a) \exp \{ -d_{(p,j),(p',j')} \} \quad \text{for } i \neq i' \quad (6.15)$$

We could have a suitable distance measure between cell (p, j) and (p', j') , $d_{(p,j),(p',j')}$ as

$$d_{(p,j),(p',j')} = \sum_{s=1}^S K_s |\underline{x}_{p,s} - \underline{x}_{p',s}| \quad (6.16)$$

Another possibility of the distance measure between cell (p, j) and (p', j') is

$$d_{(p,j),(p',j')} = \sum_{s=1}^S K_s (\underline{x}_{p,s} - \underline{x}_{p',s})^2 \quad (6.17)$$

where $\underline{x}_{p,s}$ and $\underline{x}_{p',s}$ are the covariate vectors for covariate profile p and p' . K_s is a vector of autoregressive parameters or decay rate parameters for each covariate which weights the distance. We choose K_s as part of the prior specification. If K_s is small then we have a strong correlation between neighbouring cells. The advantage of (6.17) is that it can be

written as a matrix given by

$$d_{(p,j),(p',j')} = (\underline{x}_{p,s} - \underline{x}_{p',s})^T K (\underline{x}_{p,s} - \underline{x}_{p',s}) \quad (6.18)$$

where K is a diagonal matrix with elements K_1, \dots, K_S .

The distance $d_{(p,j),(p',j')}$ between the log-hazards of two patients with similar covariate vectors in the same time interval will be zero and $r_{(i,p,j),(i',p,j)} = 1$. However, we might have that the correlation between a patient i in cell (p, j) i.e. profile p and time period j and another patient i' in cell (p', j') , $r_{(i,p,j),(i',p',j')} \neq 1$ even if the covariates profiles are exactly the same because of frailty (see Section 6.6.2).

The value of a represents the correlation that remains when the distance becomes very big. The value of the constant a should always be positive so that we do not have $r_{(i,p,j),(i',p',j')} \rightarrow 0$ as $d_{(p,j),(p',j')} \rightarrow \infty$. If $a = 0$, $r_{(i,p,j),(i',p',j')}$ depends on only $\exp\{-d_{(p,j),(p',j')}\}$. The bigger the value of “ a ”, the more similar the value of the correlation to “ a ”. Hence, as $d_{(p,j),(p',j')} \rightarrow \infty$, $r_{(i,p,j),(i',p',j')} \rightarrow a$. Larger values for a makes the correlation depend less on distance.

6.5 Gaussian Markov random field priors

A special case of a Gaussian process is a Gaussian Markov random field. Instead of the process being defined on continuous space, it is defined on a graph so that nodes have neighbours. A Gaussian Markov random field (GMRF) is a finite-dimensional random vector that follows a Gaussian distribution which satisfies the Markov conditional independence assumptions. The simplest example of a GMRF is the autoregressive process of order 1 (AR(1) process) which is in 1 dimension. We illustrate the AR(1) process following Rue & Held (2005) by supposing that $\underline{\eta} = (\eta_1, \dots, \eta_P)^T$ is a random vector expressed as

$$\eta_p = \rho\eta_{p-1} + \epsilon_p \text{ for } \epsilon_p \stackrel{iid}{\sim} N(0, 1), \quad |\rho| < 1$$

The joint density of $\underline{\eta}$ is

$$\frac{1}{(2\pi)^{n/2}} |Q|^{1/2} \exp \left\{ -\frac{1}{2} \underline{\eta}^T Q \underline{\eta} \right\}$$

where the precision matrix Q is the tridiagonal matrix with zero entries outside the first off diagonal elements and is given by

$$Q = \begin{pmatrix} 1 & -\rho & 0 & 0 & \cdots \\ -\rho & 1 + \rho^2 & -\rho & 0 & \cdots \\ 0 & -\rho & 1 + \rho^2 & -\rho & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & & -\rho & 1 + \rho^2 & -\rho \\ 0 & 0 & \cdots & -\rho & 1 \end{pmatrix}$$

The precision matrix is formed due to the fact that η_i and $\eta_{i'}$ are conditionally independent for $|i - i'| > 1$, given the rest. Hence for a GMRF, if the elements of Q are $Q_{ii'}$ then $Q_{ii'} = 0$ unless $i = i'$ or i is a neighbour of i' and thus it is dependent on a distance measure on the graph and requires the Markov property. In one dimension, the entries of the covariance matrix $V = Q^{-1}$ are given by

$$\sigma_{ii'} = \frac{1}{1 - \rho^2} \rho^{|i - i'|}$$

and the correlation is

$$r_{ii'} = \rho^{|i - i'|} \text{ for } -1 < \rho < 1$$

In this research, we look at Markov random fields on a rectangular lattice and we define neighbours as nodes which are separated by exactly one step in one dimension. We extend the GMRF to a vector of $S + 1$ dimensions denoted by $\underline{\eta} = (\eta_1^T, \dots, \eta_P^T)^T$ leading to a GMRF of size PJ .

We illustrate a 2 dimensional case. We assume a stationary GMRF and define the equation

$$\eta_{i,j} = \rho\eta_{i,j-1} + \rho\eta_{i-1,j} + \epsilon_{i,j} \quad (6.19)$$

following Figure 6.1 where ϵ is the random change with mean 0 and variance σ^2 . In this illustration, we make the autoregressive parameter ρ equal (isotropic). Just like a generalised autoregressive model, we have the marginal variance of the parameters V , the one-step autocovariance between parameters C and the diagonal autocovariance as D .

We want to find expressions for V , C and D . We find an expression for V by multiplying both sides of Equation (6.19) by $\eta_{i,j}$ and taking expectations. We get

$$V = \rho C + \rho C + \sigma^2 = 2\rho C + \sigma^2 \quad (6.20)$$

We find an expression for C by multiplying both sides of Equation (6.19) by $\eta_{i,j-1}$ and taking expectations. We get

$$C = \rho V + \rho D \quad (6.21)$$

We find an expression for D by multiplying both sides of Equation (6.19) by $\eta_{i-1,j-1}$ and

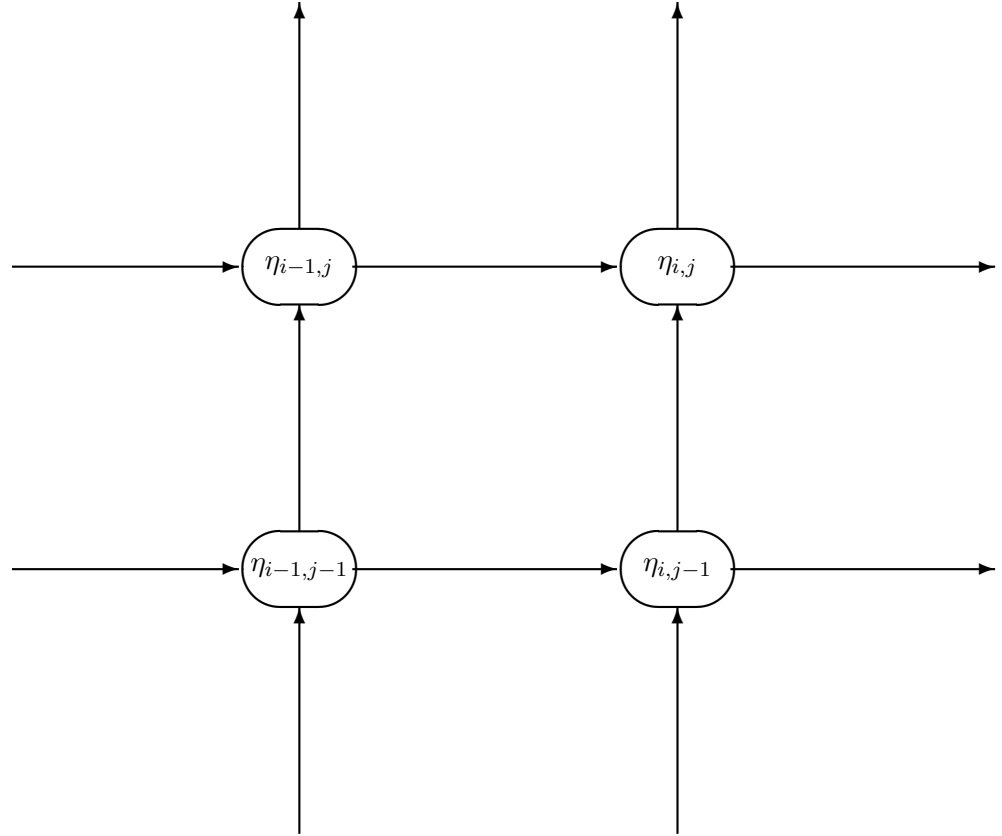


Figure 6.1: Directed acyclic graph to illustrate a 2-dimensional case of GMRF

taking expectations. We get

$$D = 2\rho C \quad (6.22)$$

Then, we solve Equations (6.20), (6.21) and (6.22) to get values for V , C and D . If we substitute D into (6.21), we get

$$C = \frac{\rho V}{1 - 2\rho^2} \quad (6.23)$$

and if we substitute (6.23) into (6.20) we get

$$V = \frac{\sigma^2(1 - 2\rho^2)}{1 - 4\rho^2}. \quad (6.24)$$

We substitute (6.24) into (6.23) and we get

$$C = \frac{\rho\sigma^2}{1 - 4\rho^2}. \quad (6.25)$$

Finally, we substitute (6.24) and (6.25) into (6.22) and we get

$$D = \frac{2\rho^2\sigma^2}{1 - 4\rho^2}$$

It is obvious that we require that for stationarity condition, $1 - 4\rho^2 > 0$ and so, $\rho^2 < 1/4$.

We might also be interested in the proportion of the variance of $\eta_{i,j}$ explained by $\eta_{i-1,j}$ and $\eta_{i,j-1}$ and it is given as

$$R^2 = 1 - \frac{\sigma^2}{V} = \frac{2\rho^2}{1 - 2\rho^2}$$

since the biggest value of $\rho^2 = 1/4$, $R^2 = 1$ and when $\rho^2 = 0$, $R^2 = 0$.

6.6 GPCH model with Gaussian Markov random field prior

We discuss the generalised piecewise constant hazard model using the Gaussian Markov random field prior in which the covariate profiles are allowed to depend on their neighbouring covariate profiles. We recall that for each cell we have a corresponding log-hazard of individuals who have that particular covariate profile. Thus, each covariate profile associated with a log-hazard corresponds to a cell in a multidimensional array. If we use a Markov random field prior, the logarithm of the hazard of a particular covariate profile will be conditionally independent of all other log-hazards given its neighbouring log-hazards. A good reason for categorising the covariates could be that when using the Gibbs sampling, we can sample a particular log-hazard (cell) by conditioning just on the neighbours. Thus, we exploit the Markov property.

6.6.1 Construction of Gaussian Markov random field prior

We recall the prior means and prior covariance matrix of $\underline{\eta}$ in Equations (6.12) and (6.13). We treat the time intervals as covariates and hence $\underline{\eta} = (\eta_1, \dots, \eta_c)^T$, and we make X be the *design matrix* with c rows and S' columns. We construct the correlation matrix R and the covariance matrix following Section 6.4.

6.6.2 Frailty

If we include frailty in the model then, the correlation between a patient i in cell (p, j) i.e. profile p and time period j and another patient i' in cell (p', j') i.e. profile p' and time period j' , $r_{(i,p,j),(i',p',j')}$ is given by

$$r_{(i,p,j),(i',p',j')} = \Lambda_{(i,j),(i',j')} \{a + (1 - a) \exp[-d_{(p,j),(p',j')}]\} \quad \text{for } i \neq i' \quad (6.26)$$

The factor Λ , where $0 \leq \Lambda \leq 1$ gives a provision for the inclusion of frailty.

We consider two different frailty models. In one of them, each patient has one random frailty in all time intervals and the factor Λ is given by

$$\Lambda_{(i,j),(i',j')} = \begin{cases} = 1 & \text{if } i = i' \\ < 1 & \text{otherwise} \end{cases}$$

The other case is if we have different frailty value in each time interval and the factor Λ is given by

$$\Lambda_{(i,j),(i',j')} = \begin{cases} = 1 & \text{if } i = i' \text{ and } j = j' \\ < 1 & \text{otherwise} \end{cases}$$

In this second case, the frailties are effectively random patient-time period interaction effects.

6.6.3 MCMC sampling of the log-hazards (η) using Gaussian Markov random field prior

We can sample for example using Metropolis within Gibbs algorithm which involves sampling the log-hazards one at a time. The conditional prior distribution of each log-hazard when we have a Gaussian Markov random field prior is the conditional distribution of that log-hazard given its neighbouring log-hazards. We use the basic properties of the normal distribution in Rue & Held (2005). For instance, from standard results for multivariate normal distribution, the conditional distribution of the p^{th} log-hazard given the neighbouring log-hazards $\eta_p | \eta_{p'}$ has a normal distribution given by

$$\eta_p | \eta_{p'} \sim N(M_{p|p'}, V_{p|p'})$$

The conditional mean $M_{p|p'}$ is given as

$$M_{p|p'} = \mu_p + C_{pp'} V_{p'}^{-1} (\underline{\eta}_{p'} - \underline{\mu}_{p'})$$

and the conditional covariance matrix $V_{p|p'}$ is given as

$$V_{p|p'} = V_{pp} - C_{pp'} V_{p'}^{-1} C_{p'p}$$

where $\underline{\eta}_{p'}$ is a $p' \times 1$ column matrix of the neighbouring log-hazards and $\underline{\mu}_{p'}$ is the vector of means of the neighbouring log-hazards. We get $C_{pp'}$ and $V_{p'}$ from partitioning the variance covariance matrix V given below as

$$\left[\begin{array}{c|c} V_{pp} & C_{pp'} \\ \hline C_{p'p} & V_{p'}^{-1} \end{array} \right]$$

and $C'_{p'p}$ is the transpose of $C_{pp'}$.

The prior density for the p^{th} log-hazard η_p given its neighbours $\eta_{p'}$ is then given by

$$\pi(\eta_p) \propto \exp \left\{ -\frac{1}{2} \frac{(\eta_p - \mu_{p|p'})^2}{V_{p|p'}} \right\}$$

The logarithm of the conditional prior density of the p^{th} log-hazard is

$$\log [\pi(\eta_p)] = (\text{constant}) - \frac{1}{2} \frac{(\eta_p - \mu_{p|p'})^2}{V_{p|p'}}$$

The likelihood contribution from the p^{th} log-hazard is $L(\eta_p|D)$. Again, we use the Metropolis-Hastings algorithm since we can not sample directly from the full conditional distribution.

The full conditional distribution of the p^{th} log-hazard η_p is given by

$$\begin{aligned} \pi(\eta_p|D) &\propto \text{prior} \times \text{likelihood} \\ &= (\text{constant})\pi(\eta_p)L(\eta_p|D) \end{aligned}$$

We propose a new value of η_p , η_p^* from a normal distribution with mean η_p . We compute the proposal density of η_p^* given η_p , $q(\eta_p^*|\eta_p)$ and the proposal density of η_p given η_p^* , $q(\eta_p|\eta_p^*)$. From the Metropolis-Hastings algorithm, the proposed log-hazard η_p^* is accepted with probability

$$A = \min \left\{ 1, \frac{\pi(\eta_p^*|D) q(\eta_p|\eta_p^*)}{\pi(\eta_p|D) q(\eta_p^*|\eta_p)} \right\}$$

The acceptance probability reduces to

$$A = \min \left\{ 1, \frac{\pi(\eta_p^*|D)}{\pi(\eta_p|D)} \right\}$$

since the proposal densities cancel out if a symmetric proposal is used.

The full conditional distribution of each log-hazards only depends on the neighbours. Experience has shown that mixing will be poor since there is strong correlation in the posterior between the log-hazards at neighbouring points in the parameter space. We suggest numerical approaches in Section 4.7 for improving the poor mixing in the GPCH model. We note that we could decide to split the log-hazards into two parts for us to use the forward backward algorithm to sample the log-hazards.

6.7 Splitting the log-hazards

When the parameters of the posterior distribution are correlated, joint sampling of these parameters can improve the performance of the sampling relative to updating each parameter independently (Roberts & Sahu, 1997). We could aim at moving parameters together (*i.e blocking*) in order to improve mixing.

We suppose that the joint prior density of the log-hazards $\underline{\eta}$ is

$$\pi(\underline{\eta}) \propto \exp \left\{ -\frac{1}{2}(\underline{\eta} - \underline{\mu})^T Q(\underline{\eta} - \underline{\mu}) \right\}$$

where the prior mean of $\underline{\eta}$ is $\underline{\mu}$ and the prior precision matrix is Q . We recall that the overall logarithm of the likelihood contribution from all covariate profiles is given in (6.5) and the full conditional density of the log-hazards is

$$\pi(\underline{\eta}|D) \propto \pi(\underline{\eta})L(\underline{\eta}|D) \quad (6.27)$$

The difficulty of sampling directly from this full conditional distribution of the log-hazards is that it is not a multivariate normal distribution because the likelihood is not normal. We might consider splitting the log-hazard of an individual into two parts. One part will carry the dependence between log-hazards and the other part is the logarithm of the frailty of the individual which does not carry dependence since every individual's frailty differs. Hence, the i^{th} log-hazard η_i is given by,

$$\eta_i = \theta_{p(i)} + Z_i \quad (6.28)$$

where $\theta_{p(i)}$ is the log-hazard without frailty and carries the dependence, $p(i)$ is the profile to which the individual i belongs and Z_i is the logarithm of the frailty of the individual. The log-hazard η_i of individual i is conditionally independent of the log-hazard η_j of individual j given the log-hazard without frailty θ_i (see Figure 6.2). The reason for splitting the log-hazards is so that we separate the parts that do not have a normal distribution and they become conditionally independent and we can sample them one at a time. Then, the parts that are correlated have a multivariate normal full conditional distribution and it is easier to sample them all at once (blocking). Again, we recall from Section 6.6.2 that we could consider two different frailty models. In one model, each patient has one random frailty in all time intervals and the other case we have a different frailty value in each time interval.

We suppose that the vector of log-hazards without frailty $\underline{\theta}$ has a multivariate normal prior distribution with mean vector \underline{M} and covariance matrix C given by

$$\underline{\theta} \sim N(\underline{M}, C).$$

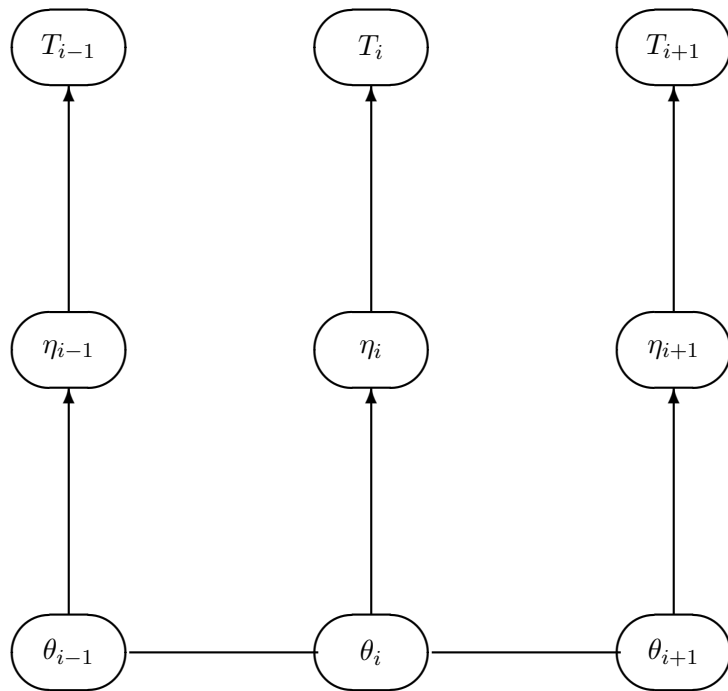


Figure 6.2: Directed acyclic graph to explain the idea of splitting the log-hazards

The vector of log-frailty \underline{Z} also has a multivariate normal prior distribution with mean $\underline{0}$ and covariance matrix $\sigma_f^2 I$ given by,

$$\underline{Z} \sim N(\underline{0}, \sigma_f^2 I).$$

where I is an identity matrix and σ_f^2 is the variance of the log-frailties.

Then, the vector of log-hazards $\underline{\eta}$ will have a multivariate normal prior distribution with mean vector \underline{M} and covariance matrix $C + \sigma_f^2 I$ given by

$$\underline{\eta} \sim N(\underline{M}, C + \sigma_f^2 I).$$

The log-hazard of each individual is independent of the log-hazard of another individual given the log-hazard without frailty. We sample the log-hazard of the i^{th} individual, η_i given both the log-hazard excluding the frailty and the individual's survival time, $\eta_i | \theta_i, T_i$.

We have to sample the precision of the log-frailty, the log-hazards of the individuals η_1, \dots, η_n and the vector of log-hazards without frailties $\underline{\theta}$. We sample η_1, \dots, η_n one at a time using the Metropolis within Gibbs algorithm (discussed in Section 4.6.5) since they are conditionally independent of each other given the log-hazards without frailty $\underline{\theta}$. We note that $\theta_{p(i)}$ is specific to the covariate profile to which the individual belongs.

Firstly, we might think of sampling the precision of log-frailties τ_f while we fix $\underline{\eta}$ and $\underline{\theta}$. We recall that from 6.28, the vector of log-frailties Z is

$$\underline{Z} = \underline{\eta} - \underline{\theta} \tag{6.29}$$

We suppose that the precision of the log-frailties τ_f has a gamma prior distribution with parameters a_f and b_f , $\tau_f \sim \text{Gamma}(a_f, b_f)$ given by

$$\pi(\tau_f) \propto \tau_f^{a_f-1} \exp\{-b_f \tau_f\}$$

The likelihood from the precision of the log-frailties is given by

$$L(\tau_f) \propto \tau_f^{n/2} \exp\left\{-\frac{\tau_f}{2} \sum_{i=1}^n Z_i^2\right\}$$

The full conditional distribution of the precision of the log-frailties is

$$\pi(\tau_f | a_f, b_f) \propto \tau_f^{a_f+n/2-1} \exp\left\{-\tau_f\left(b_f + \frac{1}{2} \sum_{i=1}^n Z_i^2\right)\right\} \tag{6.30}$$

We sample τ_f directly using Equation 6.30.

We will sample the log-hazards including frailties η 's one at a time while we fix τ_f

and $\underline{\theta}$ using Metropolis-Hastings steps. The conditional variance of $\underline{\eta}|\underline{\theta}$, $\text{Var}(\underline{\eta}|\underline{\theta})$ is the variance of the log-frailty τ_f^{-1} and the conditional mean is $\underline{\theta}$. The prior distribution of the i^{th} log-hazard, η_i is then given by

$$\pi(\eta_i) \propto \exp \left\{ -\frac{\tau_f}{2} (\eta_i - \theta_i)^2 \right\}$$

We recall that the likelihood contribution from the i^{th} individual is

$$L(\eta_i|D) = \{\exp[\eta_i]\}^{\delta_i} \exp\{-\exp[\eta_i]t\}$$

where

$$\delta_i = \begin{cases} 1 & \text{if the individual died} \\ 0 & \text{if censored} \end{cases}$$

We propose a new value of η_i , η_i^* from a normal distribution. The proposed value of the log-hazard is accepted with probability is given as

$$A = \min \left\{ 1, \frac{\pi(\eta_i^*|\tau_f, D)}{\pi(\eta_i|\tau_f, D)} \right\}$$

When we have sampled values for τ_f and $\underline{\eta}$, $\underline{\theta}$ now has multivariate normal full conditional distribution. We could sample $\underline{\theta}$ from the multivariate normal distribution either using the Cholesky decomposition or the forward backward algorithm discussed in Section 4.7. This will only work if we use the patient-time period interactions rather than the standard frailties.

6.8 Sampling the principal components of the log-hazards

Instead of sampling the correlated log-hazards one at a time, we can transform to a new set of parameters which are not correlated in the prior. One way to do this is to use the principal components. We could think of sampling the principal components of the log-hazards when we have poor mixing in survival modelling. The reason for sampling the principal components of the log-hazards is that the principal components given the prior covariance matrix are linear functions of the log-hazards. In addition, the principal components are independent and uncorrelated in the prior which could improve mixing.

Finding the principal components will involve the eigen values and eigen vectors (Jolliffe, 2002). Suppose \mathcal{Y} is a matrix where the columns are eigen vectors of $\text{Var}(\underline{\eta})$. Let vector of principal component \underline{P} be given by

$$\underline{P} = \mathcal{Y}\underline{\eta}$$

The covariance matrix of the principal components $\text{Var}(\underline{P})$ is given by

$$\text{Var}(\underline{P}) = \Upsilon \text{Var}(\underline{\eta}) \Upsilon^T \quad (6.31)$$

The covariance matrix of the principal components $\text{Var}(\underline{P})$ is a diagonal matrix of the eigen values and hence, the principal components are independent. The vector of means of the principal components $\text{E}(\underline{P})$ of the log-hazards is given by

$$\text{E}(\underline{P}) = \Upsilon \text{E}(\underline{\eta}) \quad (6.32)$$

The prior distribution of the p^{th} principal component is a normal distribution with mean $\text{E}(P_p)$, variance $\text{Var}(P_p)$ and density

$$\pi(P_p) \propto \exp \left\{ -\frac{1}{2} \frac{(P_p - \text{E}(P_p))^2}{\text{Var}(P_p)} \right\}$$

The likelihood contribution from the observations is $L(\underline{\eta}|D)$ which is the overall likelihood contribution from all covariate profiles where

$$\underline{\eta} = \Upsilon^{-1} \underline{P}. \quad (6.33)$$

In order to get the likelihood of the principal component, we convert the principal components \underline{P} back to the log-hazards ($\underline{\eta}$) using the relationship between $\underline{\eta}$ and \underline{P} .

Again, we could use the Metropolis within Gibbs algorithm since we can not sample directly from the full conditional distribution. We sample each of the principal components P_p from its full conditional distribution. The full conditional distribution of the p^{th} principal components P_p is then given by

$$\pi(P_p|D) = (\text{constant})\pi(P_p)L(\underline{\eta}|D)$$

We propose a new value of P_p, P_p^* from a normal distribution. Since the principal components are a function of the log-hazards, the proposed values of the log-hazards are computed again using (6.33). So,

$$\underline{\eta}^* = \Upsilon^{-1} \underline{P}^*$$

The proposal densities cancel out if we use a symmetric proposal (i.e. if the variance of the normal proposal distribution is constant). From the Metropolis-Hastings algorithm, the proposed log-hazard η_p^* is accepted with probability

$$A = \min \left\{ 1, \frac{\pi(P_p^*|D)}{\pi(P_p|D)} \right\}$$

If the proposed principal component of the p^{th} log-hazard is accepted then the accepted log-hazard value will be transformed from the proposed principal component value. Otherwise it stays the same.

6.9 Some comments on the computations using the GPCH model

In the generalised piecewise constant hazard model, the number of covariate profiles will depend on the number of covariates and categories. In general, if we have P covariate profiles and J time intervals, then we will have JP cells. For example, the number of covariates in the SNLG data set discussed in Section 2.2.4 was 14 and the cutpoints of the categories corresponding to the lower quartile, median and upper quartile of the continuous covariates. The number of covariate profiles is $2^8 \times 4^4 \times 5 \times 3 \approx 1$ million. If time is partitioned into 10 time intervals for example, the number of cells for the SNLG data set is ≈ 10 million. This will involve fitting a model with so many parameters. Clearly, 10 million is much bigger than the number of patients in the data set. We will have too many cells but only 636 patients and some of these may share covariate profiles so that the number of distinct covariate profiles among observed patients may be smaller. In fact many (almost all) of the approximately 1 million possible covariate profiles are not observed. The observed covariate profiles form only a tiny proportion of the possible total. Most of the cells will not have patients in the neighbouring cells because nearly all of the cells have no data. The GPCH model is very practicable in the case of the leukaemia data where we have only four covariates, 128 covariate profiles and 10 time periods and hence, a total of 1280 cells. We have 1043 patients. The log-hazard for a cell in the multidimensional array depends on those in the neighbouring cells and hence the Markov property could be used. This is unrealistic in the case of the SNLG data because of too many combinations of covariate. It would not be feasible to do MCMC with log-hazards for covariates cell if there were a large number of covariates.

We could consider the alternative of only updating the log-hazard in cells where there were observed patients since most of the neighbours will be missing (not being used). However, by doing this, we would lose the computational advantage given by the Markov property.

An alternative is to use a modified prior correlation structure. The correlation structure will be defined to have the property that the full conditional distribution for a log-hazard depends only on the log-hazards for covariate profiles within a distance. This has a link to the K nearest neighbours (KNN) method (Prijs *et al.*, 2007).

Another alternative which is linked to the problem of large numbers of covariate profiles in the generalised piecewise constant hazard model can be tackled by abandoning the

piecewise constant hazard feature with respect to the covariates (not making the covariates ordinal) and using a continuous parameter space model with a Gaussian process prior for the logarithm of the hazards at the observed patient-points in covariate space. There could be as many hazard values as the number of patients. In other cases, the number of hazard values are less than the number of patients because two individuals may have the same covariate profile.

6.10 Summary

In this chapter, we have discussed introducing flexibility into the form of the dependence of the hazard function on the covariate effects using the generalised piecewise constant hazard model. We discussed it using the prior distribution based on both a parametric model and a Markov random field. The two prior distributions have different structures. We suggested some numerical methods to improving mixing in cases where we have poor mixing. These include sampling the principal components of the log-hazards because the principal components are independent in their prior, Gaussian approximation of the proposal distribution (see Section 4.7.2) and splitting the log-hazards and sampling the part that carries the dependence using either the forward backward algorithm (4.7.1) or the Cholesky decomposition (4.6.6). This method can be illustrated using a data set with a smaller number of covariates since the number of covariate profiles is dependent on the number of covariates.

Chapter 7

Continuous parameter space model

7.1 Introduction

In this chapter, we discuss a new approach to flexible modelling in survival by relaxing the parametric model for the covariate effects using the continuous parameter space model where we leave the continuous covariates as they are. We introduce flexibility into the model by giving the logarithm of the hazards a Gaussian process prior.

The form of the baseline hazard could possibly be a piecewise constant hazard model or we could use a parametric survival model such as the Weibull distribution. For illustration in this chapter, we will use a Weibull distribution lifetime. Each observed covariate profile thus has a separate logarithm of the hazard and these are given a correlation based on a distance measure in covariate space in the prior. We construct the prior distribution of the parameters of the continuous parameter space model by using the prior information to set up a systematic way of obtaining the means and variances of the log-hazards for patients with different covariates. The continuous parameter space model could be an alternative to the GPCH model discussed in Chapter 6.

7.2 Basic ideas and notation

Let us suppose that we are fitting a parametric survival model, such as a Weibull distribution and each patient has one hazard (or hazard multiplier) λ_i . If we have n patients, labelled $i = 1, \dots, n$ then each patient has a hazard which depends on the patient's covariate vector \underline{X}_i . In the case of the generalised piecewise constant hazard model where we make the quantitative covariates categorical, we expect many patients to share the same covariate vector with others (see Section 6.2). It could be that some patients have

exactly the same covariate values even when we do not make the covariates categorical. In practice, it is possible that two patients share the same covariate vector because covariates are only recorded to a small number of significant digits. For example, age is in whole years. Therefore, to allow for this possibility, let us say that the number of *distinct* covariate vectors is n' where $n' \leq n$. Hence, we have n' different hazard (or hazard multiplier) values $\lambda_1, \dots, \lambda_{n'}$. We will let $\eta_p = \log(\lambda_p)$.

Suppose that we have a linear model for $\eta_1, \dots, \eta_{n'}$ with $S + 1$ coefficients, β_0, \dots, β_S . Usually $S + 1 < n'$ and we will assume that this is the case. The coefficients corresponding to columns of the design matrix which could be quantitative covariates or contrasts for the categorical covariates. We include β_0 as an intercept. Let $\underline{\eta} = (\eta_1, \dots, \eta_{n'})^T$. Let X be the *design matrix* with n' rows and $S + 1$ columns where row j is $(1, x_{j,1}, x_{j,2}, \dots, x_{j,S})$. We will note that we could also have categorical covariates (see discussions on Section 6.3.1 on the suggested ways to include these covariate in the design matrix). Let $\underline{\beta} = (\beta_0, \beta_1, \dots, \beta_S)^T$. Then, in a parametric model for the covariate effects

$$\underline{\eta} = X\underline{\beta}.$$

Note that, if $S + 1 > n'$ then the parameters β_0, \dots, β_S are not identifiable. We will deal with the case where $S + 1 = n'$ when we introduce the Gaussian process prior in Section 7.2.1.

Now, the fact that $S + 1 < n'$ imposes linear constraints on the values of $\eta_1, \dots, \eta_{n'}$ since there are only $S + 1$ *degrees of freedom* available. This means that the values of certain linear functions of $\eta_1, \dots, \eta_{n'}$ are fixed by the model. It is these constraints which we want to relax in this project.

We refer to (6.9) and (6.10) for the expectation and covariance matrix of $\underline{\eta}$ respectively. Because of the linear constraints, the $n' \times n'$ matrix $\text{Var}(\eta^*)$ can not have rank greater than $S + 1$ and since $S + 1 < n'$, it is singular.

7.2.1 Unconstrained η (Gaussian process prior)

In order not to have the constraints implied by a parametric model as in (6.7) we increase the number of parameters to n' . There is then exactly one parameter for each covariate vector. The parameters are $\eta_1, \dots, \eta_{n'}$ themselves. This is different from the linear model shown in (6.7) since we increase the number of the degrees of freedom until there is one degree of freedom for every logarithm of the hazard.

Now, we need a suitable prior distribution for $\underline{\eta}$.

7.3 Bayesian modelling of the log-hazards (η) using Gaussian process prior and Weibull lifetime distribution

We suppose that the lifetime random variable has a Weibull distribution with parameters (λ, α) . Suppose that we have n' different covariate profiles. Let the number of cases with covariate profile p be n_p , for $p = 1, \dots, n'$, and let the number of these cases where the event time is observed be $n_{d,p}$. Let the number of all cases where the event is observed be

$$n_d = \sum_{p=1}^{n'} n_{d,p} .$$

Let the hazard multiplier for profile p be

$$\lambda_p = \exp \{ \eta_p \} .$$

The likelihood contribution from profile p is

$$L(\lambda_p, \alpha | D) = \left\{ \prod_{k=1}^{n_p} \left[\alpha \lambda_p t_{p,k}^{\alpha-1} \right]^{d_{p,k}} \right\} \exp \left\{ -\lambda_p \sum_{k=1}^{n_p} t_{p,k}^\alpha \right\} \quad (7.1)$$

where the event or censoring time for patient k in profile p is $t_{p,k}$ and $d_{p,k} = 1$ if this is an event but $d_{p,k} = 0$ if it is a censoring time.

The logarithm of the likelihood contribution from profile p is therefore

$$\ell_p = n_{d,p} \log \alpha + n_{d,p} \eta_p + (\alpha - 1) \sum_{k=1}^{n_p} d_{p,k} \log t_{p,k} - \lambda_p \sum_{k=1}^{n_p} t_{p,k}^\alpha \quad (7.2)$$

and the overall logarithm of the likelihood is

$$\ell = n_d \log \alpha + \sum_{p=1}^{n'} n_{d,p} \eta_p + (\alpha - 1) \sum_{p=1}^{n'} \sum_{k=1}^{n_p} d_{p,k} \log t_{p,k} - \sum_{p=1}^{n'} \left\{ \lambda_p \sum_{k=1}^{n_p} t_{p,k}^\alpha \right\} . \quad (7.3)$$

We will suppose that from Section 7.2.1, the vector of log-hazards $\underline{\eta}$ has a multivariate normal prior distribution with vector of means $E(\underline{\eta})$ and covariance matrix which we have to construct.

7.3.1 Constructing prior distribution for the parameters of the continuous parameter space model

We could make the prior means of all of $\eta_1, \dots, \eta_{n'}$ equal. In this way, we are not using prior information about which hazards are likely to be greater or smaller. Sometimes this might be preferred. We could also make use of prior information by setting up a systematic way of obtaining the means and variances of the log-hazards for patients with different covariates. We set up a preliminary linear model and give prior means and variances to all the coefficients of the covariates. Hence, the prior information depends on the effects of the covariates on the log-hazard of the patients. In an attempt to make the prior mean and variances of the log-hazard different, we can express the log-hazard for each covariate profile as a linear function of these coefficient of the covariates and it helps us to form the prior information. We follow the way that the expectation of $\underline{\eta}$ was constructed in Section 6.6.1.

We will want a suitable covariance matrix for $\underline{\eta}$. We can not use $V_{\underline{\eta}}^*$ (see 6.10) because of the implied constraints. This is because $V_{\underline{\eta}}^*$ is singular and the number of parameters in the linear model is less than the number of log-hazards of the patients ($S + 1 < n'$). We recall from Section 6.6 that the covariance matrix of the vector of log-hazards is

$$\text{Var}(\underline{\eta}) = S^* R S^* \quad (7.4)$$

and the elements of the correlation matrix R are

$$r_{i,j} = \Lambda \{a + (1 - a) \exp \{-d_{i,j}\}\} \quad \text{for } i \neq j. \quad (7.5)$$

One possibility could be to form a correlation matrix based on the “distances” between the covariate vectors. We could combine this with the marginal variances which are given by the diagonal elements of $V_{\underline{\eta}}^*$. In this way, the variances of $\eta_1, \dots, \eta_{n'}$ reflect a reasonable assessment of our prior uncertainties.

We refer to (6.14) for the construction of the correlation matrix. In this case, $d_{i,j}$ is the distance measure between the i^{th} and j^{th} covariate profiles following Schmidt & Rodriguez (2011) and is given by

$$d_{i,j} = \sqrt{(\underline{x}_i - \underline{x}_j)^T D^* (\underline{x}_i - \underline{x}_j)} \quad (7.6)$$

where \underline{x}_i and \underline{x}_j are the covariate vectors for the i^{th} and j^{th} covariate profiles. The symmetric positive definite matrix D^* rescales the covariates, for example to give them the same standard deviation, since some of the covariates might be in different scales. If we assume that D^* is diagonal then the diagonal elements define the scales of the covariates so that one unit of all covariates counts equally. The corresponding diagonal element of D^* for each variable is the inverse of the square of the standard deviation of the variable.

Again, we think of values for Λ and a as in Section 6.6.

7.3.2 Sampling from the posterior distribution of the log-hazards

One possibility for MCMC sampling is to sample each of the log-hazards one at a time. The conditional prior distribution of each log-hazard given the others will be a normal distribution. For instance, from standard results for multivariate normal distribution, the conditional distribution of the p^{th} log-hazard η_p given the other remaining log-hazards $\eta_p|\underline{\eta}_{p'}$, has a normal distribution with mean $M_{p|p'}$ and variance $V_{p|p'}$. So,

$$\eta_p|\eta_{p'} \sim N(M_{p|p'}, V_{p|p'}) .$$

The conditional mean $M_{p|p'}$ is

$$M_{p|p'} = \mu_p + C_{pp'} V_{p'}^{-1} (\underline{\eta}_{p'} - \underline{\mu}_{p'})$$

and the conditional covariance matrix $V_{p|p'}$ is

$$V_{p|p'} = V_{pp} - C_{pp'} V_{p'}^{-1} C_{p'p}$$

where $\underline{\eta}_{p'}$ is a $(n' - 1) \times 1$ column matrix without η_p ,

$\underline{\mu}_{p'}$ is the vector of means of log-hazards without the mean of η_p .

We get $C_{pp'}$ and $V_{p'}$ from partitioning the variance covariance matrix V as

$$\left[\begin{array}{c|c} V_{pp} & C_{pp'} \\ \hline C_{p'p} & V_{p'}^{-1} \end{array} \right]$$

where $C_{p'p}$ is the transpose of $C_{pp'}$. The other conditional distributions of the other log-hazards can be computed in a similar way. The conditional prior density for the p^{th} log-hazard η_p is then given by

$$\pi(\eta_p) \propto \exp \left\{ -\frac{1}{2} \frac{(\eta_p - M_{p|p'})^2}{V_{p|p'}} \right\}$$

where $M_{p|p'}$ and $V_{p|p'}$ are the conditional mean and variance of the p^{th} log-hazard.

The logarithm of the prior density of the p^{th} log-hazard is

$$\log [\pi(\eta_p)] = (\text{constant}) - \frac{1}{2} \frac{(\eta_p - M_{p|p'})^2}{V_{p|p'}}$$

The posterior density will not have a closed or analytical form but we can simulate from it using MCMC techniques. We will use the Metropolis within Gibbs algorithm since we can not sample directly from the full conditional distribution. In this case, we fix the Weibull shape parameter α and then we sample each of the log-hazard from its full conditional distribution. The full conditional density of the p^{th} log-hazard is

$$\pi(\eta_p, \alpha | D) = (\text{constant})\pi(\eta_p)L(\eta_p, \alpha | D)$$

We propose a new value of η_p , η_p^* from a normal distribution. We compute the proposal density of η_p^* given η_p , $q(\eta_p^*|\eta_p)$ and the proposal density of η_p given η_p^* , $q(\eta_p|\eta_p^*)$. From the Metropolis-Hastings algorithm, the proposed log-hazard η_p^* is accepted with probability

$$A = \min \left\{ 1, \frac{\pi(\eta_p^*, \alpha | D) q(\eta_p | \eta_p^*)}{\pi(\eta_p, \alpha | D) q(\eta_p^* | \eta_p)} \right\}$$

which reduces to

$$A = \min \left\{ 1, \frac{\pi(\eta_p^*, \alpha | D)}{\pi(\eta_p, \alpha | D)} \right\}$$

and the proposal densities cancel out if they are symmetric. We will now fix the accepted values of all the sampled log-hazards $\underline{\eta}$ and sample the Weibull shape parameter α . Using a gamma prior for α , the prior density for α is

$$\begin{aligned} \pi(\alpha | a, b) &= \frac{(b)^a}{\Gamma(a)} \alpha^{a-1} \exp \{-b\alpha\} \\ &\propto \alpha^{a-1} \exp \{-b\alpha\} . \end{aligned}$$

The posterior density of α is

$$\pi(\alpha | \underline{\eta}, D) = \kappa \pi(\alpha | a, b) L(\alpha | \underline{\eta}, D) .$$

So, we propose a new value $\alpha^* \sim Ga(a^*, b^*)$ for some specified a^* , b^* . We refer to Section 5.5.2 for details on this.

The proposed value of α is accepted with probability

$$A = \min \left\{ 1, \frac{\pi(\alpha^* | \underline{\eta}, D) q(\alpha | \alpha^*)}{\pi(\alpha | \underline{\eta}, D) q(\alpha^* | \alpha)} \right\}$$

Again, experience has shown that mixing might be poor as there is strong correlation in the posterior between the log-hazards at neighbouring points in the parameter space. We will suggest the numerical approaches discussed in Section 4.7 and others which have already been discussed in Chapter 6 for improving the poor mixing. There are no additional numerical approaches to improving poor mixing in the continuous parameter space model.

7.4 Summary

In this chapter, we followed our discussion on the comments from Chapter 6, Section 6.9 to consider only updating the log-hazards of the observed patient-points using the continuous parameter space model with a Gaussian process prior. It is possible that the form of the baseline hazard should be a piecewise constant hazard. We illustrated this method by supposing a Weibull lifetime distribution. We will suggest some numerical methods to improving mixing in cases where we have poor mixing. These include sampling the principal components of the log-hazards, Gaussian approximation of the full conditional distribution as a proposal distribution (see Section 4.7.2) and splitting the log-hazards and sampling the part that carries the dependence and sampling using the Cholesky decomposition (4.6.6). It was impossible to split the log-hazards and sample using the forward backward algorithm since the log-hazards in the continuous parameter space model do not have the Markov property.

Chapter 8

Applications to Example Data sets and use of alternative computational methods

8.1 Introduction

The investigation of flexible Bayesian modelling is intended to relax the standard proportional hazard model. In this chapter, we will apply the methodology discussed in Chapters 6 and 7 to our example data sets discussed in Chapter 2.

We will begin by considering the leukaemia data set using the GPCH model with priors based on a parametric model and a Markov random field as discussed in Chapter 6. The GPCH model was not used on the SNLG data set for the reasons discussed in Section 6.9. In both cases, we discussed some alternative computational methods which might be useful for improving mixing. These numerical approaches include splitting the log-hazards and sampling using the forward backward algorithm (see Section 6.7 and Section 4.7.1) or Cholesky decomposition, Gaussian approximation of the fcd used as a proposal distribution (see Section 4.7.2) and sampling the principal components of the log-hazards discussed in Section 6.8. We then follow the first alternative approach to the comments on Section 6.9 where we consider only updating the combinations of the covariates (cells) where there were observed patients (continuous parameter space model) since most of the neighbours will be missing (not being used).

We analysed the SNLG data set using the continuous parameter space model discussed in Chapter 7. We also discussed some alternative computational methods to improve mixing in MCMC. These approaches used to improve mixing include sampling the principal components (see Section 6.8), splitting the log-hazards and sampling the log-hazards from a multivariate normal distribution using the Cholesky decomposition (see Section 6.7 and

Section 4.6.6) and Gaussian approximation of the fcd to form a proposal distribution of the parameters.

8.2 Application 1: GPCH model with priors based on parametric model

We revisit the leukaemia data set collected by North West Leukaemia Register in the United Kingdom in Section 2.3. The data set includes 1043 patients with 4 completely observed covariates (see Appendix A.3.2 for a partial display of data). The time variable was measured in units of days. The ordering of the covariates is given in Table 8.1.

Table 8.1: Ordering of covariates for the leukaemia data set for GPCH model

Covariates	variable
Age	x_1
Sex	x_2
Wbc	x_3
Depsc	x_4
Time variable	x_5

The covariates x_1 , x_3 and x_4 were categorised into four groups and we give the categories values 1, 2, 3 and 4. The cut points of the categories correspond to the lower quartile, median and upper quartile of the covariates. Covariate x_2 is binary. We refer to the cut points for the time variable used in Section 5.8 and we categorise the time variable x_5 based on these cut points. The number of covariate profiles is 128. Here, the linear predictor for the p^{th} covariate profile in the j^{th} time interval is defined as

$$\eta_{p,j} = \beta_{0,j} + \beta_{a',j}, x_{p,1} + \beta_{s',j}, x_{p,2} + \beta_{w',j}, x_{p,3} + \beta_{d',j}, x_{p,4} + \gamma_{x_{p,1}, x_{p,2}, x_{p,3}, x_{p,4}, j}$$

where $\beta_{a',j}$, $\beta_{s',j}$, $\beta_{w',j}$ and $\beta_{d',j}$ depend on the categorical value of “age”, “sex”, “Wbc” and “Depsc” respectively. Thus, the vector of covariate coefficients of the parametric model is $\underline{\beta} = (\beta_0, \beta_{a'}, \beta_{s'}, \beta_{w'}, \beta_{d'})^T$.

We include these categorical covariates in the design matrix X following the discussion in Section 6.3.1. There will be 11 parameters in the linear model, including an intercept β_0 . Hence, the design matrix X is a 128 by 11 matrix.

Construction of prior distributions

We construct the prior distributions of covariate effects following the discussion in Section 4.4.3.5 and 5.3.1.4. We constrain the parameters for the levels of the covariate effects to zero sum in order to avoid over-parameterisation since all covariates are categorical. We have $\delta_{a',1}$, $\delta_{a',2}$ and $\delta_{a',3}$ for the coefficient of age, $\delta_{s'}$ for the coefficient of sex, $\delta_{w',1}$, $\delta_{w',2}$ and $\delta_{w',3}$ for the coefficient of Wbc and $\delta_{d',1}$, $\delta_{d',2}$ and $\delta_{d',3}$ for the coefficient of Depsc.

We follow the discussions in Section 5.3.1.1 for the construction of the prior for the baseline parameter. We get prior distribution for β_0 will give us a mean of $\mu_0 = -6.0$ and variance $\sigma_0^2 = 0.12$. The means and variances of the parameters of the coefficients of the ordinal covariates are given in Table 8.2.

Table 8.2: The prior means and variances for the coefficients of the ordinal covariates of the leukaemia data set

Parameter	prior mean	prior variance
β_0 (baseline parameter)	-6	0.12
$\delta_{a',1}$	0	0.02
$\delta_{a',2}$	0	0.06
$\delta_{a',3}$	0	0.003
$\delta_{s'}$	0	0.0625
$\delta_{w',1}$	0	0.06
$\delta_{w',2}$	0	0.02
$\delta_{w',3}$	0	0.001
$\delta_{d',1}$	0	0.012
$\delta_{d',2}$	0	0.004
$\delta_{d',3}$	0	0.002

We will assume that the parameters of the linear model are not independent and so we construct the prior covariance matrix following the discussions in Section 4.4.1. The prior covariance matrix of the parameters of the linear model is given in Table 8.3.

We would want to construct the vector of prior means $\underline{\mu}^*$ and covariance matrix $\text{Var}(\underline{\eta}^*)$ of the log-hazards of the possible combination of covariates $\underline{\eta}^*$ given in Equation (6.9) and (6.10) respectively. In this case, we have 128 covariate profiles excluding the time variable. We make $\underline{\mu}^*$ and $\text{Var}(\underline{\eta}^*)$ be the prior mean and covariance matrix of the log-hazards in the first time interval. We will assume a stationary first order autoregressive process for the time periods. We choose an autoregressive parameter by thinking of how much the variance of the log-hazards in the next time period will reduce if we learnt about the log-hazards in one time period. We suppose that 64% of the variance is explained and we have the coefficient of determination as $r^2 = 0.64$. Hence, the correlation $r = 0.8$.

We follow the discussion in Section 6.3.1 and we construct the nested covariance matrix

Table 8.3: Covariance matrix of the parameters of the ordinal covariates of the leukaemia data set

β_0	$\delta_{a',1}$	$\delta_{a',2}$	$\delta_{a',3}$	$\delta_{s'}$	$\delta_{w',1}$	$\delta_{w',2}$	$\delta_{w',3}$	$\delta_{d',1}$	$\delta_{d',2}$	$\delta_{d',3}$
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0	1.0	0.9	0.9	0.0	0.7	0.7	0.7	0.7	0.7	0.7
0	0.9	1.0	0.9	0.0	0.7	0.7	0.7	0.7	0.7	0.7
0	0.9	0.9	1.0	0.0	0.7	0.7	0.7	0.7	0.7	0.7
0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
0	0.7	0.7	0.7	0.0	1.0	0.9	0.9	0.7	0.7	0.7
0	0.7	0.7	0.7	0.0	0.9	1.0	0.9	0.7	0.7	0.7
0	0.7	0.7	0.7	0.0	0.9	0.9	1.0	0.7	0.7	0.7
0	0.7	0.7	0.7	0.0	0.7	0.7	0.7	1.0	0.9	0.9
0	0.7	0.7	0.7	0.0	0.7	0.7	0.7	0.9	1.0	0.9
0	0.7	0.7	0.7	0.0	0.7	0.7	0.7	0.9	0.9	1.0

of the covariate profiles $\text{Var}(\tilde{\eta})$ which is a covariance matrix for the covariate profiles within the time intervals. We will also have a diagonal matrix $\text{Var}(\gamma)$ which is the covariance matrix of the interaction effects. For convenience in this thesis, we would want to rearrange the vector of means $\tilde{\eta}$ and the covariance matrix $\text{Var}(\tilde{\eta})$, to follow time intervals within the covariate profiles using Equation (6.3) and (6.11) and we stick to this style throughout the modelling. The vector of log-hazards $\underline{\eta}$ now has a prior distribution with mean $\underline{\mu}$ and covariance matrix $\text{Var}(\underline{\eta})$. The number of log-hazards corresponding to the cells will be 1280 since we have 10 time intervals. We have partially displayed the correlation between some covariate vectors in Table 8.4 and 8.5 and covariance matrices of some of the log-hazards in Table 8.6 and 8.7 respectively using the two alternatives discussed in Section 6.3.1.

Table 8.4: Partial display of the correlation between ordinal covariate vectors with priors based on parametric model

\underline{x}_1	\underline{x}_2	\underline{x}_3	\underline{x}_4	\underline{x}_5
1.000	0.724	0.579	0.463	0.370
0.724	1.000	0.724	0.579	0.463
0.579	0.724	1.000	0.724	0.579
0.463	0.579	0.724	1.000	0.724
0.370	0.463	0.579	0.724	1.000

Table 8.4 shows that the correlations between the pair of log-hazards are not very strong.

Table 8.5 shows very strong correlations between the pair of log-hazards.

Table 8.5: Partial display of the correlation between ordinal covariate vectors with (an alternative) prior based on parametric model

x_1	x_2	x_3	x_4	x_5
1.000	0.907	0.895	0.885	0.877
0.907	1.000	0.907	0.895	0.885
0.895	0.907	1.000	0.907	0.895
0.885	0.895	0.907	1.000	0.907
0.877	0.885	0.895	0.907	1.000

Table 8.6: Partial display of the prior covariance matrix of the log-hazards with priors based on parametric model

1.047	0.758	0.606	0.485	0.388
0.758	1.047	0.758	0.606	0.485
0.606	0.758	1.047	0.758	0.606
0.485	0.606	0.758	1.047	0.758
0.388	0.485	0.606	0.758	1.047

These two forms of the covariance matrices are not equivalent and so we expect that the posteriors will also be slightly different. We might want to consider what happens if we have a lot of time steps and the time lag becomes large. In the first form, the covariance matrix between two time periods tend to a zero matrix because ρ_k tends to zero as $k \rightarrow \infty$. In the second form, the limit is $\text{Var}(\underline{\eta}^*)$. We think that the second form may be more realistic because it suggests that we have non-zero prior covariances between the log-hazards at different time intervals, no matter how far apart the time intervals are.

Posterior distributions

The joint posterior distribution has been discussed in Chapter 6, Section 6.3.2 and we use both forms of the covariance matrices. The Metropolis within Gibbs algorithm can be applied here. The R functions to do this computation using the first form of covariance matrix can be found in Appendix A.5.1 and A.5.3. Following a burn-in of 20000 iterations, 60000 iterations were taken. Convergence was assessed by visual inspection of trace plots of the parameters and the mixing was quite good. The trace plots and density plots of some of the parameters (log-hazards) are shown in Figure 8.1. Visual inspection of the trace plots of the posterior values of the parameters revealed no convergence problems. The posterior means and standard deviations of some of the log-hazards are given in Table 8.8 (see basic Gibbs sampler I).

Figure 8.1 shows good mixing of the trace plots of the log-hazards of three covariate

Table 8.7: Partial display of the prior covariance matrix of the log-hazards with (an alternative) priors based on parametric model

	1.295	1.175	1.159	1.147	1.136
	1.175	1.295	1.175	1.159	1.147
	1.159	1.175	1.295	1.175	1.159
	1.147	1.159	1.175	1.295	1.175
	1.136	1.147	1.159	1.175	1.295

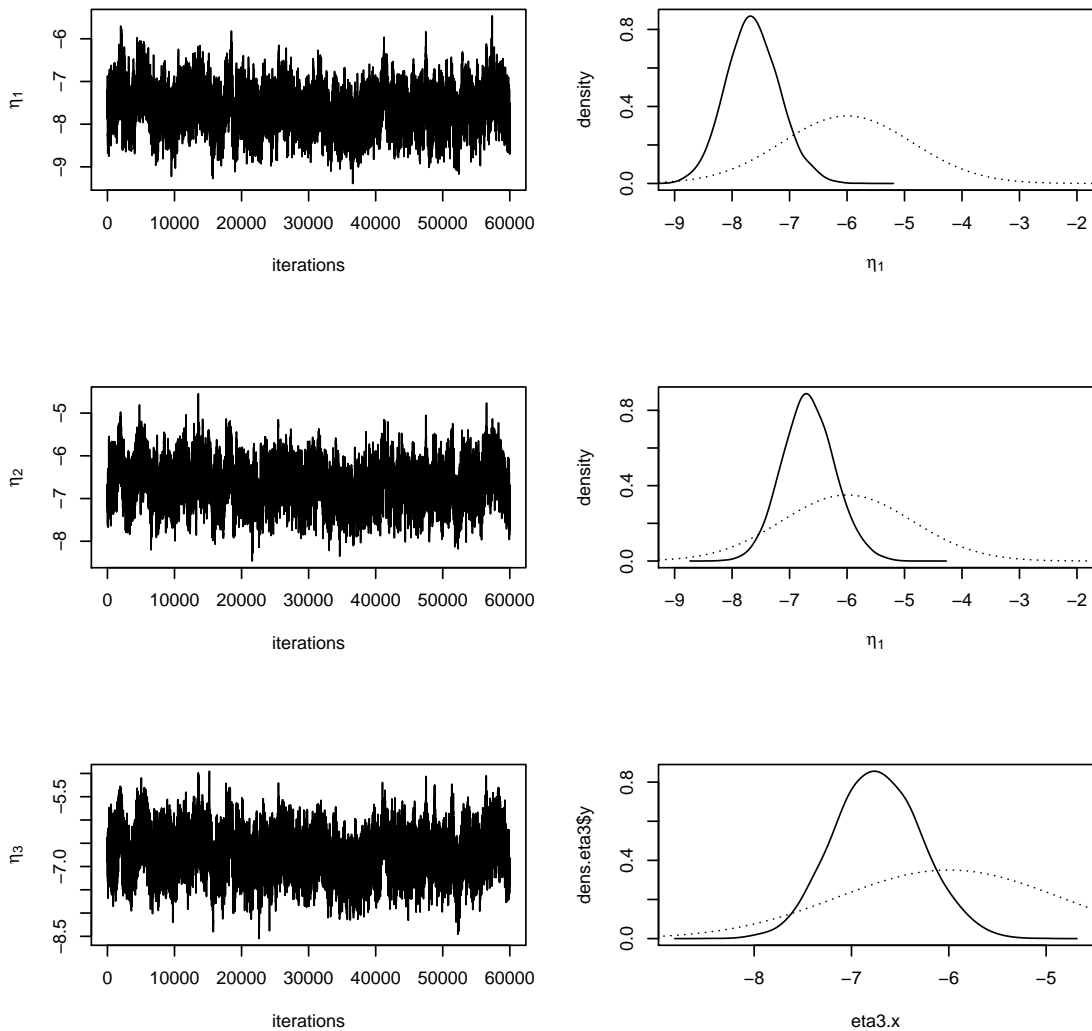


Figure 8.1: Trace plot and density plot (with dash lines representing the prior density and straight lines representing the posterior density) for three log-hazards using the GPCH model and priors based on the parametric model for the leukaemia data set

profiles. The density plots also show that we have learnt something from both data and prior.

Figure 8.2 shows timeplots of the posterior means and ± 2 standard deviation intervals for three log-hazards using the GPCH model and we shifted the time intervals to avoid overlapping. There was little difference between log-hazards among these three covariate profiles. The plot showed changes in the log-hazards within different time periods. Figure 8.3 shows three covariate profiles where the log-hazards are different. The effect of the covariates are obvious.

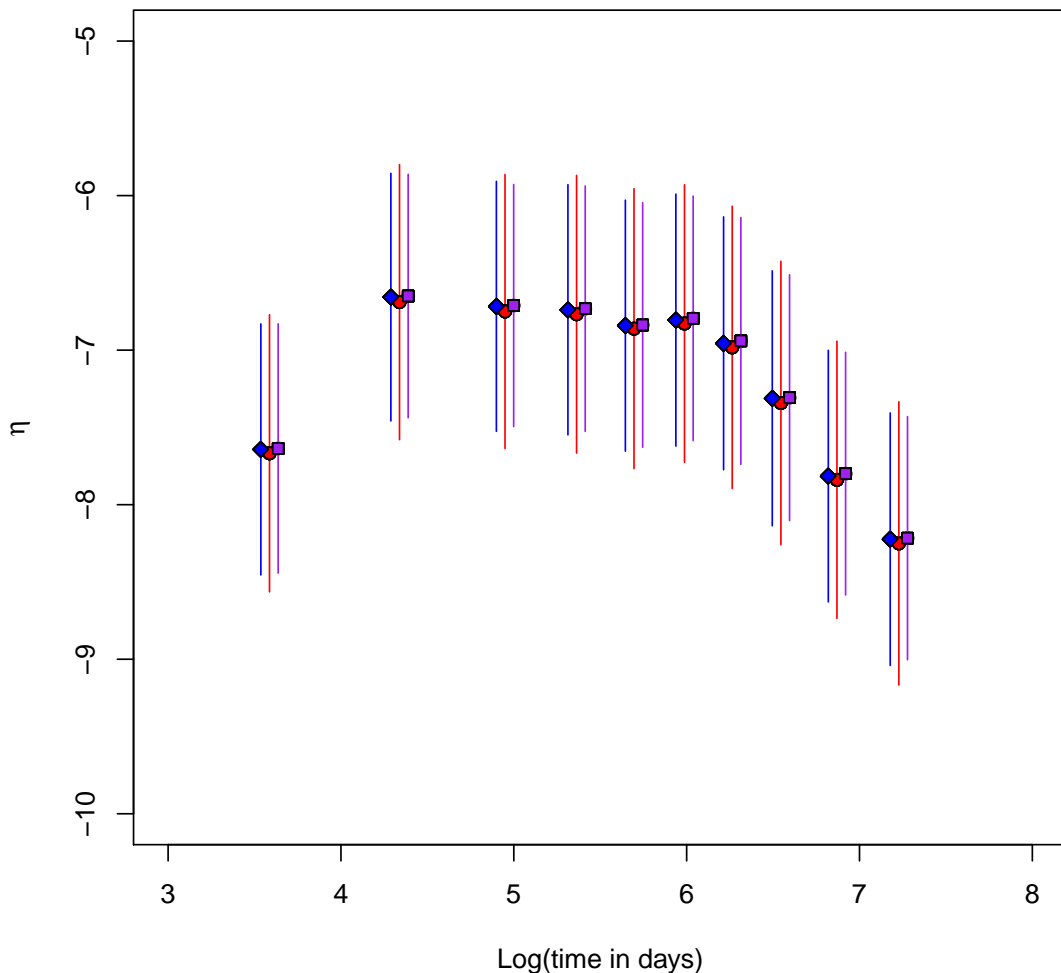


Figure 8.2: Time plots of the posterior means and ± 2 standard deviation intervals for three log-hazards using the GPCH model where priors are based on parametric model with overlapping log-hazards for the leukaemia data set

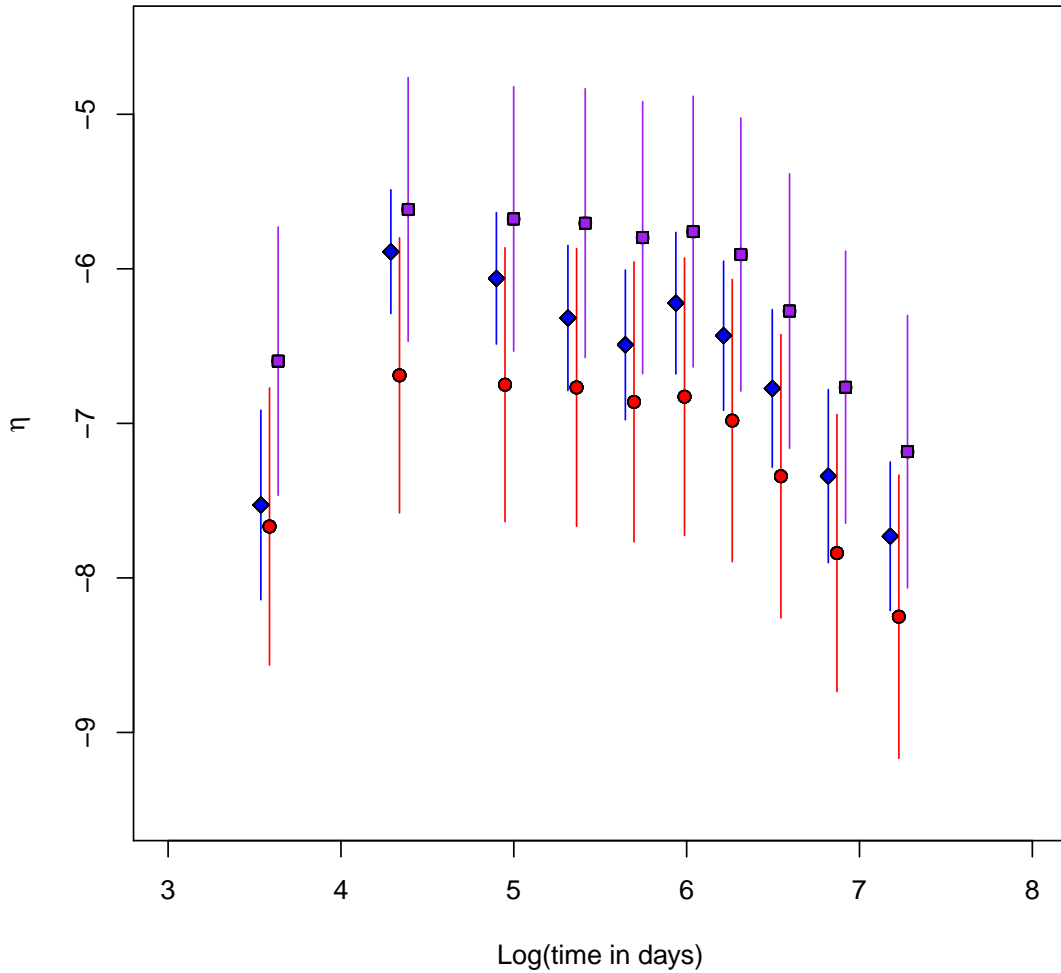


Figure 8.3: Time plots of the posterior means and ± 2 standard deviation intervals for three log-hazards using the GPCH model where priors are based on parametric model without overlapping log-hazards for the leukaemia data set

The R functions to do this computation using the second form of covariance matrix is also given in Appendix A.5.2 and A.5.3. Following a burn-in of 20000 iterations, 60000 iterations were taken. The posterior means and standard deviations of some of the log-hazards are given in Table 8.8 (see basic Gibbs sampler II). The posterior summaries using both forms are slightly different.

Table 8.8: Posterior means and standard deviation of some chosen parameters using the GPCH model with priors based on parametric model

Parameter	Basic Gibbs sampler I		Basic Gibbs sampler II	
	Post.means	Post.std	Post.means	Post.std
η_1	-6.961	0.669	-7.667	0.448
η_2	-7.363	0.595	-6.689	0.445
η_3	-7.080	0.611	-6.750	0.443
η_4	-6.950	0.609	-6.767	0.449
η_5	-6.848	0.574	-6.861	0.453
η_6	-6.445	0.579	-6.827	0.449
η_7	-6.436	0.594	-6.982	0.457
η_8	-6.496	0.664	-7.342	0.458
η_9	-7.126	0.678	-7.839	0.448
η_{10}	-7.786	0.643	-8.251	0.458

8.3 Some alternative numerical approaches in GPCH model with priors based on parametric model

There is a need to suggest alternative numerical approaches to improve the mixing when using the priors based on parametric model. Sometimes, the reason for the poor mixing could be that the parameters of the posterior distribution are correlated and hence, we could aim at block sampling of the parameters. We would want to improve the mixing using some numerical approaches which include splitting the log-hazards and sampling using either the forward backward algorithm or Cholesky decomposition and by using a Gaussian approximation of the fcd as a proposal distribution. We could also sample the principal components of the log-hazards.

8.3.1 Sampling principal components using priors based on parametric model

We follow the discussions in Section 6.8 to sample the principal components of the log-hazards. We use a function in R package to compute eigen values and vectors. Thus, we have the vector of means and the covariance matrix of the principal component following Equation (6.32) and (6.31) respectively. The functions to sample the principal components of the log-hazards are given in Appendix A.5.6 and A.5.3.

Following a burn-in of 3000 iterations, 10000 samples were collected. Convergence was assessed by visual inspection of trace plots of the parameters and the mixing was good.

8.3.2 Gaussian approximation of the fcd as a proposal distribution using priors based on parametric model

We recall the overall logarithm of the likelihood given in Equation 6.6 as

$$\ell = \log L = \sum_{j=1}^J \sum_{p=1}^P n_{d,j,p} \lambda_{j,p} - \sum_{j=1}^J \sum_{p=1}^P \lambda_{j,p} T_{j,p}^* \quad (8.1)$$

The prior density of $\underline{\eta}$ is

$$\pi(\underline{\eta}) \propto \exp \left\{ -\frac{1}{2} (\underline{\eta} - \underline{\mu})^T Q (\underline{\eta} - \underline{\mu}) \right\}$$

where Q is the precision matrix. Then,

$$\log (\pi(\underline{\eta})) = (\text{constant}) - \frac{1}{2} (\underline{\eta} - \underline{\mu})^T Q (\underline{\eta} - \underline{\mu}) .$$

We wish to approximate the function

$$\begin{aligned} g &= \log [\pi(\underline{\eta}|D)] \\ &= \ell + \log (\pi(\underline{\eta})) \end{aligned}$$

The partial first derivative with respect to η_p is

$$g'_p = \frac{\partial g}{\partial \eta_p} = n_{d,j,p} - \lambda_{j,p} T_{j,p}^* - [Q(\underline{\eta} - \underline{\mu})]_p .$$

Let $g' = (g'_1, \dots, g'_P)^T$, let $\underline{n}_d = (n_{d,j,1}, \dots, n_{d,j,P})^T$ and let $\underline{\lambda}^* = (\lambda_1^*, \dots, \lambda_P^*)^T$ where

$$\lambda_p^* = \lambda_{j,p} T_{j,p}^* .$$

So

$$g' = \underline{n}_d - \underline{\lambda}^* - Q(\underline{\eta} - \underline{\mu}) .$$

The partial second derivatives are

$$g''_{p,p} = \frac{\partial^2 g}{\partial \eta_p^2} = -\lambda_{j,p} T_{j,p}^* - Q_{p,p}$$

and

$$g''_{p,p'} = \frac{\partial^2 g}{\partial \eta_p \partial \eta_{p'}} = -Q_{p,p'}$$

for $p \neq p'$, where the element in row p and column p' of Q is $Q_{p',p}$. Hence, the matrix of

partial second derivatives is

$$G'' = -Q - \text{diag}(\underline{\lambda}^*) . \quad (8.2)$$

Let this matrix, when evaluated at the posterior mode, be G''_{mode} . Then, the posterior precision matrix is approximately

$$Q_1 = -G''_{\text{mode}} = Q + \text{diag}(\underline{\lambda}^*_{\text{mode}}) \quad (8.3)$$

where $\underline{\lambda}^*_{\text{mode}}$ is $\underline{\lambda}^*$ evaluated at the posterior mode. We could choose the maximum likelihood estimates as the starting parameter values of the log-hazards. The maximum likelihood estimate of η_p is computed from logarithm of the likelihood given in Equation (6.4) and is

$$\eta_0 = \frac{n_{d,j,p}}{T_{j,p}^*}$$

The approximated proposal distribution now has mean $\underline{\lambda}^*_{\text{mode}}$ and the precision matrix as Q_1 . We use the approximated multivariate normal distribution as the proposal distribution. We propose a new value of $\underline{\eta}$, $\underline{\eta}^*$ from a multivariate normal distribution using the Cholesky decomposition discussed in Section 4.6.6. We compute the proposal density of $\underline{\eta}^*$ given $\underline{\eta}$, $q(\underline{\eta}^*|\underline{\eta})$ and the proposal density of $\underline{\eta}$ given $\underline{\eta}^*$, $q(\underline{\eta}|\underline{\eta}^*)$. The Metropolis-Hastings algorithm will be used and the functions used are given in Appendix A.5.7 and A.5.3.

We have taken 20000 iterations and convergence was assessed by visual inspection of trace plots of the parameters and the mixing was good. The trace plots of some of the parameters (log-hazards) are shown in Figure 8.4.

Figure 8.4 shows some very short periods where the sampler sticks.

8.4 Splitting the log-hazards

In the context of the GPCH model, we introduce the frailties as usual, one frailty per patient, then η_i for patient i is a vector (over time intervals). In this case, each patient has one random frailty in all time intervals. In another case, we could have that each patient has a different frailty in each time interval. These frailties could be referred to as random patient-time period interactions (see Section 6.7).

We split the log-hazard of an individual η_i into two parts so that one part θ_i , carries the dependence and the other carries the individual's log-frailty Z_i since the full conditional distribution of the log-hazards $\pi(\underline{\eta}|D)$ given in Equation (6.27) will not be a multivariate normal distribution. Let the vector of log-frailties of the individuals be \underline{Z} . We recall the construction of the prior distribution of the precision of log-frailty with parameters a_f and b_f discussed in Section 5.7.4. We use the Metropolis within Gibbs algorithm to sample the parameters. First, we sample the precision of the log-frailty, τ_f from its full

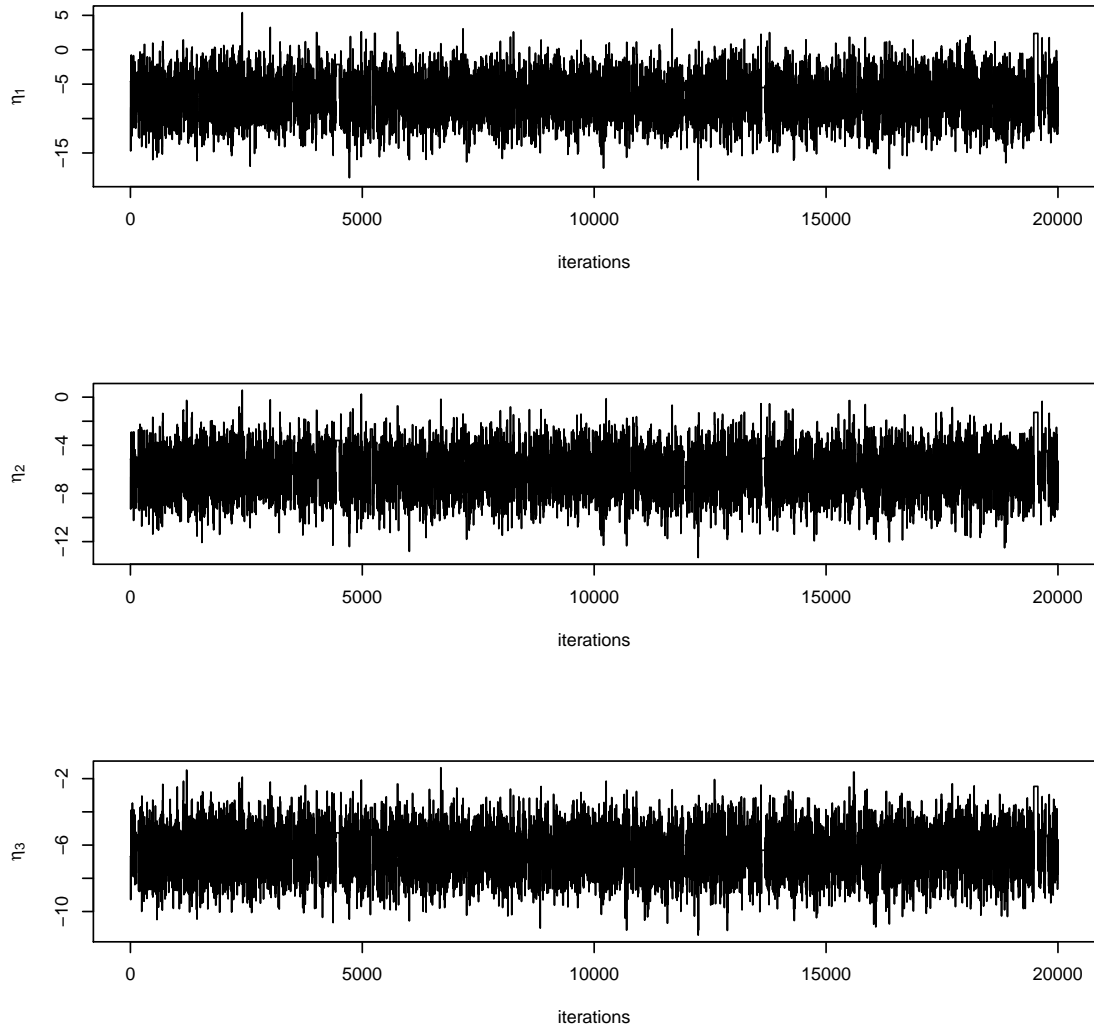


Figure 8.4: Trace plot for three log-hazards when the proposal distribution was approximated using the GPCH model and priors based on the parametric model for the leukaemia data set

conditional distribution given in Equation (6.30). Then, we sample the η 's one at a time and the vector of log-hazards, $\underline{\theta}$ will now have a multivariate normal distribution. We can use either the forward backward algorithm discussed in Section 4.7.1 by extending it to the case of a vector or we can also use the Cholesky decomposition to sample the vector of log-hazards from the multivariate normal distribution.

In the case where we have one random frailty per patient, we can not use the forward backward algorithm in the time direction but we could have the forward backward algorithm in the direction of another covariate. In the other case where each patient has a different random frailty in each time period, the forward backward algorithm is

practicable. We illustrate using the latter case.

We suppose that we have “J” time intervals or cutpoints and we have a vector of means and a covariance matrix for the vector of log-hazards for each time period. We use the forward backward recursion to sample the $\underline{\theta}$'s over the time periods since we have observed data for $\underline{\eta}_1, \dots, \underline{\eta}_t$. For instance, in the forward recursion, we change the distribution of 1st vector of log-hazard $\underline{\theta}_1$ by conditioning on the observed vector of $\underline{\eta}_1$. The distribution of $f(\underline{\theta}_1|\underline{\eta}_1)$ has a conditional precision matrix $Q_{\underline{\theta}_1|\underline{\eta}_1}$ given by

$$Q_{\underline{\theta}_1|\underline{\eta}_1} = Q_{\underline{\theta}_1} + Q_{data}$$

and a conditional mean vector $\underline{\mu}_{\underline{\theta}_1|\underline{\eta}_1}$ given by

$$\frac{Q_{\underline{\theta}_1}\underline{\mu}_1 + Q_{data}\underline{\eta}_1}{Q_{\underline{\theta}_1} + Q_{data}}$$

where Q_{data} is a data precision matrix given as

$$Q_{data} = M\tau_f$$

where M is a diagonal matrix with elements corresponding to the number of individuals with that covariate profile. We have that $Q_{\underline{\theta}_1}$ is the prior precision matrix of $\underline{\theta}_1$ and $\underline{\mu}_1$ is the prior mean of $\underline{\theta}_1$.

We store the distribution of $f(\underline{\theta}_1|\underline{\eta}_1)$ which will in turn change the distribution of the vector $\underline{\theta}_2$ since $\underline{\theta}_2 = \underline{\theta}_1 + W$ for $W \sim N(0, V_w)$. We then condition on $\underline{\eta}_2$ and get the distribution of $\underline{\theta}_2, f(\underline{\theta}_2|\underline{\eta}_1, \underline{\eta}_2)$. We keep conditioning until we have conditioned on all the $\underline{\eta}_t$'s for each time period and we get the distribution of $\underline{\theta}_J, f(\underline{\theta}_J|\underline{\eta}_1, \dots, \underline{\eta}_t)$. We store all conditional distributions in the forward recursion.

We start the backward recursion by sampling $\underline{\theta}_J$ from the multivariate normal distribution $f(\underline{\theta}_J|\underline{\eta}_1, \dots, \underline{\eta}_J)$ using any convenient method like Cholesky decomposition (see Section 4.6.6) and this in turn changes the value of $\underline{\theta}_{J-1}$. We will go back to the distribution of $\underline{\theta}_{J-1}$ given $\underline{\eta}_1, \dots, \underline{\eta}_{J-1}, f(\underline{\theta}_{J-1}|\underline{\eta}_1, \dots, \underline{\eta}_{J-1})$. We condition on $\underline{\theta}_J$ since we have now observed $\underline{\theta}_J, f(\underline{\theta}_{J-1}|\underline{\eta}_1, \dots, \underline{\eta}_J, \underline{\theta}_J)$. Then, we sample $\underline{\theta}_{J-1}$ from $f(\underline{\theta}_{J-1}|\underline{\eta}_1, \dots, \underline{\eta}_J, \underline{\theta}_J)$. We will keep conditioning on the previous known values of $\underline{\theta}_J$'s until we have the distribution of $\underline{\theta}_1$ given all other $\underline{\theta}_J$'s, $f(\underline{\theta}_1|\underline{\eta}_1, \dots, \underline{\eta}_J, \underline{\theta}_2, \dots, \underline{\theta}_J)$ and we sample $\underline{\theta}_1$ from this distribution. Thus, we have samples for the vectors of log-hazards for all time periods.

The R functions to do this are given in Appendix A.5.4 and A.5.3. Following a burn-in of 5,000 iterations, 20,000 iterations were taken. Convergence was assessed by visual inspection of trace plots of the parameters and the mixing was good. The posterior summaries of some of the parameters are given in Table 8.9 (see Split model I).

We could sample the vector of log-hazards $\underline{\theta}$ from the multivariate normal distribution using the Cholesky decomposition. The R functions to do this is given in Appendix A.5.5 and A.5.3. Following a burn-in of 5,000 iterations, 20,000 iterations were taken. Convergence was assessed by visual inspection of trace plots of the parameters and the mixing was good. The posterior summaries of some of the parameters are given in Table 8.9 (see Split model II).

Table 8.9: Posterior means and standard deviation of some chosen parameters when we split the log-hazards using the GPCH model with priors based on parametric model

Parameter	Split model I		Split model II	
	Post.means	Post.std	Post.means	Post.std
η_1	-7.027	1.419	-7.148	0.678
η_2	-7.023	1.684	-6.603	0.647
η_3	-7.024	2.007	-7.515	0.862
η_4	-7.021	2.269	-7.073	0.635
η_5	-7.036	2.522	-6.861	0.597
η_6	-7.042	2.760	-7.013	0.619
η_7	-7.051	2.965	-6.944	0.590
η_8	-7.052	3.155	-6.823	0.619
η_9	-7.050	3.351	-6.766	0.619
η_{10}	-7.055	3.527	-7.180	0.681

8.5 Application 2: GPCH model with priors based on Gaussian Markov random field

We use similar ordering for the covariates in Table 8.1 since it is the same model but we have different priors. We use all categorical covariates including time interval as a covariate in the design matrix X following our discussion in Section 6.3.1. There will be 20 parameters in the linear model, including the intercept β_0 . Hence, the design matrix X is a 1280 by 20 matrix.

Prior distributions

We refer to Table 8.2 for the means and standard deviations of the coefficients of the ordinal covariates. We construct prior distribution for the time variable in a similar way. Again, we construct the vector of means $\underline{\mu}$ and covariance matrix V_{η}^* of the log-hazards corresponding to the cells given in Equation (6.12) and Equation (6.13) respectively. We construct a suitable covariance matrix for $\underline{\eta}$, $\text{Var}(\underline{\eta})$ given in Equation (6.14). We find the distances between the log-hazards corresponding to the cells using Equation (6.16). We

think of the value of a in Equation 6.15. The value of a corresponds to the correlation between the log-hazards irrespective of how near their covariate profiles are. It could also be thought of as the overall uncertainty of how long the individuals with leukaemia will live irrespective of their covariates. We will suppose the value 0.2 for a . Hence, we have the covariance matrix $\text{Var}(\eta)$. We have partially displayed the correlation between some covariate vectors in Table 8.10 and covariance matrix of some of the log-hazards in Table 8.11.

Table 8.10: Partial display of the correlation between ordinal covariate vectors using the Markov random field prior with the leukaemia data set

x_1	x_2	x_3	x_4	x_5
1.000	0.924	0.855	0.793	0.736
0.924	1.000	0.924	0.855	0.793
0.855	0.924	1.000	0.924	0.855
0.793	0.855	0.924	1.000	0.924
0.736	0.793	0.855	0.924	1.000

Table 8.10 shows that the correlations between the pair of log-hazards are very strong.

Table 8.11: Partial display of the prior covariance matrix of the log-hazards using the Markov random field prior with the leukaemia data set

2.422	1.996	1.749	1.569	1.115
1.996	1.927	1.686	1.510	1.071
1.749	1.686	1.727	1.545	1.094
1.569	1.510	1.545	1.618	1.144
1.115	1.071	1.094	1.144	0.947

Posterior distributions

The joint posterior distribution has been discussed in Section 6.6.3. The Metropolis within Gibbs algorithm can be applied to sampling the log-hazards one at time using the Gaussian Markov random field prior and the functions to do this are given in Appendix A.5.9 and A.5.8. Following a burn-in of 5000 iterations, 30000 iterations were taken. Convergence was assessed by visual inspection of trace plots of the parameters and the mixing was good. The trace plots and density plots of some of the parameters are shown in Figure 8.5 and the summary of the log-hazards of one of the covariate profiles at different intervals are also given in Table 8.12 (see Basic Gibbs sampler).

Again, Figure 8.5 shows good mixing of the trace plots of the log-hazards of three

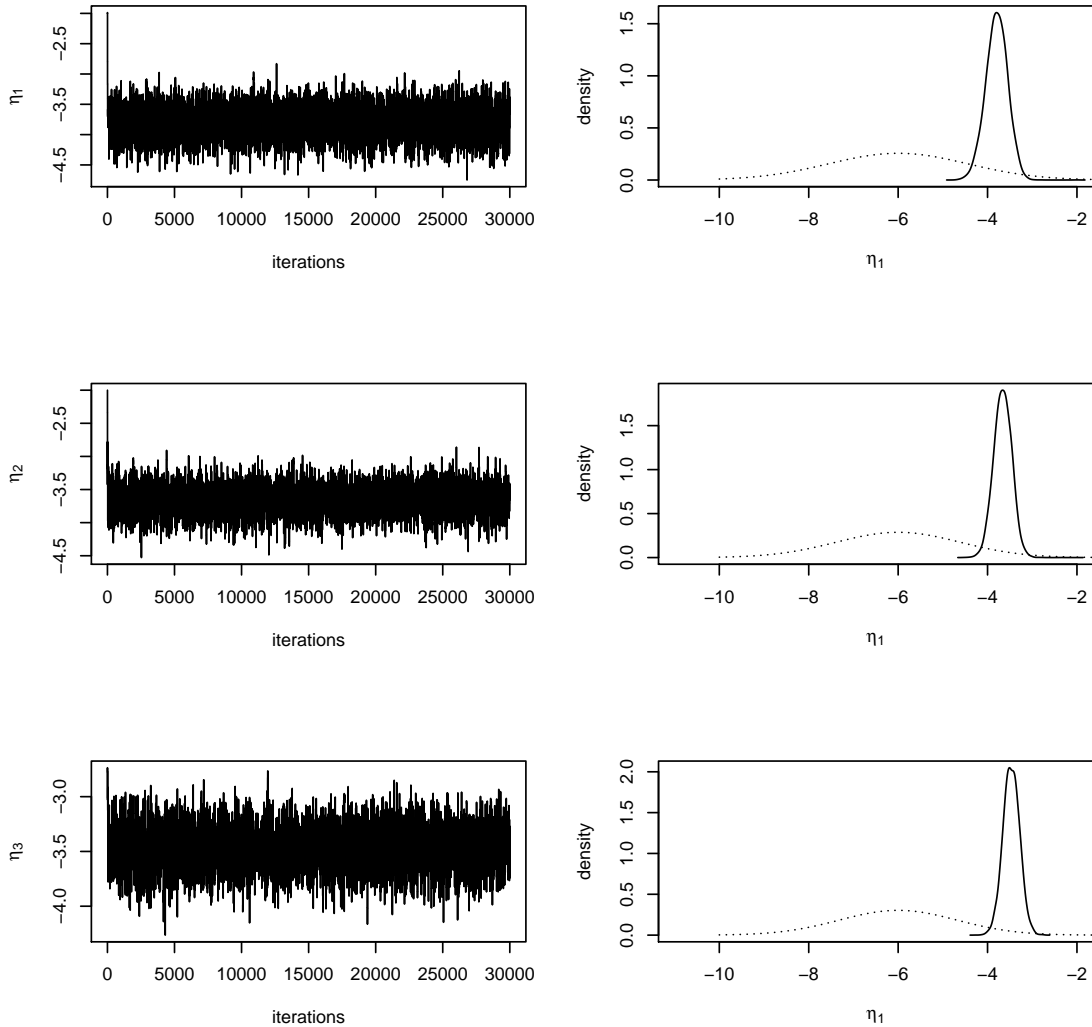


Figure 8.5: Trace plot and density plot (with dash lines representing the prior density and straight lines representing the posterior density) for three log-hazards using the GPCH model with Markov random field priors for the leukaemia data set

covariate profiles. The density plots also show that we have learnt something from the data.

We have also plotted the posterior means and ± 2 standard deviation intervals for three log-hazards at different time intervals using the GPCH model with Markov random field priors in Figure 8.6.

In Figure 8.6, we have started at different time intervals to avoid overlapping. We can see clear differences among the log-hazards from the three different covariate profiles.

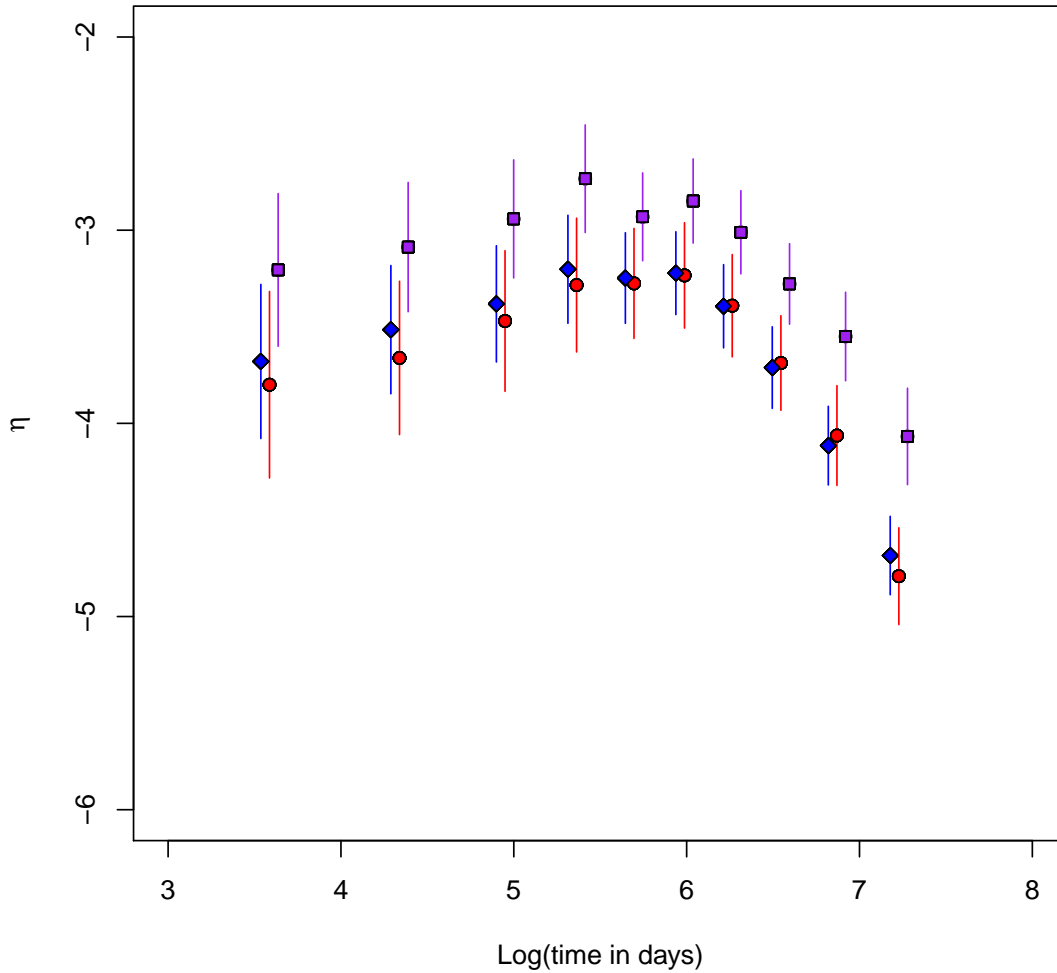


Figure 8.6: Time plots of the posterior means and ± 2 standard deviation intervals for three log-hazards using the GPCH model with Markov random field priors without overlapping log-hazards for the leukaemia data set

8.6 Some alternative numerical approaches in GPCH model with Markov random field prior

There is also a need to suggest alternative numerical approaches to improve the mixing in the model using the Markov random field prior like in the case of using priors based on the parametric model. We follow all approaches discussed in Section 8.3.

8.6.1 Sampling principal components using Markov random field prior

We follow a similar approach discussed in Section 8.3.1 to sample the principal components of the log-hazards. The functions to sample the principal components of the log-hazards are given in Appendix A.5.10 and A.5.8 using the Markov random field prior. Following a burn-in of 5000 iterations, 10000 iterations were taken. Convergence was assessed by visual inspection of trace plots of the parameters and the mixing was good.

8.6.2 Gaussian approximation of the proposal distribution using Markov random field prior

We follow a similar approach discussed in Section A.5.13 to get a Gaussian approximation of the proposal distribution and sample the log-hazards. The functions to do this are given in Appendix A.5.11 and A.5.8. Following a burn-in of 5000 iterations, 20000 iterations were taken. Convergence was assessed by visual inspection of trace plots of the parameters. The trace plots of some of the parameters (log-hazards) are shown in Figure 8.7. Figure 8.7 shows periods where the sampler sticks.

8.7 Splitting the log-hazards

We follow a similar approach discussed in Section 8.4 to sample the log-hazards with the forward backward algorithm using the Markov random field prior. The functions to do this are given in Appendix A.5.11 and A.5.8. Following a burn-in of 5000 iterations, 20,000 iterations were taken. Convergence was assessed by visual inspection of trace plots of the parameters and the mixing was good. The posterior summaries of some of the parameters are also given in Table 8.12 (see Split model I).

Again, we could sample the vector of log-hazards θ from the multivariate normal distribution using the Cholesky decomposition. The functions to do this are given in Appendix A.5.12 and A.5.8. Following a burn-in of 5,000 iterations, 20,000 iterations were taken. Convergence was assessed by visual inspection of trace plots of the parameters and the mixing was good.

The method of splitting the log-hazard is a different model and we will not expect to have similar results.

8.8 Application 3: Continuous parameter space model with Gaussian process prior

We revisit the SNLG data set collected by the Scotland and Newcastle Lymphoma Group in Section 2.2.4. The dataset includes 636 patients with 14 completely observed covariates

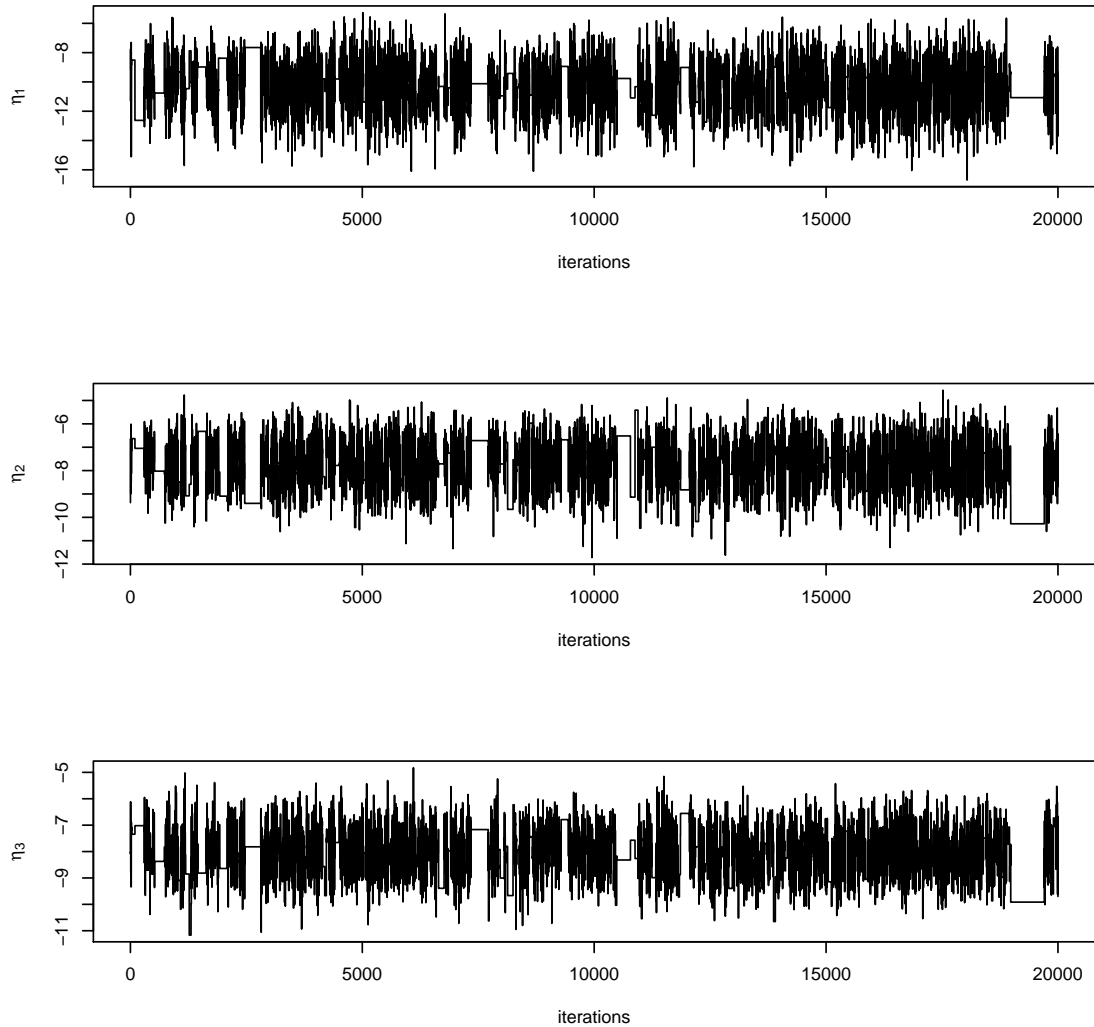


Figure 8.7: Trace plots for three log-hazards with approximated proposal distribution with Markov random field priors for the leukaemia dataset

(see Appendix A.3.1 for a partial display of data). We follow the ordering of the covariates which are given in Chapter 5, Table 5.1. We refer to the Chapter 5, Table 5.2 for the prior distribution of the parameters of the linear model. Again, we will assume that the parameters of the linear model are not independent and so we construct the prior covariance matrix using the discussion in Section 4.4.1. We construct the design matrix X following our discussion in Section 6.3.1 for the categorical covariates and in Section 7.2 for continuous covariates. There will be 21 parameters in the linear model, including an intercept β_0 . We note that we have 636 distinct covariate profiles in the SNLG data set and hence, the design matrix X is a 636 by 21 matrix.

Table 8.12: Posterior means and standard deviation of some chosen parameters using the GPCH model with Markov random field prior

Parameter	Basic Gibbs sampler		Split model I	
	Post.means	Post.std	Post.means	Post.std
η_1	-3.800	0.241	-8.490	0.963
η_2	-3.661	0.198	-8.553	1.154
η_3	-3.470	0.182	-8.588	1.362
η_4	-3.284	0.173	-8.746	1.574
η_5	-3.275	0.142	-8.909	1.714
η_6	-3.234	0.136	-9.044	1.802
η_7	-3.391	0.132	-9.203	1.860
η_8	-3.688	0.122	-9.369	1.872
η_9	-4.063	0.129	-9.391	1.888
η_{10}	-4.791	0.125	-9.717	2.076

8.8.1 Construction of prior distribution for the log-hazards

We recall the means and standard deviations of the coefficients of the covariates given in Table 5.2. We construct the vector of means $\underline{\mu}$ and covariance matrix V_{η}^* of the log-hazards of the distinct covariate profiles given in Equation (6.12) and Equation (6.13) respectively. We construct a suitable covariance matrix for $\underline{\eta}$, $\text{Var}(\underline{\eta})$ given in Equation (6.14). We construct the distances between the covariate profiles using Equation (7.6). We would think that the proportion of the variance representing the overall uncertainty of how long the individuals with non-Hodgkins lymphoma will live irrespective of their covariates is 0.2. Hence, we take the value of a in Equation (7.5) as 0.2. Hence, we have the covariance matrix $\text{Var}(\underline{\eta})$. We have partially displayed the correlation between some covariate vectors in Table 8.13 and covariance matrix of some of the log-hazards in Table 8.14.

Table 8.13: Partial display of the correlation between covariate vectors of the SNLG data set

\underline{x}_1	\underline{x}_2	\underline{x}_3	\underline{x}_4	\underline{x}_5
1.000	0.889	0.861	0.838	0.877
0.889	1.000	0.874	0.867	0.891
0.861	0.874	1.000	0.888	0.881
0.838	0.867	0.888	1.000	0.871
0.877	0.891	0.881	0.871	1.000

Table 8.13 shows that the correlations between the pair of log-hazards are strong.

Table 8.14: Partial display of the prior covariance matrix of the log-hazards of the SNLG data set

3.117	0.809	1.114	2.764	0.948
0.809	0.266	0.331	0.834	0.281
1.114	0.331	0.538	1.216	0.396
2.764	0.834	1.216	3.486	0.995
0.948	0.281	0.396	0.995	0.375

8.8.2 Posterior distributions

The joint posterior distribution of the log-hazards has been discussed in Section 7.3.2. We sample the log-hazards one at a time using the Metropolis within Gibbs algorithm and the functions to do this are given in Appendix A.5.15 and Appendix A.5.14. Following a burn-in of 5,000 iterations, 30,000 iterations were taken. Convergence was assessed by visual inspection of trace plots of the parameters and the mixing was good. The trace plots of some of the parameter are shown in Figure 8.8. The posterior summaries of some of the parameters are also given in Table 8.15 (see basic Gibbs sampler).

Figure 8.8 shows good mixing of the trace plots of the log-hazards of three covariate profiles. Again, the density plot also show that we have learnt something from the data.

8.9 Some alternative numerical approaches in continuous parameter space model

In some cases, the parameters of the posterior distribution might be strongly correlated and hence we have poor mixing. We would want to improve the mixing using some numerical approaches by using block sampling of the parameters. These approaches include splitting the log-hazards and sampling them using the Cholesky decomposition and Gaussian approximation of the fcd as a proposal distribution. We could also sample the principal components of the log-hazards. We would not be able to use the forward backward algorithm to improve the mixing as the log-hazards in the continuous parameter space model using a Gaussian process prior do not follow a Markov process.

8.9.1 Improving mixing in continuous parameter space model by sampling the principal components of the log-hazards

Again, we follow the discussions in Section 6.8 to sample the principal components of the log-hazards. We use a function in R package to compute eigen vectors and we have the vector of means and the covariance matrix of the principal component following Equation (6.32) and (6.31) respectively. We could use the Metropolis within Gibbs algorithm to

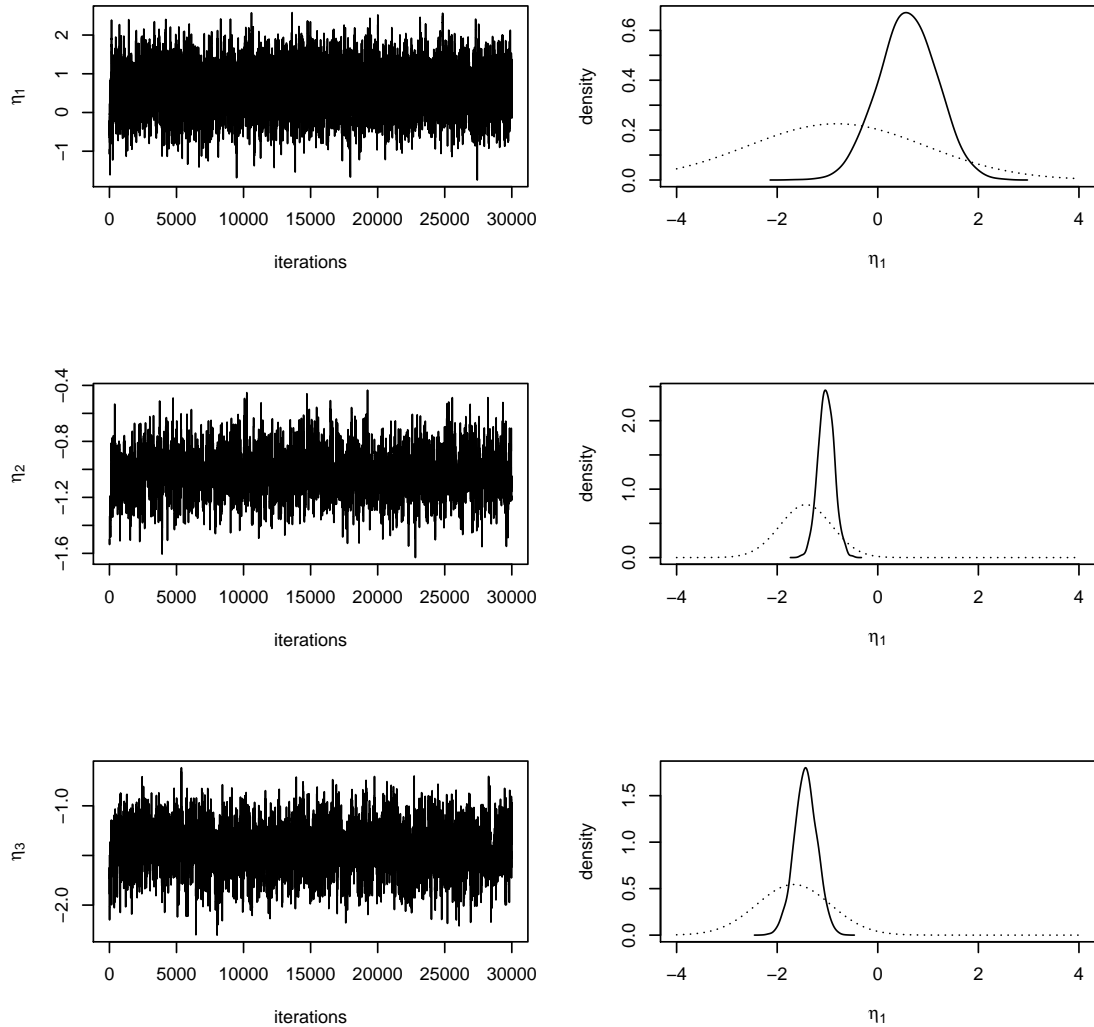


Figure 8.8: Trace plot and Density plot for three log-hazards using the continuous parameter space model with Gaussian process priors for the SNLG data set

sample since we can not sample from the full conditional distribution. We fix the Weibull shape parameter α while we follow the discussions in Section 6.8 to sample the principal components of the log-hazards. We then fix the sampled values of the log-hazards and we sample α . The functions to do this are given in Appendix A.5.16 and A.5.14. Following a burn-in of 3000 iterations, 7000 iterations were taken. Convergence was assessed by visual inspection of trace plots of the parameters and the mixing was good (see Figure 8.9).

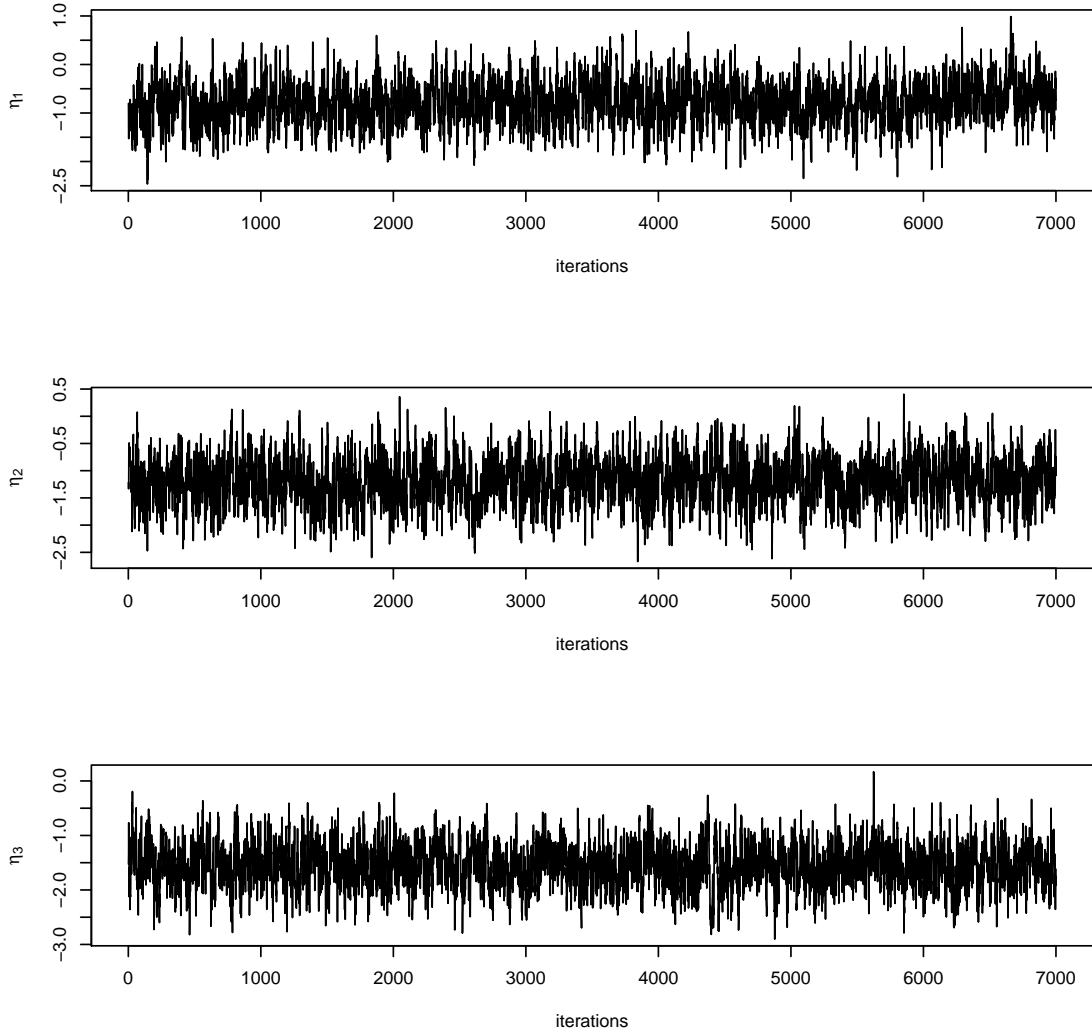


Figure 8.9: Trace plot when we sample the principal components of the log-hazards using the continuous parameter space model with Gaussian process priors for the SNLG data set

8.9.2 Improving mixing in continuous parameter space model by Gaussian approximation fcd of the log-hazards

We recall the overall log likelihood given in Equation (7.3) as

$$\ell = n_d \log \alpha + \sum_{p=1}^{n'} n_{d,p} \eta_i + (\alpha - 1) \sum_{p=1}^{n'} \sum_{k=1}^{n_p} d_{p,k} \log t_{p,k} - \sum_{p=1}^{n'} \left\{ \lambda_p \sum_{k=1}^{n_p} t_{p,k}^\alpha \right\}. \quad (8.4)$$

The prior density of $\underline{\eta}$ is

$$\pi(\underline{\eta}) \propto \exp \left\{ -\frac{1}{2}(\underline{\eta} - \underline{\mu})^T Q(\underline{\eta} - \underline{\mu}) \right\}$$

where Q is the precision matrix. Then,

$$\log [\pi(\underline{\eta})] = (\text{constant}) - \frac{1}{2}(\underline{\eta} - \underline{\mu})^T Q(\underline{\eta} - \underline{\mu})$$

We wish to approximate the function

$$g = \ell + \log [\pi(\underline{\eta})]$$

The partial first derivative with respect to η_p is

$$g'_p = \frac{\partial g}{\partial \eta_p} = n_{d,p} - \lambda_p \sum_{k=1}^{n_p} t_{p,k}^\alpha - [Q(\underline{\eta} - \underline{\mu})]_p.$$

Let $g' = (g'_1, \dots, g'_p)^T$, let $\underline{n}_d = (n_{d,1}, \dots, n_{d,p})^T$ and let $\underline{\lambda}^* = (\lambda_1^*, \dots, \lambda_p^*)^T$ where

$$\lambda_p^* = \lambda_p \sum_{k=1}^{n_p} t_{p,k}^\alpha.$$

So

$$g' = \underline{n}_d - \underline{\lambda}^* - Q(\underline{\eta} - \underline{\mu}).$$

The partial second derivatives are

$$g''_{p,p} = \frac{\partial^2 g}{\partial \eta_p^2} = -\lambda_p \sum_{k=1}^{n_p} t_{p,k}^\alpha - Q_{p,p}$$

and

$$g''_{p,p'} = \frac{\partial^2 g}{\partial \eta_p \partial \eta_{p'}} = -Q_{p,p'},$$

for $p \neq p'$, where the element in row p and column p' of Q is $Q_{p',p}$. Again, the posterior precision matrix follow Equation (8.3) and $\underline{\lambda}_{\text{mode}}^*$ is $\underline{\lambda}^*$ evaluated at the posterior mode. The maximum likelihood estimate of η_p is

$$\eta_0 = \frac{n_{d,p}}{\sum_{k=1}^{n_p} t_{p,k}^\alpha}$$

We use a Metropolis within Gibbs algorithm to sample the log-hazards and the Weibull shape parameter α . We fix the value of α while we use the approximated multivariate

normal distribution as the proposal distribution. We propose a new value of $\underline{\eta}$, $\underline{\eta}^*$ from a multivariate normal distribution using the Cholesky decomposition discussed in Section 4.6.6. We compute the proposal density of $\underline{\eta}^*$ given $\underline{\eta}$, $q(\underline{\eta}^*|\underline{\eta}, \alpha)$ and the proposal density of $\underline{\eta}$ given $\underline{\eta}^*$, $q(\underline{\eta}|\underline{\eta}^*, \alpha)$. The Metropolis-Hastings algorithm will be used to either accept or reject the proposed log-hazard η_p^* . We then fix the accepted value of the log-hazards and we sample α . The functions to do this are given in Appendix A.5.18 and Appendix A.5.14. Following a burn-in of 10,000 iterations, 30,000 iterations were taken. Convergence was assessed by visual inspection of trace plots of the parameters. The trace plots of some of the parameters (log-hazards) are shown in Figure 8.10 and it shows that there are some periods where the sampler sticks. It is obvious that we might need a much longer number of iterations. Further work is needed to address this problem.

Furthermore, we will try reducing the number of parameters to two so that we would possibly plot a contour plot of the posterior and proposal density. This will enable us know if the posterior density is very close to our approximated proposal density. If both plots are similar, then we have a good approximation for our proposal distribution. We will illustrate this using the case where we assume that we had only one covariate such as sex. This reduces the number of covariate profiles to two and we can have a 2-dimensional plot such as a contour plot while we fix the value of the Weibull shape parameter. We have contour plots of the posterior and proposal distribution in Figure 8.11. The plots of both distributions look very much similar. It is very obvious that the posterior and proposal distributions look very much alike. We also have trace plots of the two log-hazards (η_1 and η_2) and Weibull shape parameter α in Figure 8.12.

8.10 Splitting the log-hazards and sampling from the block of the log-hazards in the continuous parameter space model

We follow the similar approach discussed in Section 8.4 to split the log-hazard of an individual, η into two parts. Again, we could use the Metropolis within Gibbs algorithm to sample the parameters. We sample the precision of the log-frailty τ_f from its full conditional distribution given in Equation (6.30). Then, we sample the η 's one at a time. We will also sample the Weibull shape parameter since we assume a Weibull parametric distribution and then the vector of log-hazards $\underline{\theta}$ will now have multivariate normal distribution. We could use the Cholesky decomposition to sample from the multivariate normal distribution.

The functions to do this are given in Appendix A.5.17 and A.5.14. Following a burn-in of 5,000 iterations, 20,000 iterations were taken. Convergence was assessed by visual inspection of trace plots of the parameters and the mixing was good. The summaries of

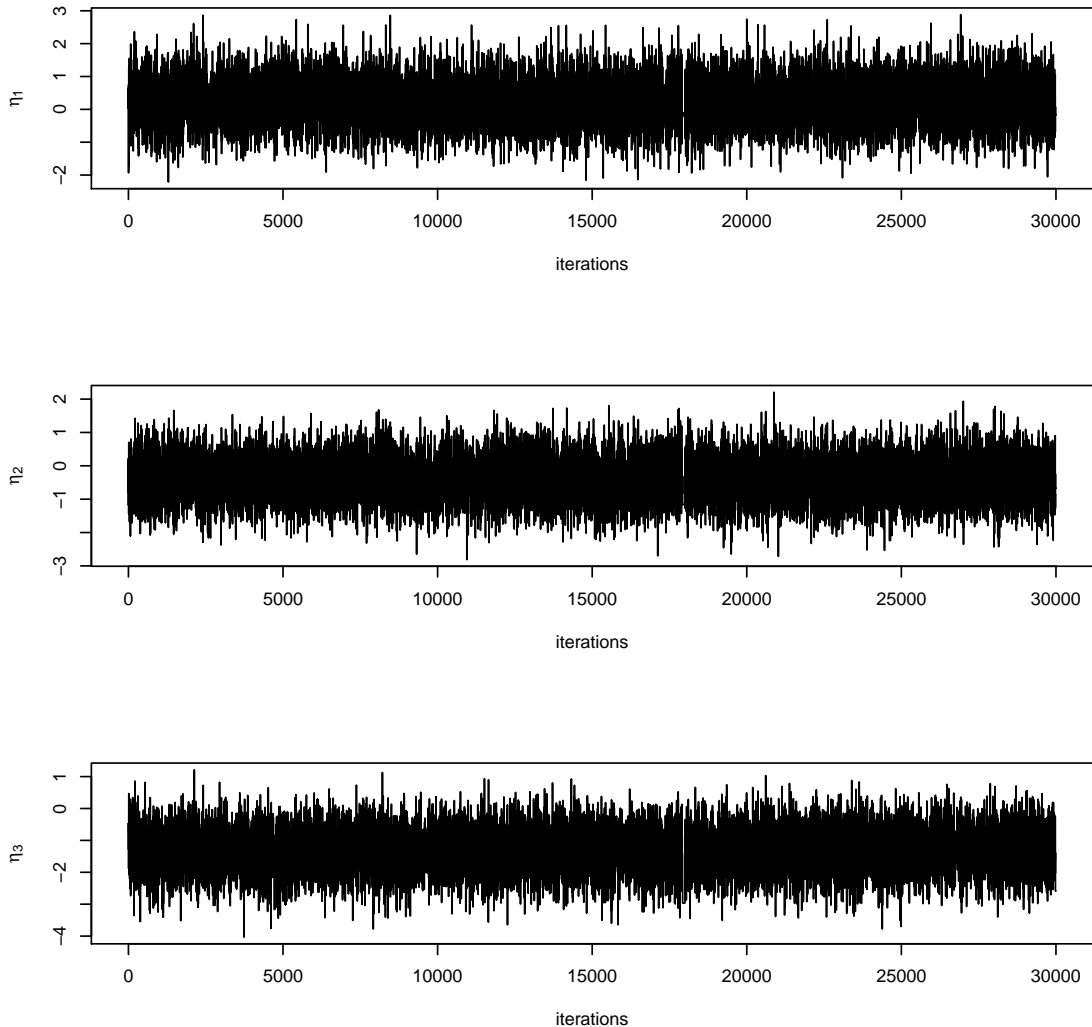


Figure 8.10: Trace plot for three log-hazards with Gaussian approximation of the proposal distribution using the continuous parameter space model with Gaussian process priors for the SNLG data set

some of the parameter are given on Table 8.15 (see Split model).

8.11 Calculation of effective sample size

We refer to Section 4.6.10 for the calculation of the effective sample size of the parameters of models. Effective sample size allows us to work out how to compare the numerical methods in terms of time and get the same accuracy. We have written the functions (see Appendix A.5.21) to calculate the effective sample size for all parameters of the model. We

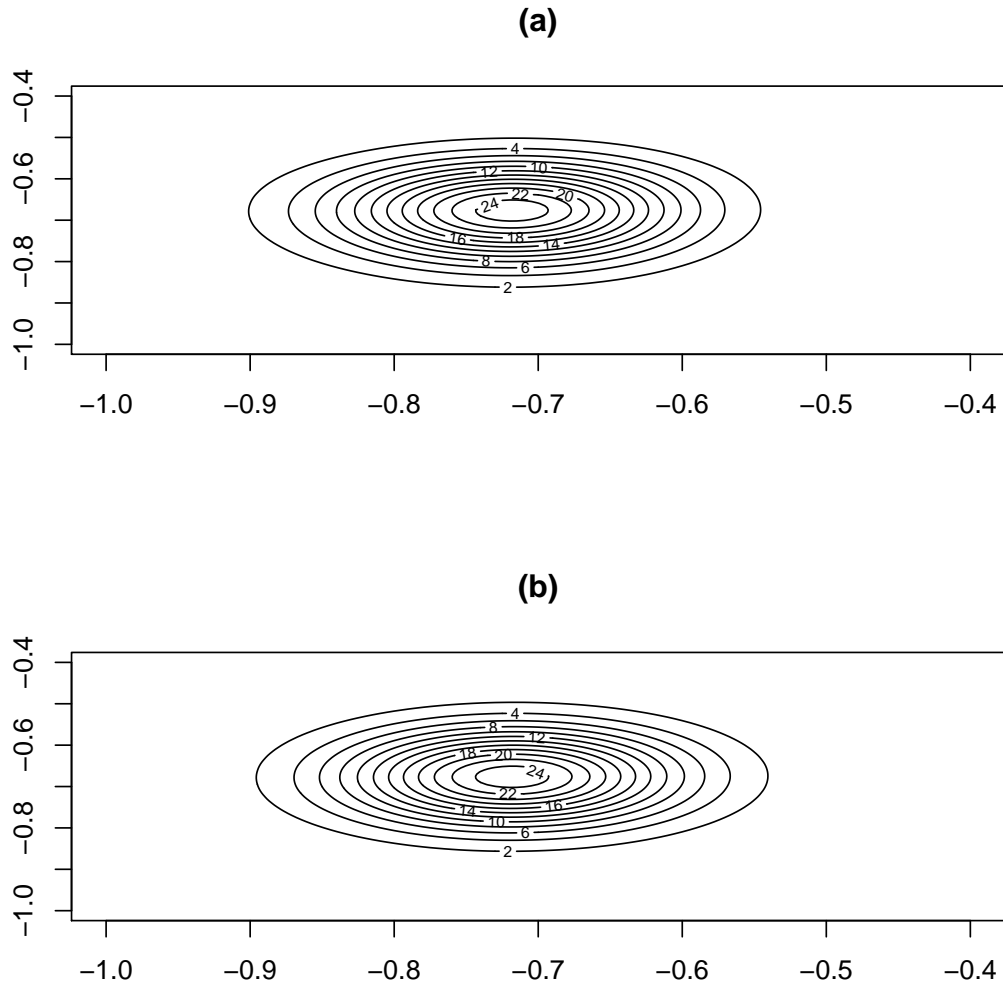


Figure 8.11: Contour plots of the posterior (a) and proposal density (b) of the log-hazards using only sex as covariate

have also calculated the average effective sample size (average ESS) over all parameters using the different numerical approaches. The time taken (see “time”) to run the number of iteration (n.iter) was also given in Table 8.16. We will want to know the time per unit of effective sample size. Thus, we divide the time taken to run the iterations by the average effective sample size and the unit is seconds. This gives time per unit of effective sample size.

We compare the time per unit of effective sample size of the basic Gibbs sampler, sampling the principal components of the log-hazards (Principal components), splitting the log-hazards and sampling using the forward backward algorithm (Splitting FBA) or

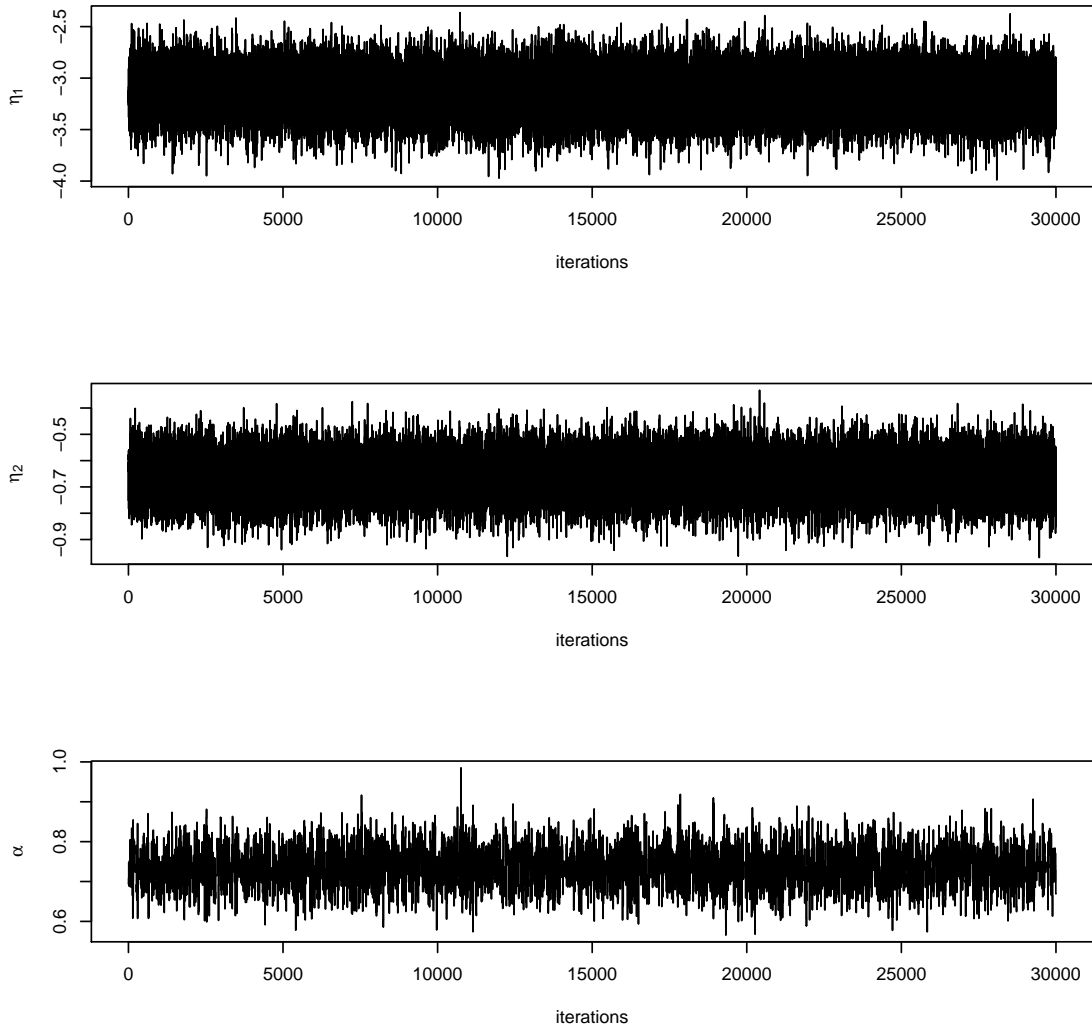


Figure 8.12: Trace plot of the log-hazards using only sex as covariate

Cholesky decomposition (Splitting CD) and Gaussian approximation of the fcd of the log-hazards (Gaussian approximation) for the methods in Table 8.16. In the GPCH model, the basic Gibbs sampler using the Markov random field prior has time per effective sample of 1.18 secs which is the smallest time and thus, it is the best method. On the other hand, the basic Gibbs sampler in the Continuous parameter space model has a time per effective sample of 4.25 secs which is also the best in this case.

Table 8.15: Posterior means and standard deviation of some chosen parameters using the continuous parameter space model and SNLG data set

Parameter	Basic Gibbs sampler		Split model	
	Post.means	Post.std	Post.means	Post.std
η_1	0.596	0.575	0.293	0.564
η_2	-1.024	0.159	-1.077	0.187
η_3	-1.438	0.227	-1.453	0.253
η_4	-2.173	0.517	-2.010	0.521
η_5	-1.238	0.181	-1.288	0.217
η_6	-1.581	0.405	-1.556	0.409
η_7	-0.792	0.431	-0.920	0.452
η_8	-1.542	0.374	-1.486	0.383
η_9	-1.433	0.778	-1.493	0.852
η_{10}	-1.000	0.370	-1.078	0.390
α	0.887	0.042	0.574	0.015

8.12 General discussion on methods and alternative numerical approaches

The investigation of flexible Bayesian modelling for relaxing the assumption of the standard proportional hazard model led to two approaches which were discussed in Chapters 6 and 7. These methods have been applied to our example data sets.

We illustrated the GPCH model with priors based on parametric model using two methods of constructing our covariance matrices. The correlations between the log-hazards in the first form of construction of the covariance matrix were much smaller than those from the latter. For instance, the correlation between the first and second log-hazards was 0.724 in the first form whereas it was 0.907 in the second form of construction of covariance matrix (see Tables 8.4 and 8.5 respectively). Tables 8.6 and 8.7 also show that the two forms of covariance matrices are not equivalent and hence the posterior summaries displayed in Table 8.8 slightly differ. We observed that in the approach where we sampled the principal components to improve the mixing, we did not need too many iterations for the MCMC algorithm to converge. The problem with this approach is that it takes a long time for each iteration. We worked out the number of iterations we needed to have particular effective sample size. Table 8.16 shows that the time per effective sample is 168.66 seconds. In the approach where we had block sampling of the parameters using Gaussian approximation of the fcd, the time per effective sample is 16.07 seconds but we need more iterations because of poor mixing. The approach where we split the log-hazards is a different model and we sampled the log-hazards either using the forward backward algorithm or Cholesky decomposition. The posterior summaries are not very different

Table 8.16: Computational and statistical efficiency

GPCH with priors based on parametric model				
Numerical approach	Average ESS	time (secs)	n.iter	Effect of sample size
Basic Gibbs sampler	957.81	10703.15	60000	11.17
Principal components	2085.583	211843.28	10000	101.58
Splitting FBA	1000.347	41788.42	20000	41.77
Splitting CD	160.71	101444.97	20000	631.23
Gaussian approximation	3342.25	53695.83	20000	16.07
GPCH with Markov random field prior				
Numerical approach	Average ESS	time	n.iter	Effect of sample size
Basic Gibbs sampler	1465.551	1728.09	30000	1.18
Principal components	1590.085	188383.90	10000	118.474
Splitting FBA	6005.618	41440.33	20000	6.9
Splitting CD	2188.159	101454.64	20000	46.365
Gaussian approximation	608.42	191543.58	20000	314.821
Continuous parameter space model with Gaussian process random field prior				
Numerical approach	Average ESS	time	n.iter	Effect of sample size
Basic Gibbs sampler	1359.934	5775.46	30000	4.25
Principal components	1194.135	201400.07	7000	168.66
Splitting CD	1179.331	202496.71	20000	171.70
Gaussian approximation	334.73	39908.45	20000	119.22

but they are not exactly the same. The idea of splitting the log-hazards does not have much effect on the posterior in the GPCH model where the priors are based on parametric model.

We have also illustrated the GPCH model with Markov random field priors. The correlations between the log-hazards in this case were strongly correlated. For instance, in this case, the correlation between the first and second log-hazards was 0.924 whereas in the first and second cases discussed in the GPCH model with priors based on parametric model, it was 0.724 and 0.907 respectively. The covariances between the log-hazards were different from the GPCH model using priors based on parametric model. In this case, the mean and variance of the first log-hazard are -6 and 2.422 respectively whereas it was -6 and 1.047 in the first case and -6 and 1.295 in the second case of the GPCH model using priors based on parametric model. The priors are different in the terms of the covariance structures. Table 8.16 shows that the time per effective sample is 1.18 seconds. We would want to see how much effect the differences in the covariance structure has affected the posterior summaries of log-hazards in the different choices of priors. The posterior summaries in the case where we used the Markov random field prior are different from any of the two methods in the case of the GPCH model with priors based on parametric

model. The posterior summaries are very much concentrated on the prior. The approach using splitting the log-hazards involves introducing frailties which gives a different model. The posterior summaries are therefore not exactly the same but, in fact, they are not very different. The idea of splitting the log-hazards does not have much effect on the posterior in the GPCH model when we had the Markov random field priors.

We have illustrated the continuous parameter space model using the Gaussian process prior. The correlations between the log-hazards in this case were strongly correlated. For instance, the correlation between the first and second log-hazards was 0.899 and the covariance was 0.809. Table 8.16 also shows that the time per effective sample is 4.25 seconds using this approach. Again, in the approach using splitting the log-hazards in the continuous parameter space is a different model and we sampled the log-hazards either using Cholesky decomposition, the posterior summaries are not exactly the same.

We might want to compare the strength of spatial correlation in the GPCH and continuous parameter space model. In the GPCH model, over a certain area (cell) of the parameter space, the log-hazard is exactly the same and the correlation is 1 whereas the correlation of log-hazards in neighbouring cells are less than 1. The GPCH model makes the hazard constant over a region of space in the continuous space model. We would want to find out what the correlation of the log-hazards between two patients that lie in the same cell in the GPCH model would be in the continuous parameter space model. We would illustrate this using the leukaemia data set. Suppose we consider two different covariate profiles p and p' in the continuous parameter space model, with covariate vectors $\underline{x}_p = (61, 1, 13.3, -1.96)$ and $\underline{x}_{p'} = (76, 1, 450.0, -3.39)$ for the covariates age, sex, Wbc and Depsc respectively in the same time period. These two covariate profiles were categorised into cells $\underline{x}_{p,j} = (2, 2, 3, 2, 1)$ and $\underline{x}_{p',j} = (4, 1, 4, 1, 1)$ respectively in the GPCH model. The correlation between the log-hazards in the continuous parameter space model was 0.86 whereas it was 0.736 in the GPCH model. We might also want to consider another example where the covariate profiles in the continuous parameter space model were categorised in the same cell in the GPCH model. If we suppose two covariate profiles in the continuous parameter space model, with covariate vectors $\underline{x}_p = (79, 1, 500.0, -1.40)$ and $\underline{x}_{p'} = (83, 1, 160.0, -2.59)$. These two covariate profiles were categorised into the same cell given by $\underline{x}_{p,j} = \underline{x}_{p',j} = (4, 2, 4, 2, 1)$ in the GPCH model and hence have a correlation of 1. These covariate profiles had a correlation of 0.89 in the continuous parameter space model. There are slight differences in the correlation between log-hazards in the continuous parameter space model and the GPCH model.

In our illustrations, we have quite a few patients with the same covariate profile in the GPCH model whereas we only ever had one patient in the continuous parameter space model and so if it was not for the covariances in the Gaussian process prior in this model, we really would not have learnt very much from one patient.

It is important to discuss the strength and weakness of all methods and alternative numerical methods discussed in this research. The numerical methods are simple and efficient. We will discuss some advantages and disadvantages of the numerical methods used. One major advantage to all methods used is that we had the possibility of forming and combining the marginal variances of the covariate profiles to reflect a reasonable assessment of our prior uncertainties.

We discuss some advantages and disadvantages of the generalised piecewise constant hazard model. One advantage of the GPCH model is that the corresponding hazard is constant within cells. We could also decide on the choice of the number of cutpoints and hence, avoid having too many parameters in the model. Another advantage of the GPCH model is that we try to exploit the Markov property within cells by using the Gibbs sampling where we can sample a particular log-hazard corresponding to a cell by conditioning on the neighbours. We have the advantage of using the forward backward algorithm in the GPCH model since we can simulate the hidden Markov model parameters from the posterior distribution given observed data.

One disadvantage of the GPCH model is the problem of the choice of the number of categories of the ordinal covariates. In our illustration, we have partitioned each continuous covariate into four groups and the lower quartile, median and upper quartile were used as the cut points of the covariates. We could also decide to partition the range of the observations of the covariate of interest into four intervals of equal width and we use these groups as the categories of the ordinal covariates. The problem with the choice of categories is that cells that are neighbours with each other might not be neighbours in the other choice of categories. Hence, the correlations between cells will be different in various choices of categories. We might give varying results and hence misleading conclusions. We might have to fit too many parameters if we have too many covariate and categories since the number of cells depend on the number of ordinal categories like in the case of the SNLG data set. Hence, the GPCH model is unrealistic when we have many cells and MCMC is not feasible. The disadvantage is that the GPCH model can be illustrated using only a data set with smaller number of covariates. Another disadvantage of the GPCH model where we use the Markov random field prior is that, even though we try to exploit the Markov property and the log-hazard in a corresponding cell is conditionally independent given the neighbours, we sometimes find that most of the neighbours are not available to depend upon.

The advantage of sampling the principal components is that they are independent in their prior and we only need fewer iterations because the principal components are closer to being independent. The disadvantage of sampling from the principal components one at a time is that every time a new value for a particular principal component is proposed, the values of all the log-hazards change and the likelihood calculation will involve all the

patients at all times. The simulations will spend more time in the calculations involved in each step.

A general advantage to both the GPCH models is the consideration of splitting the log-hazard corresponding to a covariate profile into two parts and when sampling only the part that carries the dependence, this has a multivariate normal distribution which is easier to sample as a block. The weakness of this method is that we have separate frailties at each time intervals in the GPCH model. We have illustrated this approach using the case where each patient has a different frailty in each time interval. The weakness of this method is that we have separate frailties at each time intervals. The method doesn't induce correlation between time intervals for an individual patient as a frailty should and we might probably learn less about the effects of the covariates.

We will also discuss some advantages and disadvantages of the continuous parameter space. An advantage in the continuous parameter space model is that it acts as an alternative to the GPCH model where the log-hazards for the patients that are closer together in covariates will be closely correlated and we are not affected by the problem of most of the neighbours being missing which occurs in the GPCH model. Another advantage of the continuous parameter space model is that we overcome the problem of having too many cells in the GPCH model but we have the log-hazards at the observed patient-points in the covariate space and it is possible that two patients share the same covariate vector since covariates are only recorded to a small number of significant digits. We also have the advantage of giving the covariates the same scales so that one unit of all the covariates count equally.

The disadvantage of the continuous parameter space model is that we lose the advantage of the Markov property (in the prior). It was impossible to use the forward backward algorithm in this case. Another disadvantage is that we could only have a maximum of as many log-hazard values as the number of patients and hence we have not readily made a provision for the log-hazards of patients not represented in the data. Predictions at other points can be made but require a bit of computation (see Section 8.14).

8.13 Prognostic index for the generalised piecewise constant hazard model

The prognostic index will no longer be obvious when we do not have the proportional hazard model or accelerated life model as in the case of the piecewise constant hazard model. We could also use either the predictive median survival time or the survival probability at a fixed time as the prognostic index as discussed in Section 5.9. In the GPCH model, we have the log-hazard corresponding to every cell and hence, we have the prognostic index for patients with that covariate vector.

We have a function to pick out the corresponding prognostic index using Appendix A.5.19. For instance, a patient with the covariate vector $\underline{x} = (61, 0, 13.3, -1.96)^T$ for the covariates age, sex, WBC, depsc respectively and will have an index of 47 in the first time interval. This will mean that the patient has an index of 47 on a scale from 0 to 100 and this is a low risk value.

8.14 Prognostic index for the continuous parameter space model

We have a similar function given in Appendix A.5.20 like in the case of the GPCH model which can be used as in the case of the GPCH model to select the corresponding linear predictor for a given covariate vector. For instance, a patient with the covariate vector $\underline{x} = (72, 1, 4, 4, 2, 123, 1.9, 2, 2, 2, 1, 1, 2, 2)^T$ for the covariates age, sex, stage, ECOG, LDH, HB, WBC, albumin, AP, urea, extranod, bulk, marrow and bsy respectively will have a log-hazard value of 0.604 and an index of 95. This will mean that the patient has an index of 95 on a scale from 0 to 100 and this is a great risk value.

We would want to think of the prognostic index of a new patient with a covariate profile which has not previously been observed. We calculate the prognostic index for a new patient with a given covariate profile by working out the conditional mean of the log-hazard of this patient given that the log-hazards of all other patients are equal to their posterior means since the prognostic index is based on the mean of the log-hazards and the conditional mean is linear in the other log-hazards.

8.15 Summary

In this chapter, we applied our methodology for relaxing the proportional hazard model to the example data sets. We discussed the advantages and disadvantages of the methods and numerical approaches to improving mixing in MCMC. We have also written functions to calculate the prognostic indices of patients using any of the methods.

The GPCH model can be used when we have very few covariates so that we do not fit a model with so many parameters and many combinations of covariates. The continuous parameter space model can be used when we have any number of covariates as it is possible that some individuals share the same covariate vector and hence the number of parameters could be less than the number of individuals. Further comments on model choice will be made in Section 9.3.

Chapter 9

Conclusion and Future Work

9.1 Summary of thesis

This thesis is concerned with approaches to flexible Bayesian modelling of covariate effects in survival. The ultimate goal of the research was to develop a methodology that can be applied to introducing flexibility into the survival model in which we can relax the form of the relationship between the values of the covariates and the hazards. It is reasonable to capture the necessary features of the data and develop an approach which is flexible and reasonably easily implemented in a standard software. In this research, we presented two fully Bayesian approaches for flexible Bayesian modelling of covariate effects in survival with options of using suitable priors using MCMC simulation for inference. These models were the generalised piecewise constant hazard and the continuous parameter space models. We assessed the performances of the MCMC simulation and several alternative numerical approaches were explored to improve mixing if we were hampered by slow mixing.

In Chapter 2, we introduced the two example data sets that were used as illustrations in this research. In Chapter 3 and 4, we discussed the basic concepts of survival analysis and Bayesian inference respectively. In Chapter 4, we discussed how best we can build good structures of prior distributions depending on the type of variables. We explained exchangeability and how we reparameterise the parameters of the model by imposing constraints on them to solve the problem of non-identifiability in MCMC. The parameters are constrained to sum to zero to avoid over parameterisation. We built a set of ordered random variables with a zero-sum constraint where the form of the correlation matrix depended on the ordering of the categories of the parameter. We constructed the random variables in such a way that neighbouring categories were more positively correlated than categories further apart.

In Chapter 5, we applied Bayesian inference to the analysis of survival. We constructed

prior distributions for the coefficients of the linear predictor and the Weibull shape parameter with illustrations. We illustrated a parametric approach to survival modelling using a Weibull distribution. The proportional hazard model was used to relate the covariates to the survival distribution. Posterior distributions for the covariate effects were evaluated using RJAGS software. We also wrote a function to calculate the prognostic index using the Weibull distribution using R.

We relaxed the form of the baseline hazard and the assumption of a proportional hazard model using the piecewise constant hazard model. We have also illustrated modelling in survival using a piecewise constant hazard model where we discretised the time. The cut points were chosen by supposing that we would expect equal portions of the event in each time interval. We fitted two versions of the model, one without and one with log-normal frailties. We demonstrated time dependent effects of covariates using both example data set. Again, the posterior distributions for the covariate effects were evaluated using RJAGS software. In some cases, the covariate effects either decreases or increases from one interval to another. The age effect decreases from one interval to another while the sex effects increases from one interval to another in the SNLG data set (see Table 5.5). The age effect decreases and later slightly increases from one interval to another and sex effect decreases over time in the leukaemia data set (see Table 5.6). The dependence on time was not obvious in some cases when frailty was included in the model. Sometimes, introducing frailty in the model reduces the apparent dependent covariate effect. We illustrated using both the predictive median survival time and fixed time survival probability to calculate the prognostic indices of a piecewise constant hazard model. There was a close relationship between the predictive median survival time and fixed time survival probability. The fixed time survival probability is simpler and can act as an index but has the problem of the choice of a time which is most informative for all individuals.

We proposed two methods of relaxing the form of the relationship between the covariates values and the hazards using the generalised piecewise constant hazard and the continuous parameter space model in Chapter 6 and 7 respectively. In Chapter 6, we discussed the first model, the generalised piecewise constant hazard model where the covariates were made ordinal and we had a finite number of possible covariate profiles. We would have a “cell” when we combine the covariate profile and the time interval. We first used a prior distribution for the log-hazards of the covariate profiles which was based on the parametric model which allowed for main and interaction effects in the first time interval. We had two forms of construction of the covariance matrices. We then considered the dependence between parameters in the different time intervals by supposing a first order autoregressive model by choosing an autoregressive parameter which gives the strength of relationships to the log-hazards. The posterior summaries using both forms of construction of covariance matrix were quite close.

We also used a Markov random field prior in the GPCH model. We constructed the covariance matrix of the parameters based on a correlation matrix which was also based on the distances between the log-hazards corresponding to the cells. We discussed the inclusion of frailties in the model. We suggested and gave illustrations on some numerical approaches which will help in the practicalities involved in this model. We also discussed overcoming these computational difficulties using some numerical approaches. These numerical approaches include sampling the principal components of the log-hazards, Gaussian approximation of the fcd of the log-hazards and splitting the log-hazards into two parts and sampling the part without the log-frailties using block sampling such as the forward backward algorithm.

The second method used was the continuous parameter space model using the Gaussian process prior. The generalised piecewise hazard model was unrealistic with cases where we have too many covariates and MCMC would involve a large number of parameters. We proposed this method in which we only update the log-hazards for covariate profiles where there were observed patients. Again, we gave a correlation which was based on a distance measure in the covariate space of the Gaussian process prior of the log-hazards. Some numerical approaches were suggested which helped in the practicalities involved in this model. These numerical approaches include sampling the principal components of the log-hazards, Gaussian approximation of the joint fcd and splitting the log-hazards into two parts and sampling the part without the log-frailties using block sampling using the Cholesky decomposition.

We have used two different types of prior in the GPCH model and we have different posterior summaries. Hence, the posterior summaries are very much concentrated on the prior. From Table 8.16, we found out that the basic Gibbs sampler using the Markov random field priors has the smallest time per effective sample and hence we suggest that it is the best approach to this method. In the case of the continuous parameter space model, the basic Gibbs sampler using the Gaussian process prior also has the smallest time per effective sample and thus we suggest that it is also the best approach in this case.

9.2 A review of objectives of Research

At the beginning of the thesis we stated that the aims of the research were to find a methodology in which we will

1. provide a Bayesian framework for flexibility in incorporating covariates into the model.
2. build a model where there is no restriction in the way that the covariates enter the model.

3. develop flexible models for covariate effects that could greatly improve the accuracy of predicting the survival.

We showed in Chapter 5 how the usual proportional hazard model restricts the hazard ratio to be linear in the covariates. We have suggested two approaches to flexibility in Chapter 6 and 7 which satisfy all criteria set out above and offer adequate flexibility for the model. There were no restriction in the way in which the covariates enter the models. Techniques developed in this research can be applied in various disciplines.

At the beginning of the thesis we also set out for the following questions

1. Can we fit survival models and make useful inferences without specifying the form of the dependence on the covariates?
2. Can we, at the same time, not specify the form of the baseline hazard?
3. Can we allow non-proportionality of hazards?
4. What are the practicalities involved and how can we overcome computational difficulties?
5. How do we specify meaningful prior distributions?
6. How can we construct a prognostic index?

In Chapter 6 and 7 we showed that there was no specific form of dependence of the hazard function on the covariates. In the case of the generalised piecewise constant hazard model, the prior was based on the parametric model but the form of the dependence on the covariates was not specified. In the illustration using the continuous parameter space model with a Gaussian process prior, we supposed a Weibull life time but there were no restrictions in the relationship between the covariates and the hazard function. This satisfies question (1).

In Chapter 5, we illustrated the modelling in survival using the proportional hazard model where we specified the form of the baseline hazard. In the same Chapter, we relaxed this by having a non-proportional hazard model using the piecewise constant hazard model where the form of the baseline hazard is not specified. The generalised piecewise constant hazard model is a type of the piecewise constant hazard model and hence, the form of the baseline hazard is not specified. In the continuous parameter space model using the Gaussian process prior, the form of the baseline need not be specified. This satisfies question (2).

In Chapter 5, we illustrated survival modelling with piecewise constant hazard model. We relaxed the assumption of proportionality by allowing for non-proportional hazard. The generalised piecewise hazard model allows for non-proportionality of hazards. In the

case of the continuous parameter space model, we can also allow for non-proportionality of hazards by having the form of the baseline as a piecewise constant hazard model. This allows for the discretisation of time. This satisfies question (3).

The main focus of this thesis is to introduce flexibility into the model by relaxing the form of the relationship between the covariates values and the hazards. Our intentions are to remove the usual assumption of parametric forms made by survival analysts. We have sufficient data to allow the model to adapt to the true form of relationship and not being restricted to a particular form such as the assumption of linearity. Our approach to inference is Bayesian. We have the advantage of choosing a suitable prior distribution which allows relaxation of the assumption of a parametric form of the relationship while it imposes enough structure to exploit the information of the data sets. The prior distribution specifies the correlations between the log-hazards for the neighbouring covariate profiles. We illustrated the generalised piecewise constant hazard model using two forms of prior distribution. This satisfies question (5).

In this case of the GPCH model with priors based on a parametric model, the structure of the prior distribution of the covariate profiles allows for main and interactions effects and the neighbouring time intervals are correlated in their prior. In this thesis, we gave a stationary first order autoregressive process prior for the neighbouring time intervals. In this approach, updating was done one at a time using the Metropolis-Hastings algorithm. We suggested some numerical approaches to improving mixing. In an approach to this, we updated the principal components of the log-hazards one at a time also using the Metropolis-Hastings algorithm.

The other form of prior was the Markov random field prior. In this case, each covariate profile is associated with a log-hazard which corresponds to a cell. We allow the log-hazards corresponding to the cells to depend on the neighbouring covariate profiles.

We have also illustrated using the continuous parameter space model using the Gaussian process prior. In this model, we could have the form of the baseline hazard as a piecewise constant hazard which satisfies (3) but we have illustrated using a Weibull lifetime distribution. In this model, we allowed for the possibility of having a number of distinct covariate profiles which could be less than the number of patients. We increase the number of parameters in the model and we have exactly one parameter for each covariate vector in order that we do not have the constraints that are implied by a parametric model. Each covariate profile is associated with a log-hazard and the log-hazards are correlated based on a distance measure in the prior satisfying question (5).

We have illustrated both methods of introducing flexibility to the model and we discussed some alternative numerical approaches to modelling. We have discussed the advantages and disadvantages using the approaches. We have discussed how we overcome some computational difficulties in Section 8.12. This satisfies question (4).

We have also illustrated how we construct the prognostic index using any of the approaches to introducing flexibility in the model in Section 8.13 and 8.14. This satisfies question (6).

The posterior summaries in the cases using priors based on parametric model and the Markov random field prior showed differences in the log-hazards for the covariate profiles. The effects of the covariates were obvious. The approach using the splitting of the log-hazards which involves introducing frailties gives a different model and the summaries are different. The posterior summaries are very much concentrated on the prior. The structure of the covariance of the different choices of prior has an effect on the posterior summaries. Both methods are good and provided good predictions. In practical issues, the generalised piecewise constant hazard model is often faced with the problem of having many covariates and some other disadvantages discussed in Section 8.12. In a sense, the GPCH model is an approximation because we have assumed that the log-hazard is constant in a cell. For this reason, the continuous parameter space can be used as an alternative.

In general, relaxing the form in which the covariates enter the survival model reveals the true behaviour of the hazard function and this may help to reveal the covariate effects on the survival. The understanding of the behaviour of the hazard function may also help to reveal time related effects on hazard. The flexible models for covariate effects discussed in this research could greatly improve the accuracy of predicting the survival. The approaches are useful for health care practitioners. It will help them provide the prognosis for a patient. We suspect that flexible modelling of the covariate effects will provide for better survival estimation.

We have also attached details about the functions used as illustrations in the Appendix.

9.3 Model choice

Like all modelling approaches, the flexible modelling of covariate effects in survival using either the GPCH or continuous parameter space model has several strengths and a few weaknesses. The strengths include not having to make restrictions about the way that the survival depends on the covariates, exploring the true form of the relationship and maybe capturing complex relationship between the covariates and survival, the possibility of reflecting a reasonable assessment of our prior uncertainties by making the log-hazards of individuals that are closer together in covariate-space more strongly correlated and the clarity of the results. Given enough data, we will identify some relationships even if we did not expect the relationships. The weaknesses include the computational cost (length of simulation runs) and practical limits on the size and complexity of the model. One of the weaknesses of the GPCH model, is the difficulty of application when there are a large number of covariates and hence, we fit too many parameters. In this case, the continuous parameter space model was used to overcome this problem and the log-hazards are sampled

only at the observed patient points in the covariate space.

Using a model that has a complicated relationship and allows much more flexibility like our models, might impose less structure and cause a simple relationship to be less clear. However, we do not have to use a model that biases things in favour of a simple model (see Section 5.6) when we do not really know if there is a simple relationship or not. On the other hand, we might want to be careful about the construction of priors and so we will want to construct a prior which, with a small data set, will favour simple relationships. This is the reason that we have constructed one of our priors based on a linear model. By doing this, a lot of the probability is concentrated around this linear model. Then, we have that the small data set would not make much difference in fitting a linear model whereas in large data sets, the data will overcome the prior and allow the relationship to be something different.

We would want to persuade researchers in the field of study to use our complex models rather than the simpler models in the literature particularly with large data sets. It is possible to have a good start with a simpler model and then move to a more complex model as the second step. The computational cost of MCMC sampling for our complex models may be a disincentive to their use. However, widespread use of complex models will be helped by the development of standard software to do this. Our complex model can be made easier to perform since interested researchers can be given higher-level instructions on how to go about it. In our models, the results from the complex model are easier to understand than results from the simpler techniques since the results are directly obtained from the summaries of the posterior. The transparency of the result should allow researchers to think towards using the complex model. Future work may involve applying integrated nested Laplace approximations (INLA) instead of MCMC and this could make the computational time very much less and would also eliminate the problem of convergence.

It is reasonable either to explore new models or investigate new innovative solutions to challenges. Too much repetition does not make way for new possibilities. If only simpler models are used, researchers can become dependent on the simpler models and unquestioning of the assumptions involved. Our model gives researchers a sense of other possibilities of modelling rather than only using the simple models found in literature. It will also give researchers the opportunity to test their thoughts about the subject of interest and they develop intuition for and experience the relationship between using simple and complex models. Finally, one of the most powerful means of persuasion is directly experiencing something in person. Personal experience with complex modelling can do more to convince researchers of its value than having any amount of discussion.

9.4 Future work

This thesis has attempted to provide answers to the questions in Section 1.3. We relaxed the form of the relationship between the covariates values and the hazard function by providing a flexible framework for incorporating the covariates into the model. In addition, there is no restriction in the way the covariates enter the model. We have demonstrated that the generalised piecewise constant hazard and the continuous parameter space models provide flexibility in modelling covariate effects in survival. There is a need to extend these approaches to cases where some of the covariates are missing.

We would suggest some more areas for further work. These areas include flexible modelling when we have more than one event time, K nearest neighbours, Bayesian networks and applications of classification and regression trees to survival analysis (see Section 6.1).

9.4.1 Flexible modelling with more than one event times

An important issue is with when we have more than one event times. We might assume that the people who died have a length of time with no progression and then the disease starts to progress after which the patient eventually dies. It would be more complicated if the patient is treated again and again and then goes to remission again. The events of remission and death could be thought of as not independent because these two events are observed from the same individual. Such correlation can be modelled using random effects which are known as frailties (Henderson & Oman, 1999). We could introduce frailties in the model to account for correlation between these two event times. In survival, frailties could be also be introduced as dependence between two event times for the same individual. Earlier discussions on frailties have been on only one event. It would be necessary to model the development of disease using progression free survival (PFS) (Booth & Eisenhauer, 2012). These models involve the patient being in different states such as remission, progression or death and we model the times of transition between states. Changing from one stage to another at random time points could be treated as Markov chain and hence there is a probability of moving from one stage to another.

9.4.2 Flexible modelling of covariate effects using the idea of K nearest neighbours

Another area for future work could be using the idea of the K nearest neighbours. The interest in K nearest neighbours (KNN) is to use the outcomes values of the K nearest neighbours in covariate space to predict the outcome of an individual. The covariates used could be quantitative or qualitative (ordinal or categorical). The K nearest neighbours (KNN) algorithm is frequently used for classification problems where an individual is classified depending on the class where most of the K nearest neighbours occur. Future

work could be aimed at finding the K nearest neighbours with respect to building a regression (hazard) model since in regression, the prediction for a new individual can be made by taking the average value of the nearest neighbours to the individual.

We recall that in Section 6.9, the definition of the correlation structure of the covariate profiles might be thought to have a link to the K nearest neighbours. However, Prijs *et al.* (2007) presented a K nearest neighbours method in terms of the linear predictor and related a new patient to the histories of similar individual patients with chronic respiratory disease that were treated before by combining the K nearest neighbours methods with Cox regression. Although Prijs *et al.* (2007) has done KNN in survival analysis, to the best of our knowledge there is no Bayesian version of this.

Before applying the KNN method, decisions must be taken on the form of distance measure to be used to find the K nearest neighbours, the number (K) of the nearest neighbours and the form of weighting function to be used on the neighbouring observations. The nearest neighbours of a new patient are selected using a distance measure on the variables. Some common distance measures used are the Euclidean, Mahalanobis, absolute and Minkowski distance measures.

Suppose we let $\underline{x} = (x_1, x_2, \dots, x_S)$ and $\underline{y} = (y_1, y_2, \dots, y_S)$ be two covariate vectors with “ S ” covariates. The Euclidean distance measure is

$$d(\underline{x}, \underline{y}) = \sqrt{\sum_{i=1}^S (x_i - y_i)^2}$$

The absolute distance measure is

$$d(\underline{x}, \underline{y}) = \sum_{i=1}^S |x_i - y_i|$$

The Minkowski distance measure is

$$d(\underline{x}, \underline{y}) = \left(\sum_{i=1}^S |x_i - y_i|^p \right)^{1/p}$$

where p can not be less than 1.

The Mahalanobis distance $d(x, y)$ is

$$d(\underline{x}, \underline{y}) = \sqrt{(\underline{x} - \underline{y})^T V^{-1} (\underline{x} - \underline{y})}$$

where V is the covariance matrix.

If the covariates are continuous, we first standardize each of the covariates so that it has a mean of zero and a variance of 1. This should be done because the covariates are

measured in different units and hence we avoid the influence of the unit of measurement on the distance. The Mahalanobis distance takes care of standardization of the covariates. Yan (2013) presented weighted K nearest neighbour techniques where the distances of the nearest neighbours were taken into account. The idea was that observations within the set of interest which are close to the new observation of interest were given a higher weight than neighbours that were far away from the observation of interest.

We will have to choose the value “K” which is the number of neighbours. The value of K should be supposed moderately large because if the value of “K” is chosen to be too small, there is a risk of overfitting the data because the number of neighbours will be too small and hence the variance will be too big. If chosen to be too large, there is a risk of underfitting. As K gets large the average will get more stable. Maltamo & Kangas (1998) and Haara *et al.* (1997) used KNN based methods and obtained the predicted values for the variables of interest as weighted averages of the values of neighbouring observations. Cucala *et al.* (2009) proposed a Bayesian reassessment technique from a probabilistic model by deriving computational tools for Bayesian inferences on the parameters of the model.

9.4.3 Flexible modelling of covariate effects using Bayesian networks

In the generalised piecewise constant hazard model, we categorised the covariates. A Bayesian network can also work with categorical covariates. In a Bayesian network, we aim at constructing a structure for the network and hence we need to know which variables are conditionally independent given the other variables or else we let all variables depend on each other and we have too many parameters. We could have Dirichlet priors for the parameters of the model. It may not seem reasonable to make these Dirichlet priors independent of each other. We may want these Dirichlet priors to be correlated and hence, we suggest using some hierachial Dirichlet distribution.

Appendix A

Appendix

A.1 General Appendix

A.1.1 Software

BUGS which stands for Bayesian inference using Gibbs sampling (Lunn *et al.*, 2013) can be used to describe arbitrarily complex models using a very limited syntax. The initial step when using BUGS is to construct a directed graphical model. The directed graphical model which is also called directed acyclic graph (DAG) is the basic idea of a graphical representation used to express the relationship between the known and unknown quantities in the model through a series of simple local relationships. DAGs represent all the quantities as nodes and arrows which run into nodes from their direct parents. This decomposition will allow for simplicity as well providing the basis for computation. Nodes in the graph are of three types (see (Lunn *et al.*, 2013)). The first is the *Constant* node which are fixed by the design of the study. The second is the *Stochastic* nodes which are variables that are given in a distribution. Stochastic nodes may be observed (data) or unobserved (parameters) which may be unknown quantities underlying a model, censored observations or missing data. The third is the *Deterministic* nodes which are logical functions of their parent nodes. The strength of the Bayesian graphical modelling techniques of BUGS is in the way it represents the typical complexities of real data. There are two main versions of BUGS namely WinBUGS (Spiegelhalter *et al.*, 1999) and openBUGS (Spiegelhalter *et al.*, 1996). We can run openBUGS directly from R using BRugs package (Lunn *et al.*, 2013)

JAGS which stands for Just another Gibbs sampler (Plummer, 2012) is a program that is designed to work closely with the R language for the analysis of Bayesian models using MCMC. It uses a version of the BUGS model specification language. The RJAGS package can be used from within R to run JAGS (Plummer, 2013).

A.1.2 Using RJAGS for survival data with right censoring

We will discuss how RJAGS works for survival data with right censoring. We suppose that the data contains the event or censoring times t , an indicator variable “status” which is “1” if the event is observed or “0” if the time is censored and the covariates. The data set used by RJAGS contains four vectors as follows

- t : Element i of t is the event time if the event was observed for case i . If case i was censored then element i of t is NA .
- `is.censored`: This is an indicator taking the value 1 if the time is censored and 0 if the event is observed.
- `t.cen`: Element i of `t.cen` is the censoring time for case i if this case was censored. If the event was observed for case i then the “censoring time” must be greater than the observed event time.
- `covariates`: which are the covariates of the data.

We will note that in the model specification (e.g. Appendix A.4.1), `is.censored` is given a special distribution

$$\text{is.censored}[i] \sim \text{dinterval}(t[i], t.cen[i])$$

Experience has shown that it is usually necessary to initialise the missing values of t . This is because the specification of “`is.censored`” declares these cases to be censored. Therefore the unobserved event times must be greater than the censoring times. Allowing JAGS to choose initial values at random is likely to lead to some values for unobserved event times which are less than the censoring times and therefore inconsistent with this. And so, we should initialise the missing values of t to values greater than the corresponding values of `t.cen`.

A.1.3 Setting up the data for piecewise constant hazard model using JAGS

In order to make JAGS fit the piecewise constant hazard model, the data set would contain the covariates of the data and it will be a much bigger data that is created with a special structure. In JAGS, the censoring time (`t.cen`) takes zero when the individual is not censored and takes the actual censored values when the individual is censored in the first time period and takes the conditional survival time that the individual survived to the start of that interval. This conditional survival time will be the result of the first cutpoint subtracted from the second cutpoint. We refer to Plummer (2012) for ways of handling censoring.

A.2 Some common distributions

For reference, we discuss some distributions used as in this project.

A.2.1 Beta distribution

The Beta Distribution ($\text{Beta}(a_1, a_2)$) is a distribution with two proportions restricted to the interval $[0, 1]$. If the parameter θ has a beta distribution then it has a probability density function which is given by

$$f(\theta; a_1, a_2) = \frac{\Gamma(a_1 + a_2)}{\Gamma(a_1)\Gamma(a_2)} \theta^{(a_1-1)}(1 - \theta)^{(a_2-1)} \quad \text{for } 0 \leq \theta \leq 1$$

where $a_1 > 0, a_2 > 0$ and $\Gamma(a)$ is a gamma function.

A.2.2 Gamma distribution

A random variable θ with shape parameter $a_1 > 0$ and scale parameter $a_2 > 0$ has a gamma distribution with density given by

$$f(\theta : a_1, a_2) = \frac{a_2^{a_1}}{\Gamma(a_1)} \theta^{a_1-1} \exp\{-a_2\theta\} \quad .$$

Again, $\Gamma(a)$ is a gamma function.

A.2.3 Dirichlet distribution

The Dirichlet distribution is a distribution for a set of quantities $\theta_1, \theta_2, \dots, \theta_m$ where $\theta_i > 0$ and $\sum_{i=1}^m \theta_i = 1$. The probability density function is

$$f(\theta_1, \dots, \theta_m; a_1, \dots, a_m) = \frac{\Gamma(A)}{\prod_{i=1}^m \Gamma(a_i)} \prod_{i=1}^m \theta_i^{a_i-1}$$

where $A = \sum_{i=1}^m a_i$ and a_1, \dots, a_m are parameters with $a_i > 0$ for $i = 1, \dots, m$. When $m=2$, we obtain a Beta ($\text{Beta}(a_1, a_2)$) distribution.

A.2.4 Binomial distributions

A random variable θ has a binomial distribution with parameter $N_i \in N$ and $p \in [0, 1]$ where N is the population size. The probability of getting exactly a successes in N_i trials has a probability mass function which is given by

$$f(a; N_i, p) = \Pr(\theta = a) = \binom{N_i}{a} p^a (1 - p)^{N_i-a} \quad .$$

The conjugate prior of a binomial observation is a beta distribution.

A.2.5 Multinomial observation

Suppose that we observe X_1, \dots, X_m , we fix $N = \sum_{i=1}^m X_i = 1$ and we have probabilities for these as $\theta_1, \dots, \theta_m$. Then given $\theta = (\theta_1, \dots, \theta_m)^T$, X_1, \dots, X_m has multinomial distribution given by

$$Pr(X_1 = x_1, \dots, X_m = x_m) = \frac{N!}{\prod_{i=1}^m x_i!} \prod_{i=1}^m \theta_i^{x_i} .$$

If $m=2$, we have a binomial distribution. The conjugate prior of multinomial distribution is a Dirichlet distribution.

A.3 Appendix to Chapter 2

A.3.1 Partial display of SNLG data

Table A.1: A Partial display of SNLG data

t	died	age	albumin	ap	bsy ...
0.016438356	1	72	2	2	2 ...
0.250228311	1	63	1	1	2 ...
0.168949772	1	58	1	2	1 ...
2.285844749	0	39	1	1	2 ...

A.3.2 Partial display of leukaemia data

Table A.2: A Partial display of leukaemia data

t	t.cen	age	sex	wbc	depsec
1	0	61	1	13.3	-1.96
1	0	76	1	450.0	-3.39
1	0	74	1	154.0	-4.95
1	0	79	2	500.0	-1.40
1	0	83	2	160.0	-2.59
1	0	81	2	30.4	0.03
1	0	76	1	41.3	3.95
1	0	87	1	280.0	1.91

A.4 Appendix to Chapter 5

A.4.1 Bayesian survival modelling to SNLG data set

```
#Weibull fit of the SNLG data with the complete observed cases

model
{
for (i in 1:636)
{
is.cen[i] ~ dinterval (t[i], t.cen[i])
t[i] ~dweib(alpha,lambda[i])
eta[i]<-beta0 +beta.age*age[i]+beta.hb*hb[i]+beta.wbc*wbc[i]
+beta.sex[sex[i]]+beta.albumin[albumin[i]]+beta.ap[ap[i]]+
beta.urea[urea[i]]+beta.extranod[extranod[i]]
+beta.bulk[bulk[i]]+beta.marrow[marrow[i]]+beta.bsy[bsy[i]]
+ beta.stage[stage[i]]+beta.ecog[ecog[i]]+beta.ldh[ldh[i]]
lambda[i]<-exp(eta[i])
}
delta.ecog[1]~dnorm(0,500)
delta.ecog[2]~dnorm(0,333.3)
delta.ecog[3]~dnorm(0,166.6)
delta.ecog[4]~dnorm(0,50)

beta.ecog[1]<- -delta.ecog[1]    -delta.ecog[2]    -delta.ecog[3]
    -delta.ecog[4]
beta.ecog[2]<-  delta.ecog[1]    -delta.ecog[2]    -delta.ecog[3]
    -delta.ecog[4]
beta.ecog[3]<-                                2*delta.ecog[2]    -delta.ecog[3]
    -delta.ecog[4]
beta.ecog[4]<-                                3*delta.ecog[3]    -delta.ecog[4]
beta.ecog[5]<-                                4*delta.ecog[4]

delta.stage[1]~dnorm(0,333.3)
delta.stage[2]~dnorm(0,166.6)
delta.stage[3]~dnorm(0,50)

beta.stage[1]<- -delta.stage[1]    -delta.stage[2]    -delta.stage[3]
beta.stage[2]<-  delta.stage[1]    -delta.stage[2]    -delta.stage[3]
```

```

beta.stage[3]<-          2* delta.stage[2]-delta.stage[3]
beta.stage[4]<-          3*delta.stage[3]

delta.ldh[1]~dnorm(0,500)
delta.ldh[2]~dnorm(0,166.6)

beta.ldh[1]<- -delta.ldh[1] -delta.ldh[2]
beta.ldh[2]<- delta.ldh[1] -delta.ldh[2]
beta.ldh[3]<-          2*delta.ldh[2]

delta.sex~dnorm(0.05,44.4)
delta.albumin~dnorm(0,33.41)
delta.ap~dnorm(0,33.41)
delta.urea~dnorm(0,33.41)
delta.extranod~dnorm(0,33.41)
delta.bulk~dnorm(0,33.41)
delta.marrow~dnorm(0,33.41)
delta.bsy~dnorm(0,33.41)

beta.sex[1]<-delta.sex
beta.sex[2]<- -delta.sex
beta.albumin[1]<-delta.albumin
beta.albumin[2]<- -delta.albumin
beta.ap[1]<-delta.ap
beta.ap[2]<- -delta.ap
beta.urea[1]<-delta.urea
beta.urea[2]<- -delta.urea
beta.extranod[1]<-delta.extranod
beta.extranod[2]<- -delta.extranod
beta.bulk[1]<-delta.bulk
beta.bulk[2]<- -delta.bulk
beta.marrow[1]<-delta.marrow
beta.marrow[2]<- -delta.marrow
beta.bsy[1]<-delta.bsy
beta.bsy[2]<- -delta.bsy
beta.age~dnorm(0.04,1111.1)
beta.hb~dnorm(0.02,5102.04)
beta.wbc~dnorm(0.08,277.7)

```

```

alpha~dgamma(4,4)
beta0~dnorm(-1.5,6.25)
}

```

A.4.2 Function to calculate the prognosis indices using the SNLG data set

```

1 index<-function()
2 {#source("coefficientsweibnew.txt")
3 beta.stage <-c(-0.048,-0.048,0.037,0.055)
4 beta.ecog <-c(-0.134,0.143,0.148,0.169,-0.041)
5 beta.sex <- c(-0.034,0.034)
6 beta.age <- 0.009
7 beta.albumin <- c(-0.020,0.020)
8 beta.ldh <- c(-0.113, -0.034, 0.147)
9 beta.ap <- c(0.048, -0.048)
10 beta.urea <- c(-0.097 ,0.097)
11 beta.hb <- 0.002
12 beta.wbc <- 0.050
13 beta.extranod <- c(0.073, -0.073)
14 beta.bulk <- c(-0.154,0.154)
15 beta.bsy <- c(-0.044 ,0.044)
16 beta.marrow <- c(-0.171,0.171)
17 mean <- 0.4514261
18 std.dev <- 0.414887
19 ##### AGE
20 write(file="", "Please enter the Age in years of
21 the patient at time of diagnosis.")
22 age<-scan(n=1)
23 age<-age -60
24 ##### SEX
25 write(file="", "Please enter the Sex of the patient.
26 Enter 1 for male or 2 for female.")
27
28 sex<-scan(n=1)
29 # sex<-sex
30 ##### STAGE
31 write(file="", "Please enter the Clinical Stage
32 of the patient (1, 2, 3 or 4).")

```

```
33
34 stage<-scan(n=1)
35 ##### ECOG
36 write(file="", "Please enter the ECOG Performance
37 status for the patient ( 1, 2, 3, 4 or 5).")
38
39 ecog<-scan(n=1)
40 ecog<-ecog
41 ##### LDH
42 write(file="", "Please enter the LDH value
43 for the patient ( 1, 2 or 3).")
44
45 ldh<-scan(n=1)
46 ##### HB
47 write(file="", "Please enter the Haemoglobin
48 (hb) (g/l) measurement for the patient.")
49
50 hb<-scan(n=1)
51 hb=hb-110
52 ##### WBC
53 write(file="", "Please enter the White Blood Cell
54 count for the patient.")
55
56 wbc<-scan(n=1)
57 wbc=wbc-2
58 ##### ALBUMIN
59 write(file="", "Please enter 1 if the Serum
60 Albumin measurement for the patient is normal")
61
62 write(file="", "or 2 if it is abnormal")
63 albumin<-scan(n=1)
64 ##### AP (ALK)
65 write(file="", "Please enter 1 if the Serum Alkaline
66 Phosphatase (ap) measurement for the patient is normal")
67
68 write(file="", "or 2 if it is abnormal")
69 ap<-scan(n=1)
70 ##### UREA
```

```
71 write(file="", "Please enter 1 if the Blood Urea
72 Nitrogen measurement for the patient is normal")
73
74 write(file="", "or 2 if it is abnormal")
75 urea<-scan(n=1)
76 ##### EXTRANOD
77 write(file="", "Is there extranodal disease?
78 Please enter 1 for 'Yes' or 2 for 'No'.")
79
80 ##### BULK
81 write(file="", "Is there bulk disease? Please enter
82 1 for 'Yes' or 2 for 'No'.")
83
84 bulk<-scan(n=1)
85 ##### MARROW
86 write(file="", "Is there evidence of disease in the
87 bone marrow? Please enter 1 for 'Yes' or 2 for 'No'.")
88
89 marrow<-scan(n=1)
90 ##### BSY
91 write(file="", "Are there B-symptoms?
92 Please enter 1 for 'No' or 2 for 'yes'.")
93
94 bsy<-scan(n=1)
95 eta<-beta.age*age+beta.sex[sex]+beta.albumin[albumin]
96 +beta.ap[ap]+beta.urea[urea]+beta.bulk[bulk]+
97 beta.bsy[bsy]+beta.extranod[extranod]
98
99 eta<-eta+beta.marrow[marrow]+beta.hb*(hb)+
100 beta.wbc*(wbc)+beta.ldh[ldh]+beta.stage[stage]+beta.ecog[ecog]
101 ind<-100*pnorm(eta, mean, std.dev)
102 ind<-round(ind)
103 #ind<-c(mu, ind)
104 write(file="", "Index value is")
105 write(ind, file="")
106 write(file="", "The index is on a scale from 0 to 100,
107 Greater index values indicate greater risk.")
```

108 }

A.4.3 Bayesian survival modelling using the piecewise constant hazard model and including frailties using the SNLG data

```

#Piecewise constant hazard model of the non-Hodgkins Lymphoma data
set using 14 covariates including some missing values using JAGS
model

{
for (i in 1:n.obs)
  {
t.obs[i] ~ dexp(lambda.obs[i])

eta.obs[i]<- f[patients.obs[i]] +beta0[perj.obs[i]]+beta.age[perj.obs[i]]*
(age[patients.obs[i]])+beta.hb[perj.obs[i]]*(hb[patients.obs[i]])
+beta.wbc[perj.obs[i]]*(wbc[patients.obs[i]])+
beta.sex[perj.obs[i],sex[patients.obs[i]]]+
beta.albumin[perj.obs[i],albumin[patients.obs[i]]]
+beta.ap[perj.obs[i],ap[patients.obs[i]]]+
beta.urea[perj.obs[i],urea[patients.obs[i]]]+
beta.extranod[perj.obs[i],extranod[patients.obs[i]]]
+beta.bulk[perj.obs[i],bulk[patients.obs[i]]]+
beta.marrow[perj.obs[i],marrow[patients.obs[i]]]
+beta.bsy[perj.obs[i],bsy[patients.obs[i]]]+
beta.stage[perj.obs[i],stage[patients.obs[i]]]
+beta.ecog[perj.obs[i],ecog[patients.obs[i]]]+
beta.ldh[perj.obs[i],ldh[patients.obs[i]]]

lambda.obs[i]<-exp(eta.obs[i])
  }

for (i in 1:n.miss)
  {
is.cen[i] ~ dinterval (t.miss[i], t.cen[i])
t.miss[i] ~ dexp(lambda.miss[i])

eta.miss[i]<- f[patients.miss[i]]+ beta0[perj.miss[i]]

```



```

+beta.age[perj.miss[i]]*(age[patients.miss[i]])+
beta.hb[perj.miss[i]]*(hb[patients.miss[i]])+
beta.wbc[perj.miss[i]]*(wbc[patients.miss[i]])+
beta.sex[perj.miss[i],sex[patients.miss[i]]]+
beta.albumin[perj.miss[i],albumin[patients.miss[i]]]
+beta.ap[perj.miss[i],ap[patients.miss[i]]]+
beta.urea[perj.miss[i],urea[patients.miss[i]]]+
beta.extranod[perj.miss[i],extranod[patients.miss[i]]]+
beta.bulk[perj.miss[i],bulk[patients.miss[i]]]+
beta.marrow[perj.miss[i],marrow[patients.miss[i]]]+
beta.bsy[perj.miss[i],bsy[patients.miss[i]]]+
beta.stage[perj.miss[i],stage[patients.miss[i]]]+
beta.ecog[perj.miss[i],ecog[patients.miss[i]]]+
beta.ldh[perj.miss[i],ldh[patients.miss[i]]]

lambda.miss[i]<-exp(eta.miss[i])
}

for (c in 1:636)
{
f[c]~dnorm(0,tau)
}

tau~dgamma(1.1,0.53)

#Priors for all model parameters

delta.albumin[1]~dnorm(0,p.albumin)
delta.ap[1]~dnorm(0,p.ap)
delta.sex[1]~dnorm(0.05,p.sex)
delta.bsy[1]~dnorm(0,p.bsy)
delta.bulk[1]~dnorm(0,p.bulk)
delta.extranod[1]~dnorm(0,p.extranod)
delta.urea[1]~dnorm(0,p.urea)
delta.marrow[1]~dnorm(0,p.marrow)

beta0[1]~dnorm(-1.5,p0)

```

```

beta.sex[1,1]<- -delta.sex[1]
beta.sex[1,2]<- delta.sex[1]

beta.albumin[1,1]<- -delta.albumin[1]
beta.albumin[1,2]<- delta.albumin[1]

beta.ap[1,1]<- -delta.ap[1]
beta.ap[1,2]<- delta.ap[1]

beta.bsy[1,1]<- -delta.bsy[1]
beta.bsy[1,2]<- delta.bsy[1]

beta.bulk[1,1]<- -delta.bulk[1]
beta.bulk[1,2]<- delta.bulk[1]

beta.extranod[1,1]<- -delta.extranod[1]
beta.extranod[1,2]<- delta.extranod[1]

beta.marrow[1,1]<- -delta.marrow[1]
beta.marrow[1,2]<- delta.marrow[1]

beta.urea[1,1]<- -delta.urea[1]
beta.urea[1,2]<- delta.urea[1]

delta.ecog[1,1]~dnorm(0,p.ecog.d[1])
delta.ecog[1,2]~dnorm(0,p.ecog.d[2])
delta.ecog[1,3]~dnorm(0,p.ecog.d[3])
delta.ecog[1,4]~dnorm(0,p.ecog.d[4])

beta.ecog[1,1]<- -delta.ecog[1,1]   -delta.ecog[1,2]   -delta.ecog[1,3]
  -delta.ecog[1,4]
beta.ecog[1,2]<-  delta.ecog[1,1]   -delta.ecog[1,2]   -delta.ecog[1,3]
  -delta.ecog[1,4]
beta.ecog[1,3]<-                               2* delta.ecog[1,2]   -delta.ecog[1,3]
  -delta.ecog[1,4]
beta.ecog[1,4]<-                               3*delta.ecog[1,3]   -delta.ecog[1,4]
beta.ecog[1,5]<-                               4*delta.ecog[1,4]

```

```
beta.age[1]~dnorm(0.04,p.age)
beta.hb[1]~dnorm(0.02,p.hb)

beta.wbc[1]~dnorm(0.08,p.wbc)

delta.stage[1,1]~dnorm(0,p.stage.d[1])
delta.stage[1,2]~dnorm(0,p.stage.d[2])
delta.stage[1,3]~dnorm(0,p.stage.d[3])

beta.stage[1,1]<- -delta.stage[1,1] -delta.stage[1,2] -delta.stage[1,3]
beta.stage[1,2]<- delta.stage[1,1] -delta.stage[1,2] -delta.stage[1,3]
beta.stage[1,3]<- 2* delta.stage[1,2] -delta.stage[1,3]
beta.stage[1,4]<- 3*delta.stage[1,3]

delta.ldh[1,1]~dnorm(0,p.ldh.d[1])
delta.ldh[1,2]~dnorm(0,p.ldh.d[2])

beta.ldh[1,1]<- -delta.ldh[1,1] -delta.ldh[1,2]
beta.ldh[1,2]<- delta.ldh[1,1] -delta.ldh[1,2]
beta.ldh[1,3]<- 2*delta.ldh[1,2]

p0<-6.25
p.sex<-44.4
p.albumin<-33.41
p.ap<-33.41
p.urea<-33.41
p.extranod<-33.41
p.bulk<-33.41
p.marrows<-33.41
p.bsy<-33.41
p.age<-1111.1
p.hb<-5102.04
p.wbc<-277.7

p.ecog.d[4]<-500
p.ecog.d[3]<-333.3
p.ecog.d[2]<-166.6
p.ecog.d[1]<-50
```

```
p.stage.d[3]<-333.3
p.stage.d[2]<-166.6
p.stage.d[1]<-50

p.ldh.d[2]<-500
p.ldh.d[1]<-166.6

rho<-0.95
rf<-1-rho*rho

p0.e<-p0/rf
p.sex.e<-p.sex/rf
p.albumin.e<-p.albumin/rf
p.ap.e<-p.ap/rf
p.urea.e<-p.urea/rf
p.extranod.e<-p.extranod/rf
p.bulk.e<-p.bulk/rf
p.marrow.e<-p.marrow/rf
p.bsy.e<-p.bsy/rf

p.ecog.e[4]<-p.ecog.d[4]/rf
p.ecog.e[3]<-p.ecog.d[3]/rf
p.ecog.e[2]<-p.ecog.d[2]/rf
p.ecog.e[1]<-p.ecog.d[1]/rf

p.hb.e<-p.hb/rf
p.wbc.e<-p.wbc/rf
p.age.e<-p.age/rf

p.stage.e[3]<-p.stage.d[3]/rf
p.stage.e[2]<-p.stage.d[2]/rf
p.stage.e[1]<-p.stage.d[1]/rf

p.ldh.e[2]<-p.ldh.d[2]/rf
p.ldh.e[1]<-p.ldh.d[1]/rf

for (j in 2:10)
```

```
{
beta0m[j]<- -1.5 + rho*(beta0[j-1]+1.5)
beta0[j]~dnorm(beta0m[j],p0.e)

delta.sexm[j]<- 0.05 + rho*(delta.sex[j-1]-0.05)
delta.sex[j]~dnorm(delta.sexm[j],p.sex.e)
beta.sex[j,1]<- -delta.sex[j]
beta.sex[j,2]<- delta.sex[j]

delta.albuminm[j]<- rho*(delta.albumin[j-1])
delta.albumin[j]~dnorm(delta.albuminm[j],p.albumin.e)
beta.albumin[j,1]<- -delta.albumin[j]
beta.albumin[j,2]<- delta.albumin[j]

delta.apm[j]<- rho*(delta.ap[j-1] )
delta.ap[j]~dnorm(delta.apm[j],p.ap.e)
beta.ap[j,1]<- -delta.ap[j]
beta.ap[j,2]<- delta.ap[j]

delta.bsym[j]<- rho*(delta.bsy[j-1])
delta.bsy[j]~dnorm(delta.bsym[j],p.bsy.e)
beta.bsy[j,1]<- -delta.bsy[j]
beta.bsy[j,2]<- delta.bsy[j]

delta.bulkm[j]<- rho*(delta.bulk[j-1])
delta.bulk[j]~dnorm(delta.bulkm[j],p.bulk.e)

beta.bulk[j,1]<- -delta.bulk[j]
beta.bulk[j,2]<- delta.bulk[j]

delta.extranodm[j]<- rho*(delta.extranod[j-1])
delta.extranod[j]~dnorm(delta.extranodm[j],p.extranod.e)
beta.extranod[j,1]<- -delta.extranod[j]
beta.extranod[j,2]<- delta.extranod[j]

delta.uream[j]<- rho*(delta.urea[j-1])
delta.urea[j]~dnorm(delta.uream[j],p.urea.e)
beta.urea[j,1]<- -delta.urea[j]
```

```

beta.urea[j,2]<- delta.urea[j]

delta.marrown[j]<- rho*(delta.marrown[j-1])
delta.marrown[j]~dnorm(delta.marrown[j],p.marrown.e)
beta.marrown[j,1]<- -delta.marrown[j]
beta.marrown[j,2]<- delta.marrown[j]

delta.ecogm[j,1]<- rho*(delta.ecog[j-1,1])
delta.ecog[j,1]~dnorm(delta.ecogm[j,1],p.ecog.e[1])
delta.ecogm[j,2]<- rho*(delta.ecog[j-1,2])
delta.ecog[j,2]~dnorm(delta.ecogm[j,2],p.ecog.e[2])
delta.ecogm[j,3]<- rho*(delta.ecog[j-1,3])
delta.ecog[j,3]~dnorm(delta.ecogm[j,3],p.ecog.e[3])
delta.ecogm[j,4]<-rho*(delta.ecog[j-1,4])
delta.ecog[j,4]~dnorm(delta.ecogm[j,4],p.ecog.e[4])

beta.ecog[j,1]<- -delta.ecog[j,1] -delta.ecog[j,2] -delta.ecog[j,3]
  -delta.ecog[j,4]
beta.ecog[j,2]<- delta.ecog[j,1] -delta.ecog[j,2] -delta.ecog[j,3]
  -delta.ecog[j,4]
beta.ecog[j,3]<-
  2* delta.ecog[j,2] -delta.ecog[j,3]
  -delta.ecog[j,4]
beta.ecog[j,4]<-
  3*delta.ecog[j,3] -delta.ecog[j,4]
beta.ecog[j,5]<-
  4*delta.ecog[j,4]

beta.agem[j]<- 0.04 + rho*(beta.age[j-1]-0.04)
beta.age[j]~dnorm(beta.agem[j],p.age.e)

beta.hbm[j]<- 0.02 + rho*(beta.hb[j-1]-0.02)
beta.hb[j]~dnorm(beta.hbm[j],p.hb.e)

beta.wbcm[j]<- 0.08 + rho*(beta.wbc[j-1]-0.08)
beta.wbc[j]~dnorm(beta.wbcm[j],p.wbc.e)

delta.stagem[j,1]<-rho*(delta.stage[j-1,1])
delta.stage[j,1]~dnorm(delta.stagem[j,1],p.stage.e[1])

delta.stagem[j,2]<- rho*(delta.stage[j-1,2])

```

```

delta.stage[j,2]~dnorm(delta.stagem[j,2],p.stage.e[2])

delta.stagem[j,3]<- rho*(delta.stage[j-1,3])
delta.stage[j,3]~dnorm(delta.stagem[j,3],p.stage.e[3])

beta.stage[j,1]<- -delta.stage[j,1] -delta.stage[j,2] -delta.stage[j,3]
beta.stage[j,2]<- delta.stage[j,1] -delta.stage[j,2] -delta.stage[j,3]
beta.stage[j,3]<-                2* delta.stage[j,2] -delta.stage[j,3]
beta.stage[j,4]<-                3*delta.stage[j,3]

delta.ldhm[j,1]<- rho*(delta.ldh[j-1,1])
delta.ldh[j,1]~dnorm(delta.ldhm[j,1],p.ldh.e[1])
delta.ldhm[j,2]<- rho*(delta.ldh[j-1,2])
delta.ldh[j,2]~dnorm(delta.ldhm[j,2],p.ldh.e[2])

beta.ldh[j,1]<- -delta.ldh[j,1] -delta.ldh[j,2]
beta.ldh[j,2]<- delta.ldh[j,1] -delta.ldh[j,2]
beta.ldh[j,3]<-                2*delta.ldh[j,2]
    }
}

```

A.4.4 Bayesian survival modelling using the piecewise constant hazard model and including frailties using the leukaemia data

```

model
{
for (i in 1:n.obs)
  {
t.obs[i] ~ dexp(lambda.obs[i])
eta.obs[i]<-beta0[perj.obs[i]]+beta.age[perj.obs[i]]*age[patients.obs[i]]+
beta.sex[perj.obs[i],sex[patients.obs[i]]]+beta.wbc[perj.obs[i]]
*wbc[patients.obs[i]]+beta.depsc[perj.obs[i]]*depsec[patients.obs[i]]
lambda.obs[i]<-exp(eta.obs[i])
}
}

for (i in 1:n.miss)
  {is.cen[i] ~ dinterval (t.miss[i], t.cen[i])
}

```

```

t.miss[i] ~ dexp(lambda.miss[i])
eta.miss[i]<-beta0[perj.miss[i]]+beta.age[perj.miss[i]]*age[patients.miss[i]] +
beta.sex[perj.miss[i],sex[patients.miss[i]]]+beta.wbc[perj.miss[i]]
*wbc[patients.miss[i]]+beta.depsc[perj.miss[i]]*depsec[patients.miss[i]]
lambda.miss[i]<-exp(eta.miss[i])
}
#Priors:

delta.sex[1]~dnorm(0.05,p.sex)

beta0[1]~dnorm(-6.9,p0)
beta.age[1]~dnorm(0.04,p.age)

beta.sex[1,1]<- -delta.sex[1]
beta.sex[1,2]<- delta.sex[1]
beta.wbc[1]~dnorm(0.08,p.wbc)
beta.depsc[1]~dnorm(0.12,p.depsc)

p0<-69.4
p.age<-1111.1
p.sex<-44.4
p.wbc<-33.41
p.depsc<-82.6

rho<-0.95
rf<-1-rho*rho

p0.e<-p0/rf
p.age.e<-p.age/rf
p.sex.e<-p.sex/rf
p.wbc.e<-p.wbc/rf
p.depsc.e<-p.depsc/rf

for (j in 2:10)
{
beta0m[j]<- -6.9 + rho*(beta0[j-1]+6.9)
beta0[j]~dnorm(beta0m[j],p0.e)
}

```



```

beta.agem[j]<-0.04 + rho*(beta.age[j-1]-0.04)
beta.age[j]~dnorm(beta.agem[j],p.age.e)

delta.sexm[j]<- 0.05 + rho*(delta.sex[j-1]-0.05)
delta.sex[j]~dnorm(delta.sexm[j],p.sex.e)
beta.sex[j,1]<- -delta.sex[j]
beta.sex[j,2]<- delta.sex[j]

beta.wbcm[j]<-0.08+rho*(beta.wbc[j-1]-0.08)
beta.wbc[j]~dnorm(beta.wbcm[j],p.wbc.e)

beta.depscm[j]<-0.12+rho*(beta.depsc[j-1]-0.12)
beta.depsc[j]~dnorm(beta.depscm[j],p.depsc.e)
}
}

```

A.4.5 An iterative algorithm for finding the predictive median survival time and fixed time survival probability

```

#predictive median survival time and
#predictive survival probability for SNLG data set
cuts=c(0, 0.105, 0.223, 0.357, 0.511, 0.693, 0.916, 1.204, 1.609, 2.303)
snlgsamples<-read.table("thessnlgpch.txt",header=TRUE)
snlgarcom<-read.table("snlgdatacomp.txt",header=TRUE)

#computing differences between cut points
diff<-function(cuts)
{
ncuts<-length(cuts)
cutsd<-numeric(ncuts)
cutsd[1]<-cuts[1]
for(i in 2:ncuts)
{
cutsd[i]<-cuts[i]-cuts[i-1]
}
return (cutsd)
}

```

```
#computing the survival probability
Sprob<-function(int,startprob,starttime,lambda,t)
{
  condprob<-startprob*exp(-lambda*(t-starttime))
  prob<-mean(condprob)
  return(prob)
}

##calculating the predictive survival probabiltiy
predsurvprob=function(cuts,lambda)
{
  ncuts<-length(cuts)
  cutsd<-diff(cuts)
  nsamples<-length(lambda[,1])
  lambdacutsd<-matrix(nrow=nsamples,ncol=ncuts)
  predsuv<-matrix(nrow=nsamples,ncol=ncuts)
  avepredsurv<-matrix(nrow=1,ncol=ncuts)
  # predictive survival prob at cutpoints

  for(j in 1:nsamples)
  {
    lambdacutsd[j,]=lambda[j,] * cutsd
    predsuv[j,]=exp(-(cumsum(lambdacutsd[j,])))
  }

  # average predictive survival prob at cutpoints
  for(r in 1:ncuts)
  {
    avepredsurv[,r]=mean(predsuv[,r])
  }
  return(avepredsurv)
}

#predictive median survival time for patients in SNLG data
PMST<-function(cuts,snlgdcat121,snlgsamples,kmax=10,delta=0.00000001)
{
```

```
ncuts<-length(cuts)
agegp<-snlgdccat121$age
albumingp<-snlgdccat121$albumin
apgp<-snlgdccat121$ap
bsygp<-snlgdccat121$bsy
bulkgp<-snlgdccat121$bulk
ecoggp<-snlgdccat121$ecog
extranodgp<-snlgdccat121$extranod
hbgp<-snlgdccat121$hb
ldhgp<-snlgdccat121$ldh
marrowgp<-snlgdccat121$marrow
sexgp<-snlgdccat121$sex
stagegp<-snlgdccat121$stage
ureagp<-snlgdccat121$urea
wbcgp<-snlgdccat121$wbc
snlgsamples<-as.matrix(snlgsamples)
beta.age<-snlgsamples[,2:11]
nsamples<-length(beta.age[,1])
beta.albumin<-snlgsamples[,12:31]
dim(beta.albumin)<-c(nsamples,10,2)
beta.ap<-snlgsamples[,32:51]
dim(beta.ap)<-c(nsamples,10,2)
beta.bsy<-snlgsamples[,52:71]
dim(beta.bsy)<-c(nsamples,10,2)
beta.bulk<-snlgsamples[,72:91]
dim(beta.bulk)<-c(nsamples,10,2)
beta.ecog<-snlgsamples[,92:141]
dim(beta.ecog)<-c(nsamples,10,5)
beta.extranod<-snlgsamples[,142:161]
dim(beta.extranod)<-c(nsamples,10,2)
beta.hb<-snlgsamples[,162:171]
beta.ldh<-snlgsamples[,172:201]
dim(beta.ldh)<-c(nsamples,10,3)
beta.marrow<-snlgsamples[,202:221]
dim(beta.marrow)<-c(nsamples,10,2)
beta.sex<-snlgsamples[,222:241]
dim(beta.sex)<-c(nsamples,10,2)
beta.stage<-snlgsamples[,242:281]
```

```

dim(beta.stage)<-c(nsamples,10,4)
beta.urea<-snlgsamples[,282:301]
dim(beta.urea)<-c(nsamples,10,2)
beta.wbc<-snlgsamples[,302:311]
beta0<-snlgsamples[,312:321]
eta<-matrix(nrow=nsamples,ncol=ncuts)
for(j in 1:nsamples)
{
for(r in 1:ncuts)
{
eta[j,r]<-beta.age[j,r] *(agegp) +beta.albumin[j,r,albumingp]+beta.ap[j,r,apgp]
+beta.bsy[j,r,bsygp]+beta.bulk[j,r,bulkgp]+beta.ecog[j,r,ecoggp]
+beta.extranod[j,r,extranodgp]+beta.hb[j,r]*(hbgp)+beta.ldh[j,r,ldhgp]
+beta.marrow[j,r,marrowgp]+beta.sex[j,r,sexgp]+beta.stage[j,r,stagegp]
+beta.urea[j,r,ureagp]+beta.wbc[j,r]*(wbcgp) + beta0[j,r]
}
}
lambda<-exp(eta)
avepredsurv<-predsurvprob(cuts,lambda)
int<-sum(avepredsurv>0.5)
ttest<-numeric(kmax)
sprob<-numeric(kmax)
gsprob<-numeric(kmax)
ttest[1]<-cuts[int]
sprob[1]<-avepredsurv[int]
if(min(avepredsurv)>0.5)
{
ttest[2]=2*ttest[1]
sprob[2]=Sprob(int,avepredsurv[int],cuts[int],lambda[,int],ttest[2])
}
else
{
ttest[2]=cuts[int+1]
sprob[2]=avepredsurv[int+1]
}
gsprob[1]<-log(sprob[1])
gsprob[2]<-log(sprob[2])
for(d in 3:kmax)

```

```

{
if((ttest[d-1] - ttest[d-2])^2 >delta)
{
b=(gsprob[d-1] -gsprob[d-2])/(ttest[d-1] -ttest[d-2])
ttest[d]<- ttest[d-2]-(gsprob[d-2]+log (2))/b
sprob[d]=Sprob(int,sprob[d-1],ttest[d-1],lambda[,int],ttest[d])
gsprob[d]=log(sprob[d])
}
else
{
ttest[d]<-ttest[d-1]
sprob[d]=Sprob(int,sprob[d-1],ttest[d-1],lambda[,int],ttest[d])
gsprob[d]=log(sprob[d])
}
t.out<-ttest[kmax]
}
return(t.out)
}

#compute predictive median survival time for all patients in SNLG data
PMSTall<-function(cuts,covariates,parameters,kmax=10,delta=0.000000001)
{
npatients<-length(covariates[,1])
pmst<-numeric(npatients)
for(i in 1:npatients)
{
x<-covariates[i,]
pmst[i]<-PMST(cuts,x,parameters)
}
return(pmst)
}

#computing the predictive survival probability for SNLG data
SPROB<-function(cuts,snlgdcat121,snlgsamples,t)
{
ncuts<-length(cuts)
agegp<-snlgdcat121$age
albumingp<-snlgdcat121$albumin

```

```
apgp<-snlgdcat121$ap
bsygp<-snlgdcat121$bsy
bulkgp<-snlgdcat121$bulk
ecoggp<-snlgdcat121$ecog
extranodgp<-snlgdcat121$extranod
hbgp<-snlgdcat121$hb
ldhgp<-snlgdcat121$ldh
marrowgp<-snlgdcat121$marrow
sexgp<-snlgdcat121$sex
stagegp<-snlgdcat121$stage
ureagp<-snlgdcat121$urea
wbcgp<-snlgdcat121$wbc
patient<-snlgdcat121$patient
snlgsamples<-as.matrix(snlgsamples)
beta.age<-snlgsamples[,2:11]
nsamples<-length(beta.age[,1])
beta.albumin<-snlgsamples[,12:31]
dim(beta.albumin)<-c(nsamples,10,2)
beta.ap<-snlgsamples[,32:51]
dim(beta.ap)<-c(nsamples,10,2)
beta.bsy<-snlgsamples[,52:71]
dim(beta.bsy)<-c(nsamples,10,2)
beta.bulk<-snlgsamples[,72:91]
dim(beta.bulk)<-c(nsamples,10,2)
beta.ecog<-snlgsamples[,92:141]
dim(beta.ecog)<-c(nsamples,10,5)
beta.extranod<-snlgsamples[,142:161]
dim(beta.extranod)<-c(nsamples,10,2)
beta.hb<-snlgsamples[,162:171]
beta.ldh<-snlgsamples[,172:201]
dim(beta.ldh)<-c(nsamples,10,3)
beta.marrow<-snlgsamples[,202:221]
dim(beta.marrow)<-c(nsamples,10,2)
beta.sex<-snlgsamples[,222:241]
dim(beta.sex)<-c(nsamples,10,2)
beta.stage<-snlgsamples[,242:281]
dim(beta.stage)<-c(nsamples,10,4)
beta.urea<-snlgsamples[,282:301]
```

```

dim(beta.urea)<-c(nsamples,10,2)
beta.wbc<-snlgsamples[,302:311]
beta0<-snlgsamples[,312:321]
eta<-matrix(nrow=nsamples,ncol=ncuts)
for(j in 1:nsamples)
{
for(r in 1:ncuts)
{
eta[j,r]<-beta.age[j,r] *agegp +beta.albumin[j,r,albumingp]+
beta.ap[j,r,apgp]+beta.bsy[j,r,bsygp]+beta.bulk[j,r,bulkgp]+
beta.ecog[j,r,ecoggp]+beta.extranod[j,r,extranodgp]+beta.hb[j,r]*hbgp+
beta.ldh[j,r,ldhgp]+beta.marrow[j,r,marrowgp]+beta.sex[j,r,sexgp]
+beta.stage[j,r,stagegp]+beta.urea[j,r,ureagp]+beta.wbc[j,r]*wbcgp + beta0[j,r]
}
}
lambda<-exp(eta)
avepredsurv<-predsurvprob(cuts,lambda)
int<-sum(cuts<t)
sprobb=Sprob(int,avepredsurv[int],cuts[int],lambda[,int],t)
return(sprobb)
}

#predictive survival probability for all patients
Sproball<-function(cuts,covariates,snlgsamples,t)
{
npatients<-length(covariates[,1])
sprobb<-numeric(npatients)
for(i in 1:npatients)
{
x<-covariates[i,]
sprobb[i]<-SPROB(cuts,x,snlgsamples,t)
}
return(sprobb)
}

```

A.5 Appendix to Chapter 8

A.5.1 Functions for sampling the log-hazards using priors based on the parametric model in the GPCH model

```

1 #Generalised piecewise constant hazard model using
2 #priors based on parametric model for the leukaemia data set
3 gpch<-function(niter)
4 {
5 #data values
6 ncuts<-length(cuts)
7 combmatrixx=bigconstruct(nlevel1)
8 designmatrix<-Xmat(combmatrixx)
9 ncomm<-length(combmatrixx[,1])
10 #prior values
11 bmean<-meanvec()#1
12 #####variance vector of parameters#####
13 bvar<-varvec()#2
14 cotry<-read.table("lcovnot.txt",header=TRUE)
15 co<-as.matrix(cotry)
16 Vcov<-bbvar(bvar,co)
17 #mean of eta
18 bmean1<-as.matrix(bmean)
19 priormeaneta<-designmatrix%*%bmean1
20 priormeaneta<-as.numeric(priormeaneta)
21 mu<-wafnm(cuts,priormeaneta)
22 eta<-as.matrix(eta) #1280
23 gcovar<-diag(rep(0.0001,128))#covariance matrix
24 #for interaction effects
25 phimat<-phimatrixfn(phi,cuts) #10x10
26 etavar<-designmatrix%*%Vcov%*%t(designmatrix) #128x128
27 matrixposition<-matrixposition1(etavar,cuts)
28 bigmatt=bigmatt(matrixposition,phimat,etavar,gcovar,cuts)
29 covar=reorderedbigmat(etavar,cuts,bigmatt)
30 ncom<-length(covar[,1])#1280
31 ndtt=ndtfun(cuts,ncom,covarprof,combmatrixx)
32 nd<-ndtt[[2]]
33 nt<-ndtt[[1]]
34 condnvarall<-condnvarall(covar) #function to compute the conditional

```



```

35 #variance of a log hazard given the others for all vector of log hazards
36 sdetaall<-sqrt(condnvarall) #conditional std for all vector of log hazards
37 Bmatrix=Bstoreall(covar)
38 result<-matrix(nrow=niter,ncol=ncom)
39 for(iter in 1:niter)
40 {
41 for(i in 1:ncom)
42 {
43 loglike=loglikew1(eta[i],nd[i],nt[i])
44 othereta<-eta[-i,1]
45 othermu<-mu[-i,1]
46 condmean<-mu[i,1]+Bmatrix[i,]%*(othereta-othermu)
47 logprior<- -1/2*(((eta[i]-condmean)/sdetaall[i])^2)
48 logpost<-logprior+ loglike
49 eta.prop<-rnorm(1,eta[i],sdetaprop)
50 loglike.prop<-loglikew1(eta.prop,nd[i],nt[i])
51 logprior.prop<- -1/2*(((eta.prop-condmean)/sdetaall[i])^2)
52 logpost.prop<-logprior.prop+ loglike.prop
53 A<-logpost.prop-logpost
54 aprob<-min(0,A)
55 u=runif(1)
56 if(log(u)<aprob)
57 {
58 eta[i]<-eta.prop
59 }
60 }
61 result[iter,]<-eta
62 }
63 result
64 }

```

A.5.2 Functions for sampling the log-hazards using priors (using the second form of covariance matrix) based on the parametric model in the GPCH model

We use the function in Appendix A.5.2 but we replace line 28 with

```
bigmatt=bigmat(matrixposition,phimat,etavar,gcovar,cuts,vt)
```

A.5.3 Functions consisting of all functions used in sampling the log-hazards with priors based on the parametric model

```

1 #All functions used in the GPCH for leukaemia
2 #with priors based on parametric model
3 datainclude<-read.table("leukdd.txt",header=TRUE)
4 covarnames<-c("age"," sex", "wbc", "depvc")
5 ncovarnames<-length(covarnames)
6 times<-datainclude[,1:2]
7 covariates<-datainclude[,3:(ncovarnames+2)]
8 covariates=as.matrix(covariates)
9 n=length(covariates[,1])
10 ncov<-c(4,2,4,4,10)
11 cuts<-c(0,52.6,111.6,178.3,255.4,346.6,458.1,601.99,804.7,1151.3)
12 sdetaprop=1
13 sdpc.prop=1
14
15 #function to indicate the covariate profile which the individual belongs
16 profilenum1<-function(covariates){
17 c<-length(covariates[,1])
18 profnum<-c()
19 for(i in 1:c){
20 profnum[i]<-32*(covariates[i,1]-1)+16*(covariates[i,2]-1)+
21 4*(covariates[i,3]-1)+(covariates[i,4]-1)+1
22 }
23 profnum
24 }
25
26 nlevel1<-c(4,2,4,4)
27 profff<-profilenum1(covariates)
28 covarprof<-cbind(datainclude,profff)
29
30 #Creating a vector of means for the covariate effects
31 meanvec<-function()
32 {
33 meanb0<--6.9
34 agem1<-0
35 agem2<-0

```

```
36 agem3<-0
37 sexm<-0
38 wbcm1<-0
39 wbcm2<-0
40 wbcm3<-0
41 depsc1m<-0
42 depsc2m<-0
43 depsc3m<-0
44 meanvec<-c(meanb0,agem1,agem2,agem3,sexm,wbcm1,
45 wbcm2,wbcm3,depsc1m,depsc2m,depsc3m)
46 meanvec
47 }
48
49 #Creating a vector of variances for the covariate effects
50 varvec<-function()
51 {
52 varb0<-0.12
53 agev1<-0.02
54 agev2<-0.006
55 agev3<-0.003
56 sexv<-0.0625
57 wbcv1<-0.006
58 wbcv2<-0.002
59 wbcv3<-0.001
60 depsc1v<-0.0012
61 depsc2v<-0.0004
62 depsc3v<-0.002
63 varvec<-c(varb0,agev1,agev2,agev3,sexv,wbcv1,wbcv2,
64 wbcv3,depsc1v,depsc2v,depsc3v)
65 varvec
66 }
67
68 #Creating the covariance matrix of covariate effects
69 # using the correlation matrix "co"
70 bbvar<-function(var,co){
71 std<-sqrt(var)
72 l<-length(var)
73 bbvar<-matrix(nrow=l, ncol=l)
```

```
74 for(i in 1:l){
75 for(j in 1:l){
76 bbvar[i,j]<-std[i]*std[j]*co[i,j]
77 }
78 }
79 bbvar
80 }
81
82 #Creating contrast of categorical covariate for
83 #1 row in the design matrix
84 mattt<-function(stage,a){
85 xmat<-matrix(nrow=1,ncol=a-1)
86 for(j in 1:a-1){
87 if(stage<=j){
88 xmat[,j]<--1
89 }
90 if(stage==j+1){
91 xmat[,j]<-j
92 }
93 if(stage>j+1){
94 xmat[,j]<-0
95 }
96 }
97 xmat
98 }
99
100 #Creating contrast of categorical covariate for
101 #all rows in the design matrix
102 matttall<-function(stage,a)
103 {
104 n<-length(stage)
105 matttall<-matrix(nrow=n, ncol=a-1)
106 for(i in 1:n){
107 matttall[i,]<-mattt(stage[i],a)
108 }
109 matttall
110 }
111
```

```
112 #Creating the design matrix for the data set
113 Xmat<-function(combmatrix){
114 sex=combmatrix[,2]
115 age=combmatrix[,1]
116 wbc=combmatrix[,3]
117 depssc=combmatrix[,4]
118 n<-length(wbc)
119 X<- matrix(nrow=n,ncol=11)
120 X[,1]<-rep(1,n) #1;intercept
121 X[,2:4]<-matttall(age,4) #2;age
122 X[,5]<-ifelse(sex==1,-1,1) #5:sex
123 X[,6:8]<-matttall(wbc,4) #6-8;wbc
124 X[,9:11]<-matttall(depssc,4) #9-11
125 X<-as.matrix(X)
126 return(X)
127 }
128
129 nlevel<-c(4,2,4,4,10)
130
131 ##creates an array of possible covariate profiles
132 bigconstruct<-function(nlevel){
133 ncovprod<-prod(nlevel)
134 ncover<-length(nlevel)
135 veca<-c()
136 veca[1]<-1
137 for(i in 2:ncovar){
138 veca[i]<-veca[i-1]*nlevel[i-1]
139 }
140 vecb<-c()
141 vecb[ncovar]<-1
142 for(i in (ncovar-1) : 1){
143 vecb[i]<-vecb[i+1]*nlevel[i+1]
144 }
145
146 bigarray<-matrix(nrow=ncovprod,ncol=ncovar)
147 for(i in 1:ncovar){
148 vecc<-1:nlevel[i]
149 z<-rep(vecb[i],nlevel[i])
```

```
150 vecc<-rep(vecc,z)
151 bigarray[,i]<-rep(vecc,veca[i])
152 }
153 #write(z,file="")
154 return(bigarray)
155 }
156
157 combmatrix=bigconstruct(nlevel)
158 combmatrixx=bigconstruct(nlevel1)
159 eta<-rep(-2.0,1280)
160 phi<-0.80
161 #phi<-0.90
162
163 #function to matrix of autoregressive coefficients
164 phimatrixfn<-function(phi,cuts){
165 nperiod<-length(cuts)
166 #A<-rep(0,nperiod^{2})
167 phiall<-matrix(nrow=nperiod,ncol=nperiod)
168 #diag(A)<-rep(1,nperiod)
169 for(i in 1:nperiod)
170 {
171 for(j in 1:nperiod)
172 {
173 k<-abs(i-j)
174 phiall[i,j]<-phi^{k}
175 }
176 }
177 phiall
178 }
179
180 #creating positions for matrix
181 matrixposition1<-function(etavar,cuts){
182 l1<-length(etavar[,1])
183 c1<-length(cuts)
184 c2<-1:c1
185 lc<-l1*c1
186 b1<-l1*(c2-1)+1
187 b2<-l1*c2
```

```
188 n<-length(b1)
189 bb<-list()
190 for( i in 1:n){
191 bb[[i]]<-c(b1[[i]]:b2[[i]])
192 }
193 bb
194 }
195
196 #creating a bigger matrix from a smaller matrix
197 bigmat<-function(nww,phimat,etav,etags,cuts){
198 l1<-length(etav[,1])
199 c1<-length(cuts)
200 lc<-l1*c1
201 mmat<-matrix(nrow=lc,ncol=lc)
202 for(i in 1:c1){
203 for(j in 1:c1){
204 mmat[nww[[i]],nww[[j]]]<-phimat[i,j]*etav
205 mmat[nww[[i]],nww[[i]]]<-etav+etags
206 }
207 }
208 mmat
209 }
210
211 #creating the correct positions for the re-ordered for one point
212 newn1<-function(etavar,cuts,point){
213 l1<-length(etavar[,1])
214 nperiod<-length(cuts)
215 neworder1<-numeric(nperiod)
216 for(i in 1:nperiod){
217 neworder1[i]<-l1*(i-1) +point
218 }
219 neworder1
220}
221
222 #creating the re - ordered matrix
223 reorderedbigmat<-function(etavar,cuts,bigmat){
224 l1<-length(etavar[,1])
225 nperiod<-length(cuts)
```

```
226 lc<-l1*nperiod
227 neworder<-list()
228 rebigmat<-matrix(rep(0,(lc*lc)),nrow=lc,ncol=lc)
229 neworder1<-numeric(lc)
230 neworder[[1]]<-newn1(etavar,cuts,1)
231 neworder2<-neworder[[1]]
232 for(i in 2:l1){
233 neworder[[i]]<-newn1(etavar,cuts,i)
234 neworder2<-c(neworder2,neworder[[i]])
235 }
236 rebigmat[c(1:lc),c(1:lc)]<-bigmat[neworder2,neworder2]
237 rebigmat
238 }
239
240 #function to form the mean
241 wafnm<-function(cuts,mean){
242 n<-length(mean)
243 ncuts<-length(cuts)
244 nc<-n*ncuts
245 wmatv<-list()
246 wmat<-c()
247 for(i in 1:n){
248 wmatv[[i]]<-rep(mean[i],ncuts)
249 wmat<-c(wmat,wmatv[[i]])
250 }
251 wmat<-as.matrix(wmat)
252 wmat
253 }
254
255 ndtfun<-function(cuts,ncom,covarprof,combmatrixx){
256 died<-covarprof$died
257 p.prof<-covarprof$profff
258 ndjp<-rep(0,1280)
259 Tstar<-rep(0,1280)
260 n<-length(covarprof[,1])
261 ngroup<-length(combmatrixx[,1])
262 ncuts<-length(cuts)
263 t<-covarprof$t
```



```
264 tint<-rep(0,n)
265 for(j in 2:ncuts)
266 {
267 tint<-ifelse((t>cuts[j]),j,tint)
268 }
269 for(p in 1:ngroup)
270 {
271 for(j in 1:ncuts)
272 {
273 kstar<-10*(p-1)+j
274
275 for(i in 1:n)
276 {
277 if(p.prof[i]==p)
278 {
279 if(tint[i]==j)
280 {
281 Tstar[kstar]<-Tstar[kstar]+t[i]-cuts[j]
282 ndjp[kstar]<-ndjp[kstar]+died[i]
283 }
284 if(tint[i]>j)
285 {Tstar[kstar]<-Tstar[kstar]+cuts[j+1]-cuts[j]
286 }
287 }
288 }
289 }
290 }
291 vec<-list(Tstar,ndjp)
292 vec
293 }
294
295 #Calculating the maximum likelihood estimates
296 mle2<-function(meaneta,nd,nt)
297 {
298 numprof<-length(nd)
299 mle1<-numeric(numprof)
300 for(i in 1:numprof)
301 {
```

```
302 if(nd[i]>0)
303 {
304 mle1[i]<- log(nd[i] / nt[i])
305 }
306 else
307 {
308 mle1[i]<-meaneta[i]
309 }
310 }
311 mle1
312 }
313
314 #new function for loglikelihood for one patient
315 loglikew1<-function(eta,nd,nt){
316 lambda<-exp(eta)
317 loglikew1=nd*log(lambda) -lambda*nt
318 loglikew1
319 }
320
321 #new function for loglikelihood for all patient
322 loglikewall<-function(eta,nd,nt){
323 ncom<-length(nd)
324 loglikew=0
325 for(p in 1:ncom)
326 {
327 loglikew=loglikew+loglikew1(eta[p],nd[p],nt[p])
328 }
329 loglikew
330 }
331
332 #Approximating the proposal distribution using Newton Raphson's algorithm
333 firstsecondd<-function(mu,nd,nt,prior.prec,mle,tol=0.001,maxiter=7){
334 ndd<-as.matrix(nd)
335 eta<-mle
336 tal<-as.matrix(nt)
337 go<-TRUE
338 for(i in 1:maxiter)
339 {if(go)
```

```

340     {lambda<-exp(eta)
341       lt<-lambda*tal
342       ltn<-as.numeric(lt)
343       em<-eta-mu
344       firstd<-nnd -lt-prior.prec%%em
345       secondd<- -prior.prec - diag(ltn)
346       delta<-solve(secondd,firstd)
347       eta<- eta - delta
348       check<- - t(delta)%*%secondd%%delta
349       go<-(check>tol)
350     }
351   }
352 alll<-list(postmode=eta,postpre= -secondd)
553 alll
554 }
555
556 propdenn<-function(cov,mean,eta){
557   logdetapp<-determinant(cov)
558   logdetapp<-logdetapp$modulus[1]
559   sub<-eta-mean
560   logdetaoldetaprop<--0.5*logdetapp - 0.5*(t(sub)%*%solve(cov,sub))
561   return(logdetaoldetaprop)
562 }
563
564 #function to create fixed terms to be used when
565 #finding matrix (fixed matrix when finding the conditional variance)
566 Bstore<-function(covar,i)
567 {
568   c<-covar[,-i][i,]
569   v<-covar[-i,-i]
570   Bstore<-solve(v,c)
571   Bstore
572 }
573
574 #function to create n by n-1 matrix matrix that stores the constant
575 # part of the formula in finding a conditional mean or variance
576 Bstoreall<-function(covar)
577 {

```

```
578 nlen<-length(covar[,1])
579 Bstoreall=matrix(nrow=nlen,ncol=nlen-1)
580 for(i in 1:nlen)
581 {
582 Bstoreall[i,]<-Bstore(covar,i)
583 }
584 Bstoreall
585 }
586
587 #function to find the conditional variance of one log hazard
588 #given all other log hazards
589 condnvar<-function(covar,i)
590 {
591 c<-covar[,-i][i,]
592 cn<-covar[-i,][,i]
593 v<-covar[-i,-i]
594 condnvar<-covar[i,i]-(solve(v,c))%*%cn
595 condnvar
596 }
597
598 #function to compute the conditional variance of one log hazard
599 #given all other log hazards for all log hazards
600 condnvarall<-function(covar)
601 {
602 nlen<-length(covar[,1])
603 condnvarall=numeric(nlen)
604 for(i in 1:nlen)
605 {
606 condnvarall[i]<-condnvar(covar,i)
607 }
608 condnvarall
609 }
610
611 # grouping the log hazards by time intervals
612 thetatt<-function(theta,nperiod,combmatrix){
613 period<-combmatrix[,5]
614 thetat<-list()
615 for(i in 1:nperiod){
```

```
616 thetat[[i]]<-theta[period==i]
617 thetat[[i]]<-as.matrix(thetat[[i]])
618 }
619 thetat
620 }
621
622 ## grouping the covariances of the log hazards by time intervals
623 covariancet<-function(covar,nperiod,combmatrix){
624 period<-combmatrix[,5]
625 covariancet<-list()
626 for(i in 1:nperiod){
627 covariancet[[i]]<-covar[period==i,period==i]
628 }
629 covariancet
630 }
631
632 ## grouping the means of the log hazards by time intervals
633 meantt<-function(mu,nperiod,combmatrix){
634 period<-combmatrix[,5]
635 meantt<-list()
636 for(i in 1:nperiod){
637 meantt[[i]]<-mu[period==i]
638 meantt[[i]]<-as.matrix(meantt[[i]])
639 }
640 meantt
641 }
642
643 #number of observations in each cell
644 nproff<-function(proff,ncom){
645 np<-c()
646 for(i in 1:ncom){
647 np[i]<-sum(proff==i)
648 }
649 np
650 }
651
652 #number of observations in each profiles by time periods
653 nobst<-function(combnobs,nperiod){
```

```
654 np<-combnobs[,6]
655 period<-combnobs[,5]
656 nobst<-list()
657 for(i in 1:nperiod){
658 nobst[[i]]<-np[period==i]
659 }
660 nobst
661 }
662
663 #number of observations in each profiles by periods put in diagonal matrix
664 matnobst<-function(nobstbytime,nperiod){
665 matnobst<-list()
666 for(i in 1:nperiod){
667 matnobst[[i]]<-diag(nobstbytime[i][[1]])
668 }
669 matnobst
670 }
671
672 #regrouping the log hazards without frailty into a time period
673 singletheta<-function(theta,nperiod,t){
674 ntheta1<-length(theta)
675 ntheta<-c()
676 for(i in 1:ntheta1){
677 ntheta[i]<-t + (i-1)*nperiod
678 }
679 ntheta
680 }
681
682 #regrouping all log hazards without frailty into their respective time period
683 alltheta<-function(theta){
684 nperiod<-length(theta)
685 nalltheta<-list()
686 for(i in 1:nperiod){
687 nalltheta[[i]]<-singletheta(theta[[i]],nperiod,i)
688 }
689 nalltheta
690 }
691
```

```

692 #joining the order of time period as a vector
693 joinc<-function(theta1,theta2){
694 rlist<-c()
695 rlist<-c(theta1,theta2)
696 rlist
697 }
698
699 #joining the order of time period as vector
700 joincall<-function(theta){
701 #theta<-alltheta(theta)
702 n<-length(theta)
703 rlist<-list()
704 rlist<-theta[[1]]
705 for(i in 2:n){
706 rlist<-joinc(rlist,theta[[i]])
707
708 }
709 rlist
710 }

```

A.5.4 Functions for splitting the log-hazards and sampling using forward backward algorithm with priors based on the parametric model in the GPCH model

In this case, we have that each patient has one random frailty in all time intervals. We retain lines 1 – 30 of Appendix A.5.1 and we include the following commands

```

31 eta<-neweta
32 eta<-as.matrix(eta)
33 profbygroup<-profbygroup(proff,covart,ncom)
34 priormeanlogfrailty<-as.matrix(meanlogfrailty)
35 tauf<-1/varlogfrailty
36 thetapriormean<-mu #prior mean of log hazard without frailty
37 thetapriorcovar<-covar #prior covariance matrix
38 #of the log hazard without frailty
39 thetapriorpre<-solve(thetapriorcovar) #prior precision
40 # matrix of log hazard without frailty
41 sdcovf<-sqrt(varlogfrailty)

```

```

42 numinprof<-numm(profbygroup)
43 numinprofmat<-diag(numinprof)
44 theta<-rep(1,ncom)
45 theta<-as.matrix(theta)
46 tauparam<- af + ndata/2
47 nobsinprof<-nproff(proff,1280) #number of
48 #observations in each profiles
49 combnobs<-data.frame(combmatrix,nobsinprof)
50 nobstbytime<-nobst(combnobs,nperiod) #number of
51 #observations in each profiles by time periods
52 matnobstbytime<-matnobst(nobstbytime,nperiod) #number of
53 #observations in each profiles by time periods in a diagonal matrix form
54 #####neighbours#####
55 covtt<-rep(1,128)
56 covtt<-diag(covtt)
57 result<-matrix(nrow=niter,ncol=ncom+1)
58 for(iter in 1:niter)
59 {
60 f <- eta - theta
61 tauf<-rgamma(1,tauparam, bf + 0.5*sum(f^2))
62 result[iter,ncom+1]<-tauf
63 sdcovf<- sqrt(1/tauf)
64 for(i in 1:ncom)
65 {
66 loglike=loglikew1(eta[i],nd[i],nt[i] #loglike contribution
67 # from all individuals in the various covariate profiles
68 logprior<- -0.5*(((eta[i]-theta[i])/sdcovf)^2) #log prior of the log hazard
69 #log prior of the log hazard
70 logpost<-logprior+ loglike
71 eta.prop<-rnorm(1,eta[i],sdetaprop)
72 loglike.prop<-loglikew1(eta.prop,nd[i],nt[i])
73 #loglike contribution from all individuals in the various
74 #covariate profiles using the proposed values of the parameters
75 logprior.prop<- -0.5*(((eta.prop-theta[i])/sdcovf)^2)
76 logpost.prop<-logprior.prop + loglike.prop
77 A<-logpost.prop-logpost
78 aaprob<-min(0,A)
79 u=runif(1)

```

```

80 if(log(u)<aaprob)
81 {
82 eta[i]<-eta.prop
83 }
84 }
85 frailtypriorpre<-tauf*numinprofmat
86 thetapostpre<-thetapriorpre+frailtypriorpre
87 #posterior precision matrix of log hazard
88 thetapostcovar<-solve(thetapostpre) #posterior
89 #covariance matrix
90 #of log hazard
91 thet=thetatt(theta,nperiod,combmatrix) # grouping the
92 #log hazards by time intervals
93 etat=thetatt(eta,nperiod,combmatrix) # grouping the
94 #log hazards including frailty by time intervals
95 covt=covariancet(covar,nperiod,combmatrix) ## grouping the
96 #covariances of the log hazards by time intervals
97 meant=meant(mu,nperiod,combmatrix) ## grouping the means of
98 #the log hazards by time intervals
99 condmean<-list()
100 condvar<-list()
101 priorprecisiontheta<-list()
102 frailtyprecision<-list()
103 postprecisiontheta<-list()
104 priorprecisiontheta[[1]]<-solve(covt[[1]])
105 frailtyprecision[[1]]<-tauf*matnobstbytime[[1]]
106 postprecisiontheta[[1]]<-priorprecisiontheta[[1]]+frailtyprecision[[1]]
107 condvar[[1]]<-solve(postprecisiontheta[[1]])
108 condmean[[1]]<-priorprecisiontheta[[1]]%% meant[[1]]
109 +frailtyprecision[[1]]%%etat[[1]]
110
111 condmean[[1]]<-solve(postprecisiontheta[[1]],condmean[[1]])
112 for(i in 2:nperiod){
113 meant[[i]]<-condmean[[i-1]]
114 covt[[i]]<-condvar[[i-1]]+covtt
115 priorprecisiontheta[[i]]<-solve(covt[[i]])
116 frailtyprecision[[i]]<-tauf*matnobstbytime[[i]]
117 postprecisiontheta[[i]]<-priorprecisiontheta[[i]]+frailtyprecision[[i]]

```

```
118 condvar[[i]]<-solve(postprecisiontheta[[i]])
119 condmean[[i]]<-priorprecisiontheta[[i]]**% meant[[i]]
120 +frailtyprecision[[i]]**%etat[[i]]
121
122 condmean[[i]]<-solve(postprecisiontheta[[i]],condmean[[i]])
123 }
124 theta<-list()
125 CDmatrix<-chol(condvar[[nperiod]]) #cholesky decomposition
126 #of the covariance matrix
127 nc<-length(condvar[[nperiod]][,1])
128 epsilon<-rnorm(nc,0,1) #iid random variables with zero mean and variance 1
129 etatilde<-CDmatrix**%epsilon
130 theta[[nperiod]]<-etatilde+condmean[[nperiod]]
131 condmean1<-list()
132 condvar1<-list()
133 pretheta<-list()
134 p<-nperiod-1
135 for(i in p:1){
136 backprecisiontheta<-solve(covtt)
137 postprethetaback<-postprecisiontheta[[i]]+backprecisiontheta
138 condvar1[[i]]<-solve(postprethetaback)
139 condmean1[[i]]<-postprecisiontheta[[i]]**%condmean[[i]]
140 +backprecisiontheta**%theta[[i+1]]
141
142 condmean1[[i]]<-solve(postprethetaback,condmean1[[i]])
143 CDmatrix<-chol(condvar1[[i]]) #cholesky decomposition
144 #of the covariance matrix
145 epsilon<-rnorm(nc,0,1) #iid random variables with zero mean and variance 1
146 etatilde<-CDmatrix**%epsilon
147 theta[[i]]<-etatilde+condmean1[[i]]
148 }
149 thetaaa<-numeric(ncom)
150 alltheta<-alltheta(theta)
151 orderthetaperiod<-joincall(alltheta)
152 thetaa<-joincall(theta)
153 thetaaa[orderthetaperiod]<-thetaa
154 theta<-thetaaa
155 result[iter,1:ncom]<-theta
```

```

156 }
157 result
158 }

```

A.5.5 Functions for splitting the log-hazards and sampling using Cholesky decomposition

In this case, we have that each patient has one random frailty in all time intervals. We retain lines 1 – 42 of Appendix A.5.4 and we include the following commands

```

43 theta<-rep(1,ncom)
44 eta<-rep(0.5,ncom)
45 theta<-as.matrix(theta)
46 tauparam<- af + ndata/2
47 nobsinprof<-nproff(proff,1280) #number of
48 #observations in each profiles
49 nobsinprofmat<-diag(nobsinprof)
50 ndtt<-ndtfun(cuts,ncom,covarprof,combmatrixx)
51 nd<-ndtt[[2]]
52 nt<-ndtt[[1]]
53 result<-matrix(nrow=niter,ncol=ncom+1)
54 for(iter in 1:niter)
55 {
56 f <- eta - theta
57 tauf<-rgamma(1,tauparam, bf + 0.5*sum(f^2))
58 result[iter,ncom+1]<-tauf
59 sdcovf<- sqrt(1/tauf)
60 for(i in 1:ncom)
61 {
62 loglike=loglikew1(eta[i],nd[i],nt[i])
63 logprior<- -0.5*(((eta[i]-theta[i])/sdcovf)^2)
64 logpost<-logprior+ loglike
65 eta.prop<-rnorm(1,eta[i],sdetaprop)
66 loglike.prop<-loglikew1(eta.prop,nd[i],nt[i])
67 logprior.prop<- -0.5*(((eta.prop-theta[i])/sdcovf)^2)
68 logpost.prop<-logprior.prop + loglike.prop
69 A<-logpost.prop-logpost
70 aaprob<-min(0,A)

```

```

71 u=runif(1)
72 if(log(u)<aaprob)
73 {
74 eta[i]<-eta.prop
75 }
76 }
77 frailtypriorpre<-tauf*nobsinprofmat
78 thetapostpre<-thetapriorpre+frailtypriorpre
79 #of the log hazard
80 thetapostcovar<-solve(thetapostpre) #posterior covariance matrix
81 #of the log hazards
82 CDmatrix<-chol(thetapostcovar) #cholesky decomposition
83 #of the covariance matrix
84 thetapostmean<-thetapriorpre%%thetapriormean+frailtypriorpre%%eta
85 thetapostmean<-solve(thetapostpre,thetapostmean)
86 #posterior mean of log hazard without frailty
87 epsilon<-rnorm(ncom,0,1)
88 etatilde<-CDmatrix%%epsilon
89 theta<-etatilde+thetapostmean #sampled log hazard values
90 result[iter,]<-c(theta,tauf)
91 }
92 result
93 }

```

A.5.6 Functions for sampling the principal components of the log-hazards with priors based on the parametric model in the GPCH model

Again, we retain lines 1 – 33 of Appendix A.5.1 and we include the following commands

```

34 eigenv<-eigen(covar) #eigen values and vectors of covariance matrix
35 eigenvv<-eigenv$vectors
36 covpc<-t(eigenvv)%%covar%%eigenvv #covariance matrix
37 #of the principal components
38 covpc<-round(covpc,digits=4)
39 varpc<-diag(covpc) #variance of the principal components
40 stdpc<-sqrt(varpc)
41 meanpc<-eigenvv%%mu #mean of the principal components
42 ineigen<-solve(eigenvv)

```

```
43 pcs<-eigenv%%eta #transforming log hazards to the principal components
44 eta<-ineigen%%pcs
45 result<-matrix(nrow=niter,ncol=ncom)
46 for(iter in 1:niter)
47 {
48 for(i in 1:ncom)
49 {
50 logprior<- -0.5*(((pcs[i]-meanpc[i])/stdpc[i])^2)#logarithm of prior
51 loglike<-loglikewall(eta,nd,nt) #loglike contribution
52 #from all individuals in the various covariate profiles
53 pcs.prop<-pcs
54 pcs.prop[i]<-rnorm(1,pcs[i],sdpc.prop)#proposing a
55 #value for the principal component
56 eta.prop<-ineigen%%pcs.prop #transforming
57 #the principal components back to the log hazards
58 logprior.prop<- -0.5*(((pcs.prop[i]-meanpc[i])/stdpc[i])^2)
59 loglike.prop<-loglikewall(eta.prop,nd,nt) #loglike contribution from
60 #all individuals in the various covariate profiles
62 #using the proposed values of the parameters
63 logpost<-loglike+logprior
64 logpost.prop<-loglike.prop+logprior.prop
65 aprob<-logpost.prop-logpost
66 aprob<-min(0,aprob)
67 u=runif(1)
68 if(log(u)<aprob)
69 {
70 loglike<-loglike.prop
71 pcs[i]<-pcs.prop[i]
72 eta<-ineigen%%pcs
73 }
74 result[iter, i]<-eta[i]
75 }
76 }
78 result
79 }
```

A.5.7 Functions for sampling the log-hazards when we approximate the proposal distribution with priors based on the parametric model in the GPCH model

We retain lines 1 – 33 of Appendix A.5.1 and we include the following commands

```

34 mle=mle2(mu,nd,nt)
35 eta<-mle
36 findmode<-firstseconddt(mu,nd,nt,incovar,mle,tol=0.001,maxiter=7)
37 postmode<-findmode[[1]]
38 propcovar<- solve(findmode[[2]])
39 CDmatrix<-chol(propcovar) #cholesky decomposition
40 #of the covariance matrix
41 result<-matrix(nrow=niter,ncol=ncom)
42 for(iter in 1:niter)
43 {
44 loglike<-loglikewall(eta,nd,nt) #loglike contribution
45 #from all individuals in the various covariate profiles
46 sub<-eta-mu
47 logprior<--0.5*(t(sub)%*(incovar%*sub))
48 #log prior density of the parameters
49 logpost<-logprior+ loglike
50 epsilon<-rnorm(ncom,0,1)
51 # iid random variables with zero mean and variance 1
52 etatilde<-CDmatrix%*epsilon
53 eta.prop<-etatilde+postmode
54 logpropden1<-propdenn(propcovar,postmode, eta) #proposal density
55 #of the parameters given the proposed values of the parameters
56 logpropden2<-propdenn(propcovar,postmode, eta.prop) ##proposal
57 # density of the proposed parameters given the parameters
58 loglike.prop<-loglikewall(eta.prop,nd,nt) ##loglike contribution
59 #from all individuals in the various covariate profiles using
60 # the proposed values of the parameters
61 subb<-eta.prop-mu
63 logprior.prop<- -0.5*(t(subb)%*(incovar%*subb)) #log prior
64 #density of the proposed parameters
65 logpost.prop<-logprior.prop+ loglike.prop #log posterior
66 #density of the proposed parameters
67 A<-logpost.prop-logpost

```

```

68 logdenetaratio<-logpropden1-logpropden2
69 aprob<-min(0,A+logdenetaratio) #acceptance probability
70 u=runif(1)
71 if(log(u)<aprob)
72 {
73 eta<-eta.prop
74 }
75 result[iter,]<-eta
76 }
77 result
78 }

```

A.5.8 Functions consisting of all functions used in sampling the log-hazards with Markov random field prior in the GPCH model

This function is similar to Appendix A.5.3. We retain Appendix A.5.3 but replace lines 30 – 66 with

```

#Creating a vector of means for the covariate effects
meanvec<-function()
{
meanb0<--6
agem1<-0
agem2<-0
agem3<-0
sexm<-0
wbcm1<-0
wbcm2<-0
wbcm3<-0
depsc1m<-0
depsc2m<-0
depsc3m<-0
timem1<-0
timem2<-0
timem3<-0
timem4<-0
timem5<-0
timem6<-0

```

```
timem7<-0
timem8<-0
timem9<-0
meanvec<-c(meanb0,agem1,agem2,agem3,sexm,wbcm1,wbcm2,wbcm3,depsc1m,depsc2m,
depsc3m,timem1,timem2,timem3,timem4,timem5,timem6,timem7,timem8,timem9)
meanvec
}

#Creating a vector of variances for the covariate effects
varvec<-function()
{
varb0<-0.12
agev1<-0.02
agev2<-0.06
agev3<-0.003
sexv<-0.0625
wbcv1<-0.06
wbcv2<-0.02
wbcv3<-0.001
depsc1v<-0.012
depsc2v<-0.004
depsc3v<-0.002
timev1<-0.012
timev2<-0.004
timev3<-0.002
timev4<-0.012
timev5<-0.007
timev6<-0.005
timev7<-0.004
timev8<-0.003
timev9<-0.002
varvec<-c(varb0,agev1,agev2,agev3,sexv,wbcv1,wbcv2,wbcv3,depsc1v,depsc2v,
depsc3v,timev1,timev2,timev3,timev4,timev5,timev6,timev7,timev8,timev9)
varvec
}
```

We remove lines 163 – 253 from Appendix A.5.3.

We replace lines 564 – 609 with

```

Bstore<-function(covar,positionneig,i)
{
covar<-covar[c(i,positionneig),c(i,positionneig)]
c<-covar[,-1][1,]
v<-covar[-1,-1]
Bstore<-solve(v,c)
Bstore
}

Bstoreall<-function(covar,positionneig)
{
nlen<-length(covar[,1])
#Bstoreall=matrix(nrow=nlen,ncol=nlen-1)
Bstoreall=c()
for(i in 1:nlen)
{
Bstoreall[[i]]<-Bstore(covar,positionneig[[i]],i)
}
Bstoreall
}

#function for conditional variance for a particular etas given it's neighbours
#for one
condnvar<-function(etavar,i,positionneig)
{
covar<-etavar[c(i,positionneig),c(i,positionneig)]
c<-covar[,-1][1,]
#c=as.matrix(c)
cn<-covar[-1,][,1]
v<-covar[-1,-1]
condnvar<-covar[1,1]-(solve(v,c))%*%cn
condnvar
}

#for all
condnvarall<-function(etavar,positionneig)

```

```

{
nlen<-length(etavar[,1])
condnvarall=c()
for(i in 1:nlen)
{
condnvarall[[i]]<-condnvar(etavar,i,positionneig[[i]])
}
condnvarall
}

```

Then, we include

```

ncovp<-prod(ncov)
l<-c(0.1,0.1,0.1,0.1,0.1)
####to create neighbours#####
neigh1<-function(x,y){
diff<-x-y
neig<-sum(abs(diff))==1
return(neig)
}

neigh2<-function(x,yy){
ny<-length(yy[,1])
neighx<-c()
neight<-list()
for (i in 1:ny){
neighx[i]<-neigh1(x,yy[i,])
}
return(neighx)
}

neigh5<-function(yy){
ny<-length(yy[,1])
neighall<-list()
writneig<-list()
profnum<-list()
for (i in 1:ny){
neighall[[i]]<-neigh2(yy[i,],yy)
writneig[[i]]<-yy[which(neighall[[i]]==TRUE),]
}
}

```

```
profnum[[i]]<-profilenum(writneig[[i]])
}
profnum
}

#function to find the distance between two covariate vectors
dist<-function(x,y,l)
{
dist=sum(l*abs(x-y))
return(dist)
}

#function to find the distance among covariate vectors of
# all patients and putting them in form of a matrix
distall<-function(covariates,l)
{
ncov<-length(covariates[,1])
distall<-matrix(nrow=ncov,ncol=ncov)
for( i in 1:ncov)
{
for( j in 1:ncov)
{
distall[i,j]<-dist(covariates[i,],covariates[j,],l)
}
}
distall
}

#function to create correlation matrix
corrall<-function(distance,b,c)
{
ncov<-length(distance[,1])
corrall<-matrix(nrow=ncov,ncol=ncov)
for( i in 1:ncov)
{
for( j in 1:ncov)
{
```

```

corrall[i,j]<-b*(c +(1-c)*exp(-distance[i,j]))
corrall[i,i]<-1
}
}
corrall
}

patientalive<-function(xx,covart,listprof){
nx<-length(xx[,1])
stoppedd<-covart[,7]
writsamex<-list()
for (i in 1:nx){
writsamex[[i]]<-listprof[[i]] [which(listprof[[i]]$stopped>=xx[i,5]),]
}
return(writsamex)
}

```

A.5.9 Functions for sampling the log-hazards with Markov random field prior in the GPCH model

```

1 #Generalised piecewise constant hazard model using
2 #Markov random field priors for the leukaemia data set
3 gpch<-function(b,niter)
4 {
5   c=0.2
6   covariates=as.matrix(covariates)
7   n<-length(covariates[,1])
8   combmatrix=bigconstruct(nlevel)
9   designmatrix<-Xmat(combmatrix) ## design matrix
10  ncom<-length(combmatrix[,1])
11  profile<-1:ncom
12  combmatrix1p<-data.frame(combmatrix,profile)
13  bmean<-meanvec()
14  bvar<-varvec() ##prior variance for parameter
15  cotry<-read.table("lcovt1.txt",header=TRUE)
16  ##correlation matrix of parameters
17  co<-as.matrix(cotry)
18  bbvar<-bbvar(bvar,co) ##creating covariance matrix of parameters

```

```

19 bmean1<-as.matrix(bmean)
20 priormeaneta<-designmatrix%*%bmean1 ##prior mean for log hazard
21 mu<-as.matrix(priormeaneta) ## prior mean for log hazards
22 eta<-as.matrix(eta)
23 distance<-distall(combmatrix,1)
24 corr<- corral(distance,b,c)
25 etavar<-designmatrix%*%bbvar%*%t(designmatrix) ##covariance of log hazard
26 etavard<-diag(etavar) #diagonal elements of covariance of log hazard
27 etastd<-sqrt(etavard)
28 etaS<-diag(etastd)
29 covar<-etaS%*%corr%*%etaS
30 neighallnum=neigh5(combmatrix) #changing neighbours of
31 covariates from string to numeric
32 condnvarall<-condnvarall(covar,neighallnum) #conditional variance
33 #for a particular etas given it's neighbours
34 sdetaall<-sqrt(condnvarall) #conditional standard deviations
35 # for a particular etas given it's neighbours
36 Bmatrix=Bstoreall(covar,neighallnum) #storing values
37 #to be used in calculating the conditional mean
38 ndtt=ndtfun(cuts,ncom,covarprof,combmatrixx)
39 nd<-ndtt[[2]]
40 nt<-ndtt[[1]]
41 result<-matrix(nrow=niter,ncol=ncom)
42 for(iter in 1:niter)
43 {
44 for(i in 1:ncom)
45 {
46 loglike=loglikew1(eta[i],nd[i],nt[i])
47 positionneig<-neighallnum[[i]]
48 othereta<-eta[positionneig]
49 othermu<-mu[positionneig]
50 condmean<-mu[i,1]+Bmatrix[[i]]%*%(othereta-othermu)
51 logprior<- -1/2*(((eta[i]-condmean)/sdetaall[i])^2)
52 logpost<-logprior+ loglike
53 eta.prop<-rnorm(1,eta[i],sdetaprop)
54 loglike.prop<-loglikew1(eta.prop,nd[i],nt[i])
55 #likelihood contribution of each profile
56 logprior.prop<- -1/2*(((eta.prop-condmean)/sdetaall[i])^2)

```

```
57 logpost.prop<-logprior.prop+ loglike.prop
58 A<-logpost.prop-logpost
59 aproprob<-min(0,A)
60 u=runif(1)
61 if(log(u)<aproprob)
62 {
63 eta[i]<-eta.prop
64 }
65 }
66 result[iter,]<-eta
67 }
68 result
69 }
```

A.5.10 Functions for sampling the principal components of the log-hazards using Markov random field priors in the GPCH model

Again, we retain lines 1 – 29 of Appendix A.5.9 and we add up all of lines 34 – 79 of Appendix A.5.6.

A.5.11 Functions for splitting the log-hazards and sampling using forward backward algorithm with Markov random field priors in the GPCH model

We retain lines 1 – 29 of Appendix A.5.9 and we add up all of lines 31 – 158 of Appendix A.5.4.

A.5.12 Functions for splitting the log-hazards and sampling using Cholesky decomposition

We retain lines 1 – 29 of Appendix A.5.9 and we add up all of lines 43 – 91 of Appendix A.5.5.

A.5.13 Functions for sampling the log-hazards when we approximate the proposal distribution with Markov random field priors

We retain lines 1 – 29 of Appendix A.5.9 and 34 – 78 of Appendix A.5.7.

A.5.14 Functions consisting of all functions used in sampling the log-hazards using Gaussian process prior in the continuous parameter space model

```

1 ##All functions for flexible modelling of log hazards in the
2 ## continuous parameter space model using the Gaussian process prior
3
4 data<-read.table("snlgcompt.txt",header=TRUE) #dataset
5 ndata<-length(data[,1])
6 meanlogfrailty<-0 ## mean of log frailty
7 varlogfrailty<-1 ## variance of log frailty
8 prior<-list(aprior=1.5,bprior=1.5,meanlogfrailty=meanlogfrailty,
9 varlogfrailty=varlogfrailty) #prior for parameters
10
11 #function to group patients by profile
12 profiler<-function(x)
13 {# x is a matrix of covariate values with one row for
14 #each patient and one column for each covariate
15 n<-length(x[,1])
16 n.covars<-length(x[1,])
17 profile.list<-numeric(n)
18 profile.list[1]<-1 # A list of which profile each patient has.
19 profile.mat<-x[1,]
20 dim(profile.mat)<-c(1,n.covars)
21 n.profile<-1 # The number of distinct profiles.
22 profile.n<-1 # The numbers of cases with each profile.
23 for (i in 2:n)
24   {new<-TRUE
25     for (j in 1:n.profile)
26       {if ((sum(x[i,]==profile.mat[j,])==n.covars) & new)
27         {new<-FALSE
28           profile.n[j]<-profile.n[j]+1
29           profile.list[i]<-j
30         }
31       }
32     if (new)
33       {n.profile<-n.profile+1
34         profile.list[i]<-n.profile

```

```
35     profile.n[n.profile]<-1
36     profile.mat<-rbind(profile.mat,x[i,])
37   }
38 }
39 out<-list(n.profile=n.profile,profile.n=profile.n,
40 profile.list=profile.list,profile.mat=profile.mat)
41 return(out)
42 }
43
44 dat<-data[,-(1:2)]
45 datt<-as.matrix(dat)
46 profsearch<-profiler(datt)
47 numprof<-profsearch[[1]] # number of covariate profiles
48 numinprof<-profsearch[[2]] #number of individuals in each profiles
49 patientsbyprof<-profsearch[[3]] #list individuals according to their profiles
50 distinctprofiles<-profsearch[[4]] #The number of distinct profiles.
51 eta<-rep(-1.5,numprof)
52 neweta<-rep(-0.7,ndata)
53 alpha=1 #Weibull shape parameter
54 proposal<-list(a=4,b=4, af=1.1, bf=0.53,sdetaprop=1,sdpc.prop=1)
55
56 ##function to compute mean of log hazards by profile
57 meaneta<-function(patientsbyprof,eta){
58 n<-max(patientsbyprof)
59 meaneta<-numeric(n)
60 for(i in 1:n)
61 {
62 meaneta[i]<-mean(eta[which(patientsbyprof==i)])
63 }
64 meaneta
65 }
66
67 ##function to compute the maximum likelihood estimate
68 mle1<-function(data,numprof,patientsbyprof,alpha,meaneta)
69 {
70 datprof<-list()
71 died<-list()
72 t<-list()
```



```
73 mle1<-numeric(numprof)
74 for(i in 1:numprof)
75 {
76
77 datprof[[i]]<-data[patientsbyprof==i,]
78 died[[i]]<-sum(datprof[[i]]$died)
79 t[[i]]<-datprof[[i]]$t
80 if(died[[i]]>0)
81 {
82 mle1[[i]]<- log(died[[i]] / sum(t[[i]]^alpha))
83 }
84 else
85 {
86 mle1[[i]]<-meaneta[i]
87 }
88 }
89 mle1
90 }

91 #function to group and list patients with similar profiles
92 profbygroup<-function(patientsbyprof,data){
93 n<-max(patientsbyprof)
94 profbygroup<-list()
95 for(i in 1:n)
96 {
97 profbygroup[[i]]<-data[which(patientsbyprof==i),]
98 }
99 profbygroup
100 }
101
102 #Creating a vector of means for the covariate effects
103 meanvec<-function()
104 {
105 meanb0<--1.5
106 agem<-0.04
107 hbm<-0.02
108 wbcm<-0.08
109 sexm<-0.05
```

```
110 albuminm<-0
111 apm<-0
112 uream<-0
113 extranodm<-0
114 bulkm<-0
115 marrowm<-0
116 bsym<--0
117 stage1m<-0
118 stage2m<--0
119 stage3m<-0
120 ecog1m<-0
121 ecog2m<--0
122 ecog3m<-0
123 ecog4m<-0
124 ldh1m<-0
125 ldh2m<-0
126 meanvec<-c(meanb0,agem,hbm,wbcm,sexm,albuminm,apm,uream,
127 extranodm,bulkm,marrowm,bsym,stage1m,stage2m,stage3m,ecog1m,
128 ecog2m,ecog3m,ecog4m,ldh1m,ldh2m)
129 meanvec
130 }
131
132 #Creating a vector of variances for the covariate effects
133 varvec<-function()
134 {
135 varb0<-0.16
136 agev<-0.009
137 hbv<-0.002
138 wbcv<-0.004
139 sexv<-0.005
140 albuminv<-0.003
141 apv<-0.003
142 ureav<-0.003
143 extranodv<-0.003
144 bulkv<-0.003
145 marrowv<-0.003
146 bsyv<-0.003
147 stage1v<-0.002
```

```
148 stage2v<-0.006
149 stage3v<-0.003
150 ecog1v<-0.002
151 ecog2v<-0.006
152 ecog3v<-0.003
153 ecog4v<-0.002
154 ldh1v<-0.006
155 ldh2v<-0.002
156 varvec<-c(varb0,agev,hbv,wbcv,sexv,albuminv,apv,ureav,
157 extranodv,bulkv,marrowv,bsyv,stage1v,stage2v,stage3v,
158 ecog1v,ecog2v,ecog3v,ecog4v,ldh1v,ldh2v)
159 varvec
160 }
161
162 #Creating contrast of categorical covariates for 1
163 #row in the design matrix
164 mattt<-function(stage,a){
165 xmat<-matrix(nrow=1,ncol=a-1)
166 for(j in 1:a-1){
167 if(stage<=j){
168 xmat[,j]<--1
169 }
170
171 if(stage==j+1){
172 xmat[,j]<-j
173 }
174
175 if(stage>j+1){
176 xmat[,j]<-0
177 }
178 }
179 xmat
180 }
181
182 #Creating contrast of categorical covariate for
183 #all rows in the design matrix
184 matttall<-function(stage,a)
185 {
```

```
186 n<-length(stage)
187 mathtall<-matrix(nrow=n, ncol=a-1)
188 for(i in 1:n){
189 mathtall[i,<-matht(stage[i],a)
190 }
191 mathtall
192 }
193
194 #Creating the design matrix for the SNLG data set
195 Xmat<-function(datt){
196 age=datt[,1]
197 hb=datt[,2]
198 wbc=datt[,3]
199 sex=datt[,4]
200 albumin=datt[,5]
201 ap=datt[,6]
202 urea=datt[,7]
203 extranod=datt[,8]
204 bulk=datt[,9]
205 marrow=datt[,10]
206 bsy=datt[,11]
207 stage=datt[,12]
208 ecog=datt[,13]
209 ldh=datt[,14]
210 n<-length(datt[,1])
211 X<- matrix(nrow=n,ncol=12)
212 X[,1]<-rep(1,n) #1;intercept
213 X[,2]<-age #2;age
214 X[,3]<-hb #3;hb
215 X[,4]<-wbc #4;wbc
216 X[,5]<-ifelse(sex==1,-1,1) #5:sex
217 X[,6]<-ifelse(albumin==1,-1,1) #6:albumin
218 X[,7]<-ifelse(ap==1,-1,1) #7:ap
219 X[,8]<-ifelse(urea==1,-1,1) #8:urea
220 X[,9]<-ifelse(extranod==1,-1,1) #9:extranod
221 X[,10]<-ifelse(bulk==1,-1,1) #10:bulk
222 X[,11]<-ifelse(marrow==1,-1,1) #11:marrow
223 X[,12]<-ifelse(bsy==1,-1,1) #12:bsy
```

```
224 matstage<-matttall(stage,4)
225 matecog<-matttall(ecog,5)
226 matldh<-matttall(ldh,3)
227 X<-data.frame(X,matstage,matecog,matldh)
228 X<-as.matrix(X)
229 return(X)
230 }
231
232 #Creating the covariance matrix of covariate effects
233 # using the correlation matrix "co".
234 bbvar<-function(var,co){
235 std<-sqrt(var)
236 l<-length(var)
237 bbvar<-matrix(nrow=l, ncol=l)
238 for(i in 1:l){
239 for(j in 1:l){
240 bbvar[i,j]<-std[i]*std[j]*co[i,j]
241 }
242 }
243 bbvar
244 }
245
246 #Finding the distance between two covariate vectors
247 dist<-function(x,y,matx)
248 {
249 a=t(x-y)
250 b=x-y
251 dist= sqrt(a%*(matx%*b))
252 return(dist)
253 }
254
255 #Creating a matrix where the entries are the distances
256 # between the corresponding covariate vectors
257 distall<-function(covariates,matx)
258 {
259 ncov<-length(covariates[,1])
260 distall<-matrix(nrow=ncov,ncol=ncov)
261 for( i in 1:ncov)
```

```
262 {
263 for( j in 1:ncov)
264 {
265 distall[i,j]<-dist(covariates[i,],covariates[j,],matx)
266 }
267 }
268 distall
269 }
270
271 #Creating a correlation matrix using the distance matrix
272 corrrall<-function(distance,b,c)
273 {
274 ncov<-length(distance[,1])
275 corrrall<-matrix(nrow=ncov,ncol=ncov)
276 for( i in 1:ncov)
277 {
278 for( j in 1:ncov)
279 {
280 corrrall[i,j]<-b*(c +(1-c)*exp(-distance[i,j]))
281 corrrall[i,i]<-1
282 }
283 }
284 corrrall
285 }
286
287 #function to create fixed terms which are used when
288 # finding the conditional variance
289 Bstore<-function(covar,i)
290 {
291 c<-covar[,-i][i,]
292 v<-covar[-i,-i]
293 Bstore<-solve(v,c)
294 Bstore
295 }
296
297 #function to create n by n-1 matrix matrix that stores the constant
298 #part of the formula used to find a conditional mean or variance
299 Bstoreall<-function(covar)
```

```
300 {
301 nlen<-length(covar[,1])
302 Bstoreall=matrix(nrow=nlen,ncol=nlen-1)
303 for(i in 1:nlen)
304 {
305 Bstoreall[i,<-Bstore(covar,i)
306 }
307 Bstoreall
308 }
309
310 #function to find the conditional variance of one
311 #log hazard given all other log hazards
312 condnvar<-function(covar,i)
313 {
314 c<-covar[,-i][i,]
315 cn<-covar[-i,][,i]
316 v<-covar[-i,-i]
317 condnvar<-covar[i,i]-(solve(v,c))%*%cn
318 condnvar
319 }
320
321 #function to compute the conditional variance of one log hazard
322 #given all other log hazards for the vector of log hazards
323 condnvarall<-function(covar)
324 {
325 nlen<-length(covar[,1])
326 condnvarall=numeric(nlen)
327 for(i in 1:nlen)
328 {
329 condnvarall[i]<-condnvar(covar,i)
330 }
331 condnvarall
332 }
333
334 #loglike contribution from individuals having
335 #similar covariate profiles
336 loglikemf<-function(data,eta,alpha){
337 died<-data$died
```

```

338 t<-data$t
339 nd<-length(died[died==1])
340 loglike<--(exp(eta)*sum(t^alpha))+nd*log(alpha) + nd*eta
341 #+ sum((alpha-1)*log(t)[died==1])
342 loglike
343 }
344
345 #loglike contribution from all individuals
346 loglikeaall<-function(data,eta,alpha){
347 ndd<-length(data)
348 loglike=0
349 for(i in 1:ndd){
350 loglike=loglike+loglikemf(data[[i]],eta[i],alpha)
351 }
352 loglike
353 }
354
355 #function to compute the log density of a distribution
356 propdenn<-function(cov,mean,eta){
357 logdetapp<-determinant(cov)
358 logdetapp<-logdetapp$modulus[1]
359 sub<-eta-mean
360 logdetaoldetaprop<--0.5*logdetapp - 0.5*(t(sub)%*%solve(cov,sub))
361 return(logdetaoldetaprop)
362 }
363
364 #number of deaths in each covariate profile
365 nd<-function(profbygroup,numprof){
366 nd<-numeric(numprof)
367 for(i in 1:numprof)
368 {
369 nd[i]<-sum(profbygroup[[i]]$died==1)
370 }
371 nd
372 }
373
374 #function to approximate the proposal distribution
375 #using Newton Raphson's algorithm

```



```

376 firstseconddd<-function(profbygroup,numprof,mu,alpha,nd,prior.prec,mle,
377 tol=0.001,maxiter=7){
378 ndd<-as.matrix(nd)
379 eta<-mle
380 tal<-numeric(numprof)
381 for(i in 1:numprof)
382 {
383 tal[i]<-sum((profbygroup[[i]]$t)^alpha)
384 }
385 tal<-as.matrix(tal)
386 go<-TRUE
387 for(i in 1:maxiter)
388   {if(go)
389     {lambda<-exp(eta)
390       lt<-lambda*tal
391       ltn<-as.numeric(lt)
392       em<-eta-mu
393       firstd<-ndd -lt-prior.prec%%em
394       secondd<- -prior.prec - diag(ltn)
395       delta<-solve(secondd,firstd)
396       eta<- eta - delta
397       check<- - t(delta)%%secondd%%delta
398       go<-(check>tol)
399     }
400   }
401 alll<-list(postmode=eta,postpre= -secondd)
402 alll
403 }

```

A.5.15 Functions for sampling log-hazards with Gaussian process prior in the continuous parameter space model using SNLG data set

```

1 #continuous parameter space model using
2 #Gaussian process prior for the log-hazards for the SNLG data set
3 gpp<-function(prior,proposal,b,niter)
4 {
5 n<-length(distinctprofiles[,1])
6 matt<-Xmat(distinctprofiles) # design matrix

```

```

7 bmean<-meanvec() #prior means for covariates effects
8 bvar<-varvec() #prior variances for covariate effects
9 co<-read.table("co.txt",header=TRUE) #correlation matrix of parameters
10 co<-as.matrix(co)
11 bvarrr<-bbvar(bvar,co) #covariance matrix of parameters
12 bmean1<-as.matrix(bmean)
13 priormeaneeta<-matt%%bmean1
14 mu<-as.matrix(priormeaneeta) #prior mean for parameters
15 aprior<-prior$aprior
16 bprior<-prior$bprior
17 eta<-as.matrix(eta)
18 a=proposal$a
19 sdetaprop=proposal$sdetaprop #proposal standard deviation for log hazards
20 matx=diag((sapply(dat,sd)^(-2))) #rescaled matrix
21 distance<-distall(distinctprofiles,matx) #matrix where the entries
22 #are the distances between the corresponding covariate vectors
23 c=0.2 #constant to avoid the correlation from tending to 0
24 corr=corrall(distance,b,c) #creates correlation matrix
25 etavar<-matt%%bvarrr%%t(matt) #creates covariance of the parameters
26 # with implied constraints (singular)
27 etavard<-diag(etavar) #diagonal elements of covariance
28 etastd<-sqrt(etavard)
29 etaS<-diag(etastd)
29 covar<-etaS%%corr%%etaS ##creates covariance of
30 #the parameters without constraints
31 profbygroup<-profbygroup(patientsbyprof,data)
32 condnvarall<-condnvarall(covar) #function to compute the conditional
33 #variance of a log hazard given the others for the vector of log hazards
34 sdetaall<-sqrt(condnvarall)
35 Bmatrix=Bstoreall(covar) #n by n-1 matrix #matrix which stores
36 #the constant part of the formula in finding a conditional mean or variance
37 result<-matrix(nrow=niter,ncol=numprof+1)
38 for(iter in 1:niter)
39 {
40 for(i in 1:numprof)
41 {
42 loglike<-loglikemf(profbygroup[[i]],eta[i],alpha)
43 othereta<-eta[-i,1]

```

```

44 othermu<-mu[-i,1]
45 condmean<-mu[i,1]+Bmatrix[i,]%*%(othereta-othermu)
46 logprior<- -1/2*(((eta[i]-condmean)/sdetaall[i])^2)
47 logpost<-logprior+ loglike
48 eta.prop<-rnorm(1,eta[i],sdetaprop) #proposed value for the log hazard
49 loglike.prop<-loglikemf(profbygroup[[i]],eta.prop,alpha)
50 logprior.prop<- -1/2*(((eta.prop-condmean)/sdetaall[i])^2)
51 logpost.prop<-logprior.prop+ loglike.prop
52 A<-logpost.prop-logpost
53 aprob<-min(0,A)
54 u=runif(1)
55 if(log(u)<aprob)
56 {
57 eta[i]<-eta.prop
58 }
59 result[iter,i]<-eta[i]
60 }
61 alpha.prop<-rgamma(1,a,a/alpha)
62 loglikealpha<-loglikeaall(profbygroup,eta,alpha)
63 loglikealpha.prop<-loglikeaall(profbygroup,eta,alpha.prop)
64 logpropdenratio<-(2*a-1)*(log(alpha) - log(alpha.prop))
65 - a*((alpha/alpha.prop) - (alpha.prop/alpha))
66 logprioralpha<-(aprior-1)*log(alpha) - bprior*alpha
67 logprioralpha.prop<-(aprior-1)*log(alpha.prop) - bprior*alpha.prop
68 Aaprob1=logprioralpha.prop+loglikealpha.prop-(logprioralpha+loglikealpha)
69 +logpropdenratio
70
71 Aaprob<-min(0,Aaprob1)
72 u=runif(1)
73 if(log(u)<Aaprob)
74 {
75 alpha<-alpha.prop
76 }
77 result[iter,numprof+1]<-alpha
78 }
79 result
80 }

```

A.5.16 Functions for sampling the principal components of the log-hazards with Gaussian process priors in the continuous parameter space model

Again, we retain lines 1 – 30 of Appendix A.5.15 and we include the following commands

```
eigenv<-eigen(covar) #eigen values and vectors of covariance matrix
eigenvv<-eigenv$vectors
covpc<-t(eigenvv)%*%covar)%*%eigenvv #covariance matrix
#of the principal components
covpc<-round(covpc,digits=4)
varpc<-diag(covpc) #variance of the principal components
stdpc<-sqrt(varpc)
meanpc<-eigenvv)%*%mu #mean of the principal components
ineigen<-solve(eigenv)
profbygroup<-profbygroup(patientsbyprof,data) #group
#patients with similar covariate profiles
pcs<-eigenvv)%*%eta #transforming log hazards to the principal components
eta<-ineigen)%*%pcs #transforming the principal components to log hazards
alpha=1
result<-matrix(nrow=niter,ncol=numprof+1)
for(iter in 1:niter)
{
for(i in 1:numprof)
{
logprior<- -0.5*(((pcs[i]-meanpc[i])/stdpc[i])^2)#logarithm of prior
loglike<-loglikeaall(profbygroup,eta,alpha)
#loglikelihood contribution of involving all individuals
pcs.prop<-pcs
pcs.prop[i]<-rnorm(1,pcs[i],sdpc.prop)
#proposing a value for the principal component
eta.prop<-ineigen)%*%pcs.prop #transforming
#the principal components back to the log hazards
logprior.prop<- -0.5*(((pcs.prop[i]-meanpc[i])/stdpc[i])^2)
loglike.prop<-loglikeaall(profbygroup,eta.prop,alpha)
logpost<-loglike+logprior
logpost.prop<-loglike.prop+logprior.prop
aprobs<-logpost.prop-logpost
aprobs<-min(0,aprobs)
```

```

u=runif(1)
if(log(u)<aprob)
{
loglike<-loglike.prop
pcs[i]<-pcs.prop[i]
eta<-ineigen%*%pcs
}
}
result[iter, 1:numprof]<-eta
alpha.prop<-rgamma(1,a,a/alpha)
loglikealpha<-loglikeaall(profbygroup,eta,alpha)
loglikealpha.prop<-loglikeaall(profbygroup,eta,alpha.prop)
logpropdenratio<-(2*a-1)*(log(alpha) - log(alpha.prop))
- a*((alpha/alpha.prop) - (alpha.prop/alpha))
logprioralpha<-(aprior-1)*log(alpha) - bprior*alpha
logprioralpha.prop<-(aprior-1)*log(alpha.prop) - bprior*alpha.prop
Aaprob1=logprioralpha.prop+loglikealpha.prop-(logprioralpha+loglikealpha)
+logpropdenratio
Aaprob<-min(0,Aaprob1)
u=runif(1)
if(log(u)<Aaprob)
{
alpha<-alpha.prop
}
result[iter,numprof+1]<-alpha
}
result
}

```

A.5.17 Functions for splitting and sampling the log-hazards using Gaussian process prior in continuous parameter space model

We retain lines 1 – 30 of Appendix A.5.15 and we include the following commands

```

meanlogfrailty<-prior$meanlogfrailty #prior mean of the log frailty
#priormeanlogfrailty<-as.matrix(meanlogfrailty)
varlogfrailty<-prior$varlogfrailty #prior variance of the log frailty

```

```

tauf<-1/varlogfrailty
thetapriormean<-mu #prior mean of log hazard without frailty
thetapriorcovar<-covar #prior covariance matrix of
#the log hazard without frailty
thetapriorpre<-solve(thetapriorcovar) #prior precision matrix of
#the log hazard without frailty
sdcovf<-sqrt(varlogfrailty)
numinprofmat<-diag(numinprof)
theta<-rep(1,numprof)
theta<-as.matrix(theta)
tauparam<- af + ndata/2
profbygroup<-profbygroup(patientsbyprof,data)
result<-matrix(nrow=niter,ncol=numprof+2)
for(iter in 1:niter)
{
f <- eta - theta[patientsbyprof]
tauf<-rgamma(1,tauparam, bf + 0.5*sum(f^2))
result[iter,numprof+2]<-tauf
sdcovf<- sqrt(1/tauf)
for(i in 1:ndata)
{
loglike<-loglikemf(data[i,],eta[i],alpha)
#loglike contribution from individual
logprior<- -0.5*(((eta[i]-theta[patientsbyprof[i]])/sdcovf)^2)
#log priorof the log hazard
logpost<-logprior+ loglike
eta.prop<-rnorm(1,eta[i],sdetaprop)
loglike.prop<-loglikemf(data[i,],eta.prop,alpha) ##loglike contribution
#from individual using the proposed values of the parameters
logprior.prop<- -0.5*(((eta.prop-theta[patientsbyprof[i]])/sdcovf)^2)
logpost.prop<-logprior.prop + loglike.prop
A<-logpost.prop-logpost
aaprob<-min(0,A)
u=runif(1)
if(log(u)<aaprob)
{
eta[i]<-eta.prop
}
}
}

```

```

}
alpha.prop<-rgamma(1,a,a/alpha)
loglikealpha<-loglikemf(data,eta,alpha)
loglikealpha.prop<-loglikemf(data,eta,alpha.prop)
logpropdenratio<-(2*a-1)*(log(alpha) - log(alpha.prop))
- a*((alpha/alpha.prop) - (alpha.prop/alpha))

logprioralpha<-(aprior-1)*log(alpha) - bprior*alpha
logprioralpha.prop<-(aprior-1)*log(alpha.prop) - bprior*alpha.prop
Aaprob1=logprioralpha.prop+loglikealpha.prop-(logprioralpha+loglikealpha)
+logpropdenratio

Aaprob<-min(0,Aaprob1)
u=runif(1)
if(log(u)<Aaprob)
{
alpha<-alpha.prop
}
frailtypriorpre<-tauf*numinprofmf
thetapostpre<-thetapriorpre+frailtypriorpre #posterior precision matrix
#of the log hazard
thetapostcovar<-solve(thetapostpre) #posterior covariance matrix
#of the log hazards
CDmatrix<-chol(thetapostcovar) #cholesky decomposition #of the covariance matrix
thetapostmean<-thetapriorpre%*%thetapriormean+frailtypriorpre%*%eta
thetapostmean<-solve(thetapostpre,thetapostmean) #posterior mean of
#log hazard without frailty
epsilon<-rnorm(numprof,0,1)#iid random variables with zero mean and variance 1
etatilde<-CDmatrix%*%epsilon
theta<-etatilde+thetapostmean #sampled log hazard values
result[iter,]<-c(theta,alpha,tauf)
}
result
}

```

A.5.18 Functions for sampling the log-hazards with Gaussian process priors using Gaussian approximation of the proposal distribution in the continuous parameter space model

We retain lines 1 – 30 of Appendix A.5.1 and we include the following commands

```

incovar<-solve(covar)
profbygroup<-profbygroup(patientsbyprof,data)
mle=mle1(data,numprof,patientsbyprof,alpha,mu)
eta<-mle
nd=nd(profbygroup,numprof)
result<-matrix(nrow=niter,ncol=n+1)
for(iter in 1:niter)
{
loglike<-loglikeaall(profbygroup,eta,alpha) #loglike contribution
# from all individuals in the various covariate profiles
sub<-eta-mu
logprior<--0.5*(t(sub)%*(incovar%*sub)) #log prior density
#of the parameters
logpost<-logprior+ loglike #log posterior density of the parameters
findmode<-firstsecondd(profbygroup,numprof,mu,alpha
,nd,incovar,mle,tol=0.001,maxiter=7)

postmode<-findmode[[1]]
propcovar<- solve(findmode[[2]])
CDmatrix<-chol(propcovar) #cholesky decomposition of the covariance matrix
epsilon<-rnorm(numprof,0,1) # iid random
#variables with zero mean and variance 1
etatilde<-CDmatrix%*%epsilon
eta.prop<-etatilde+postmode
logpropden1<-propdenn(propcovar,postmode, eta) #proposal
#density of the parameters given the proposed values of the parameters
logpropden2<-propdenn(propcovar,postmode, eta.prop) ##proposal
# density of the proposed parameters given the parameters
loglike.prop<-loglikeaall(profbygroup,eta.prop,alpha) ##loglike
#contribution from all individuals in the various covariate
#profiles using the proposed values of the parameters
subb<-eta.prop-mu
logprior.prop<- -0.5*(t(subb)%*(incovar%*subb)) #log prior

```



```

#density of the proposed parameters
logpost.prop<-logprior.prop+ loglike.prop #log posterior
#density of the proposed parameters
A<-logpost.prop-logpost
logdenetaratio<-logpropden1-logpropden2
aprob<- min(0,A+logdenetaratio) #acceptance probability
u=runif(1)
if(log(u)<aprob)
{
loglike<-loglike.prop
eta<-eta.prop
}
result[iter,1:n]<-eta
alpha.prop<-rgamma(1,a,a/alpha) #proposing a
#value for weibull shape parameter
overallloglike.prop<-loglikeaall(profbygroup,eta,alpha.prop)
#log likelihood contribution from the Weibull shape parameter
logpropdenratio<-(2*a-1)*(log(alpha) - log(alpha.prop)) -
a*((alpha/alpha.prop) - (alpha.prop/alpha))
#ratio of the log proposal density of Weibull shape parameter
# and the proposed Weibull shape parameter
logprioralpha<-(aprior-1)*log(alpha) - bprior*alpha
# log prior of the Weibull shape parameter
logprioralpha.prop<-(aprior-1)*log(alpha.prop)
- bprior*alpha.prop ## log prior of the proposed Weibull shape parameter
Aa=(logprioralpha.prop+overallloglike.prop)-(logprioralpha+loglike)
+logpropdenratio

Aaprob<-min(0,Aa)
u=runif(1)
if(log(u)<Aaprob)
{
alpha<-alpha.prop
}
result[iter,n+1]<-alpha
}
result
}

```

A.5.19 Function to calculate the prognosis indices using the generalised piecewise constant hazard model with the leukaemia data set

```

indexgpp<-function()
{
  #AGE
  write(file="", "Please enter the Age in years of the patient
    at time of diagnosis.")
  age<-scan(n=1)
  age[which(age<=(49))]=1
  age[which(age>(49)&age<=65)]=2
  age[which(age>65&age<=74)]=3
  age[which(age>74)]=4

  # SEX
  write(file="", "Please enter the Sex of the patient.
    Enter 1 for male or 0 for female.")
  sex<-scan(n=1)
  sex<-sex+1
#WBC
  write(file="", "Please enter the White Blood Cell count for the patient." )
  wbc<-scan(n=1)
  wbc[which(wbc<=1.8)]=1
  wbc[which(wbc>1.8&wbc<=7.90)]=2
  wbc[which(wbc>7.90&wbc<=38.65)]=3
  wbc[which(wbc>38.65)]=4
#DEPSC
  write(file="", "Please enter the Deprivation score measurement for the patient." )
  depsc<-scan(n=1)
  depsc<-depsc
  depsc[which(depsc<=(-2.7))]=1
  depsc[which(depsc>(-2.7)&depsc<=(-0.370))]=2
  depsc[which(depsc>(-0.370)&depsc<=2.93)]=3
  depsc[which(depsc>2.93)]=4
#TIME
  write(file="", "Please enter the time for the patient." )
  time<-scan(n=1)

```

```
time[which(time<=52.6)]=1
time[which(time>52.6&time<=111.6)]=2
time[which(time>111.6&time<=178.3)]=3
time[which(time>178.3&time<=255.4)]=4
time[which(time>255.4&time<=346.6)]=5
time[which(time>346.6&time<=458.1)]=6
time[which(time>458.1&time<=601.99)]=7
time[which(time>601.99&time<=804.7)]=8
time[which(time>804.7&time<=1151.3)]=9
time[which(time>1151.3)]=10

eta<-read.table("nnnmrfmean.txt")
eta<-eta[1:1280,]
patientsdata<-c(age ,sex,wbc,depsec,time)
len<-length(patientsdata)
datt<-read.table("combmatrix.txt",header=TRUE)
ndata<-length(datt[,1])
sump<-c()
  for(i in 1:ndata){
sump[i]<-sum(patientsdata==datt[i,])
}
write(file="", "The covariate vector of the patient is")
write(patientsdata,file="")
loghazard<-eta[which(sump==len)]
  write(file="", "The logarithm of the hazard is.")
  write(loghazard,file="")
mean          <-  -3.062487
std.dev       <-   0.7382914
  ind<-100*pnorm(loghazard,mean,std.dev)
  ind<-round(ind)
# ind<-c(mu,ind)
  write(file="", "Index value is")
  write(ind,file="")
  write(file="", "The index is on a scale from 0 to 100,
  Greater index values indicate greater risk.")
}
```

A.5.20 Function to calculate the prognosis indices using the continuous parameter space model with the SNLG data set

We retain lines 1 – 94 of Appendix A.4.2 and we include the following commands

```

95 eta<-read.table("nnncpsbasicmean.txt")
96 patientsdata<-c(age,albumin,ap,bsy,bulk,ecog,extranod,hb
    ldh,marrow,sex,stage,urea,wbc)
97 eta<-eta[1:636,]
98 eta<-as.numeric(eta)
99 len<-length(patientsdata)
100 datt<-read.table("snlgarcom.txt",header=TRUE)
101 ndata<-length(datt[,1])
102 sump<-c()
103 for(i in 1:ndata){
104 sump[i]<-sum(patientsdata==datt[i,])
105 }
106 write(file="", "The covariate vector of the patient is")
107 write(patientsdata,file="")
108 loghazard<-eta[which(sump==len)]
109 write(file="", "The logarithm of the hazard is.")
110 write(loghazard,file="")
112 mean          <-  mean(eta)
113 std.dev       <-  sd(eta)
114 ind<-100*pnorm(loghazard,mean,std.dev)
115 ind<-round(ind)
116 # ind<-c(mu,ind)
117 write(file="", "Index value is")
118
119 write(ind,file="")
120 write(file="", "The index is on a scale from 0 to 100,
121 Greater index values indicate greater risk.")
122 }

```

A.5.21 Function to calculate effective sample size

```

essf<-function(time,samples,sizeb,nbatch)
{
niter<-length(samples)

```

```

overallmean<-mean(samples)
sdooverall<-sd(samples)
nnbatch<-1:nbatch
cutbs<-sizeb*nnbatch
fposition<-sizeb*(nnbatch-1)+1
sposition<-sizeb*nnbatch
#write(sposition,file="")
aq<-sizeb/(nbatch-1)
position<-list()
meanbatches<-numeric(nbatch)
for(i in 1:nbatch)
{
position[[i]]<-c(fposition[i]:sposition[i])
meanbatches[i]<-mean(samples[position[[i]])]
}
ph<-aq*sum((meanbatches-overallmean)^(2))
mcse<-sqrt(ph/niter)
ess<-(sdooverall/mcse)^(2)
ess

}

essfn<-function(time,samplesall,sizeb,nbatch)
{
nparams<-length(samplesall[1,])
ess<-c()
for(i in 1:nparams){
ess[i]<-essf(time,samplesall[,i],sizeb,nbatch)
}
aveess<-mean(ess)
aveess
}

```

A.6 List of notations

Table A.3: Glossary of notations

Symbols	Meaning
C	Censoring
R	Set of right censored individuals
E	Set of individuals that had the event.
n	Number of individuals indexed $i = 1, \dots, n$
n_d	Number of individuals who had the event
t	Survival time
$F(t)$	Lifetime distribution or cumulative distribution function
$f(t)$	Probability density function
$S(t)$	Survival function
$h(t)$	Hazard function
$H(t)$	Cumulative hazard function
$h_0(t)$	Baseline hazard function
$S_0(t)$	Baseline survival function
S	Number of covariates indexed $s = 1, \dots, S$
λ_i	Hazard multiplier which depends on the covariates of the i^{th} individual.
η_i	Linear predictor of the i^{th} individual.
$x_{i,s}$	The value of the covariate s for the i^{th} individual.
β_0	Baseline parameter
φ_i	Acceleration factor which depends on the covariates of the i^{th} individual.
λ	Weibull scale parameter
α	Weibull shape parameter
ϕ	Missing data parameter
g	Link function
$\underline{x}_i = (1, x_{i,1}, \dots, x_{i,s})$	The covariate vector of the i^{th} individual.
J	Number of intervals indexed $j = 1, \dots, J$
z_i	frailty of the i^{th} individual.
$D = (x_1, \dots, x_n)$	Set of data
$\pi(\theta)$	Prior distribution of parameter θ
$\pi(\theta D)$	Posterior distribution of parameter θ given data D .
$L(\theta D)$	Likelihood contribution
$\underline{\mu}$	Vector of prior means of the parameters
\bar{V}	Prior covariance matrix of parameters
L_t	Lower limit of interval
U_t	Upper limit of interval
m	Number of categories of a covariate
ϑ	Proportion
t_m	Median survival time
$t_{q(1)}$	Lower quartile survival time
$t_{q(3)}$	Upper quartile survival time
τ_j	Cutpoint j
τ_f	Precision of the logarithm of the frailty
b	Possible number of covariate profiles
c	Total possible number of covariate profiles including time periods.
A	Acceptance probability

Table A.4: Glossary of abbreviations

Symbols	Meaning
<i>SNLG</i>	Scotland Newcastle Lymphoma Group
<i>NHL</i>	Non-Hodgkins lymphoma
<i>DLBCL</i>	Diffuse large B-cell lymphoma
<i>OS</i>	Overall survival
<i>TFR</i>	Time to first relapse
<i>ECOG</i>	Eastern Co-operative Oncology Group
<i>LDH</i>	Serum lactate dehydrogenase
<i>HB</i>	Haemoglobin
<i>WBC</i>	White blood cell
<i>AP</i>	Alkaline phosphase
<i>Urea</i>	Blood urea nitrogen
<i>Extranod</i>	Extranodal without bone marrow
<i>Marrow</i>	Bone marrow involvement
<i>Bsy</i>	B symptoms
<i>PCH</i>	Piecewise constant hazard
<i>MCMC</i>	Markov chain Monte Carlo
<i>AaIPI</i>	Age-adjusted International prognostic index
<i>FLIPI</i>	Follicular lymphoma International prognostic index
<i>MIPI</i>	Mantle cell International prognostic index
<i>NCCN – IPI</i>	National comprehensive cancer network International prognostic index
<i>fcd</i>	Full conditional distribution
<i>MH</i>	Metropolis-Hastings
<i>BUGS</i>	Bayesian inference using Gibbs sampler
<i>JAGS</i>	Just another Gibbs sampler
<i>GPCH</i>	Generalised piecewise constant hazard model

Bibliography

- ABRAMS, M., HAIGH, P. & COMFORT, J. 1988 Small sample theory for steady state confidence intervals. In proceedings of the Winter Simulation conference.
- AMERICAN RESEARCH SOCIETY 2013 [Http://www.cancer.org/cancer/non-hodgkinlymphoma](http://www.cancer.org/cancer/non-hodgkinlymphoma).
- BAUM, L. E., PETRIE, T., SOULES, G. & WEISS, N. 1970 A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Ann. Math. Statist.* **41** (1), 164–171.
- BERGER, J. O. 1985 *Statistical Decision Theory and Bayesian Analysis (second Edition)*. New York: Springer-Verlag.
- BERGER, J. O., BERNARDO, J. M. & SUN, D. 2009 The formal definition of reference priors. *The Annals of Statistics* **37** (2), 905–938.
- BOOTH, C. M. & EISENHAEUER, E. A. 2012 Progression-free survival: meaningful or simply measurable? *Journal of Clinical Oncology* **30** (10), 1030–1033.
- BRESLOW, N. 1972 Discussion on regression models and life-tables (by D.R Cox). *Journal of Royal Statistical Society* **34**, 216–217.
- BRESLOW, N. 1974 Covariance analysis of censored survival data. *Biometrics* **30**, 89–99.
- BURZYKOWSKI, T., BUYSE, M., PICCART-GEHART, M., SLEDGE, G., CARMICHAEL, J. & LUCK, H. 2008 Evaluation of tumor response, disease control, progression-free survival, and time to progression as potential surrogate end points in metastatic breast cancer. *Journal of Clinical Oncology* **26** (12), 1987–1992.
- BUTT, A. A., MICHAELS, S., GREEN, D., CLARK, R., KISSINGER, P. & MARTIN, D. H. 2002 Serum LDH level as a clue to the diagnosis of histoplasmosis. *AIDS Read* **12** (7), 317–321.
- CANCER RESEARCH UK 2012 <http://www.cancerresearchuk.org/cancer-help/type/non-hodgkins-lymphoma/treatment/staging-non-hodgkins-lymphoma>.

- CHATFIELD, C. 2004 *The Analysis of Time Series, 6th Edition*. Chapman and Hall.
- CHESON, B. D., FISHER, R. I., BARRINGTON, S. F., CAVALLI, F., SCHWARTZ, L. H., ZUCCA, E. & LISTER, T. A. 2014 Recommendations for initial evaluation, staging, and response assessment of Hodgkin and non-Hodgkin lymphoma: The Lugano classification. *Journal of Clinical Oncology* **32** (27), 3059–3067.
- CHIH-LIN, C., STREET, W. N. & WOLBERG, W. H. 2007 Application of artificial neural network-based survival analysis on two breast cancer datasets.
- CLEMEN, R. T., FISCHER, G. W. & WINKLER, R. L. 2000 Assessing dependence: some experimental results. *Management Science* **46** (8), 1100–1115.
- COLLETT, D. 1994 *Modelling survival data in medical research*. Chapman & Hall.
- CONGDON, P. 2001 *Bayesian statistical modelling*. Wiley.
- CORTESE, G., SCHEIKE, T. H. & MARTINUSSEN, T. 2009 Flexible survival regression modelling. *Statistical Methods in Medical Research* **19** (1), 5–28.
- COWLES, M. K. & CARLIN, B. P. 1996 Markov chain Monte Carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association* **91** (434), 883.
- COX, D. R. 1972 Regression models and life-tables. *Journal of the Royal Statistical Society B* **34**, 187–220.
- CUCALA, L., MARIN, J.-M., ROBERT, C. P. & TITTERINGTON, D. M. 2009 A Bayesian reassessment of nearest-neighbor classification. *Journal of the American Statistical Association* **104** (485), 263–273.
- DENISON, G. T. D., HOLMES, C. C., MALLICK, B. K. & SMITH, A. F. M. 2002 *Bayesian methods for nonlinear classification and regression*. Wiley.
- ELFADALY, F. G. & GARTHWAITE, P. H. 2013 Eliciting Dirichlet and Connor- Mosimann prior distributions for multinomial models. *TEST* **22** (4), 628–646.
- ENDRIZZI, L., FIORENTINO, M. V., SALVAGNO, L., SEGATI, R., PAPPAGALLO, G. L. & FOSSER, V. 1982 Serum lactate dehydrogenase (LDH) as a prognostic index for non-Hodgkin's lymphoma. *European Journal of Cancer and Clinical Oncology* **18** (10), 945–949.
- FARROW, M. 2003 Practical building of subjective covariance structures for large complicated systems. *Journal of the Royal Statistical Society: Series D (The Statistician)* **52** (4), 553–573.

- FARROW, M. 2011 MAS8303, Modern Bayesian inference. Lecture notes, Newcastle University.
- FEDERICO, M., BELLEI, M., MARCHESELLI, L., LUMINARI, S., LOPEZ-GUILLERMO, A., VITOLO, U., PRO, B., PILERI, S., PULSONI, A. & SOUBEYRAN, P. E. A. 2009 Follicular lymphoma international prognostic index 2: A new prognostic index for follicular lymphoma developed by the international follicular lymphoma prognostic factor project. *Journal of Clinical Oncology* **27** (27), 4555–4562.
- GAMERMAN, D. 1991 Dynamic Bayesian models for survival data. *Applied Statistics* **40** (1), 63.
- GAMERMAN, D. 1997 *Markov chain Monte Carlo*. Chapman & Hall.
- GARTHWAITE, P. H., KADANE, J. B. & O’HAGAN, A. 2005 Statistical methods for eliciting probability distributions. *Journal of the American Statistical Association* **100** (470), 680–701.
- GEMAN, S. & GEMAN, D. 1984 Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-6** (6), 721–741.
- GENEST, C. & KALBFLEISCH, J. 1988 Bayesian nonparametric survival analysis: Comment. *Journal of the American Statistical Association* **83** (403), 780.
- GILKS, W. R., RICHARDSON, S. & SPIEGELHALTER, D. J. 1996 *Markov chain Monte Carlo in practice*. Chapman & Hall.
- GOFFINET, D. R., CASTELLINO, R. A., KIM, H., DORFMAN, R. F., FUKS, Z., ROSENBERG, S. A., NELSEN, T. & KAPLAN, H. S. 1973 Staging laparotomies in unselected previously untreated patients with non-Hodgkin’s lymphomas. *Cancer* **32** (3), 672–681.
- GOLDSTEIN, M. & SHAW, S. C. 2004 Bayes linear kinematics and Bayes linear Bayes graphical models. *Biometrika* **91**, 425–446.
- GORDON, A. D., BREIMAN, L., FRIEDMAN, J. H., OLSHEN, R. A. & STONE, C. J. 1984 Classification and regression trees. *Biometrics* **40** (3), 874.
- GOSLING, J. P. 2014 Methods for eliciting expert opinion to inform health technology assessment. . Medical Research Council.
- HAARA, A., MALTAMO, M. & TOKOLA, T. 1997 The K-nearest-neighbour method for estimating basal-area diameter distribution. *Scandinavian Journal of Forest Research* **12** (2), 200–208.

- HASTINGS, W. K. 1970 Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57** (1), 97.
- HENDERSON, R. & OMAN, P. 1999 Effect of frailty on marginal regression estimates in survival analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **61** (2), 367–379.
- HENDERSON, R., SHIMAKURA, S. & GORST, D. 2002 Modeling spatial variation in leukemia survival data. *Journal of the American Statistical Association* **97** (460), 965–972.
- HIANCE, A., CHEVRET, S. & LEVY, V. 2009 A practical approach for eliciting expert prior beliefs about cancer survival in phase iii randomized trial. *Journal of Clinical Epidemiology* **62** (4), 431–437.e2.
- HIRSH, M. P. 1990 Report of a committee convened to discuss the evaluation and staging of patients with Hodgkin's disease: Cotswolds meeting. *Journal of Pediatric Surgery* **25** (7), 831.
- HOSMER, D. W. & LEMESHOW, S. 1999 *Applied Survival Analysis*. Wiley.
- HOSTER, E., DREYLING, M., KLAPPER, W., GISSELBRECHT, C., HOOF, A., KLUIN-NELEMANS, H. C., PFREUNDSCHUH, M., REISER, M., METZNER, B., EINSELE, H., PETER, N., JUNG, W., WORMANN, B., LUDWIG, W. D., DUHRSEN, U., EIMERMACHER, H., WANDT, H., HASFORD, J., HIDDEMANN, W. & UNTERHALT, M. 2008 A new prognostic index (MIPI) for patients with advanced-stage mantle cell lymphoma. *Journal of the American Society of Hematology* **111**, 558–565.
- HOUGAARD, P. 1999 Fundamentals of survival data. *Biometrics* **55** (1), 13–22.
- IBRAHIM, J. G., CHEN, M.-H. & SINHA, D. 2001 *Bayesian survival analysis*. Springer.
- IRONY, T. Z. & SINGPURWALLA, N. D. 1997 Non-informative priors do not exist: A dialogue with Jose M. Bernardo. *Journal of Statistical Planning and Inference* **65** (1), 159–177.
- JAMES, A., CHOY, S. L. & MENGENSEN, K. 2010 Elicitor: An expert elicitation tool for regression in ecology. *Environmental Modelling & Software* **25** (1), 129–145.
- JOHNSTON, L. 1996 [Http://oreilly.com/medical/lymphoma/news/nhl4.html](http://oreilly.com/medical/lymphoma/news/nhl4.html).
- JOLLIFFE, I. T. 2002 *Principal component analysis*. Springer.
- JONES, G. 2004 On the Markov chain central limit theorem. *Probab. Surveys* **1**, 299–320.

- KALBFLEISCH, J. & PRENTICE, R. 1973 Marginal likelihoods based on Cox's regression and life model. *Biometrika* **60**, 267–278.
- KIM, S., CHEN, M.-H., DEY, D. K. & GAMERMAN, D. 2007 Bayesian dynamic models for survival data with a cure fraction. *Lifetime Data Anal* **13** (1), 17–35.
- KLEINBAUM, D. G. & KLEIN, M. 2005 *Survival Analysis*. Springer.
- LAFLEUR-BROOKS, M. 1994 *Exploring medical language*. Mosby Lifeline.
- LEE, E. T. & WANG, J. W. 2003 *Statistical Methods for Survival Data Analysis*. Wiley.
- LINK, W. A. & EATON, M. J. 2011 On thinning of chains in MCMC. *Methods in Ecology and Evolution* **3** (1), 112–115.
- LITTLE, R. J. A. & RUBIN, D. B. 1987 *Statistical analysis with missing data*. Wiley.
- LONGMORE, J. M. & LONGMORE, J. M. 2007 *Oxford handbook of clinical medicine*. Oxford University Press.
- LUNN, D., JACKSON, C., BEST, N., THOMAS, A. & SPIEGELHALTER, D. 2013 *The BUGS Book: A Practical Introduction to Bayesian Analysis*. Boca Raton, London: CRC/Chapman & Hall.
- LUNN, D. J., THOMAS, A., BEST, N. & SPIEGELHALTER, D. 2000 WinBUGS: A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing* **10** (4), 325–337.
- LYMPHOMA.ORG 2016 Home-lymphoma research foundation - welcome.
- MALLET, S., ROYSTON, P., DUTTON, S., WATERS, R. & ALTMAN, D. G. 2010 Reporting performance of prognostic models in cancer: a review. *BMC Medicine* **8**, 21.
- MALTAMO, M. & KANGAS, A. 1998 Methods based on K -nearest neighbor regression in the prediction of basal area diameter distribution. *Can. J. For. Res.* **28** (8), 1107–1115.
- MARTINUSSEN, T. & SCHEIKE, T. H. 2006 *Dynamic regression models for survival data*. Springer.
- MCKEAGUE, I. W. & TIGHIOUART, M. 2000 Bayesian estimators for conditional hazard functions. *Biometrics* **56** (4), 1007–1015.
- METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H. & TELLER, E. 1953 Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21** (6), 1087.

- MOLINARI, N., DAURES, J.-P. & DURAND, J.-F. 2001 Regression splines for threshold selection in survival data analysis. *Statist Med* **20**, 237–247.
- MORRIS, C. N. & NORMAND, S. L. 1992 Hierarchical models for combining information and for meta-analyses. In *Bayesian statistics 4* (ed. J. M. Bernardo, J. O. Berger, A. P. Dawid & A. F. M. Smith), pp. 321–344. Walton Street, Oxford: Oxford University Press.
- NAGUIB, R. N. G. & SHERBET, G. V. 2001 *Artificial neural networks in cancer diagnosis, prognosis, and patient management*. CRC Press.
- O’HAGAN, A., BUCK, C. E., DANESHKHAH, A., EISER, J. R., GARTHWAITE, P. H., JENKINSON, D. J., OAKLEY, J. E. & RAKOW, T. 2006 *Uncertain Judgements: Eliciting Experts’ Probabilities*. Wiley.
- OKEN, M. M., CREECH, R. H., TORMEY, D. C., HORTON, J., DAVIS, T. E., MCFADDEN, E. T. & CARBONE, P. P. 1982 Toxicity and response criteria of the Eastern Cooperative Oncology Group. *American Journal of Clinical Oncology* **5** (6), 649–656.
- PLUMMER, M. 2012 JAGS Version 3.3.0 user manual.
- PLUMMER, M. 2013 [Http://cran.r-project.org/web/packages/rjags/rjags.pdf](http://cran.r-project.org/web/packages/rjags/rjags.pdf).
- PRIJS, M., PEELEN, L., BRESSER, P. & N.PEEK 2007 A nearest neighbor approach to predicting survival time with an application in chronic respiratory disease. *Biostatistics* **11,3**, 77–86.
- PROCTOR, S. 2000 A practical guide to continuous population-based data collection (PACE): a process facilitating uniformity of care and research into practice. *QJM* **93** (2), 67–73.
- R DEVELOPMENT CORE TEAM 2008 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.
- REN, S. & OAKLEY, J. E. 2013 Assurance calculations for planning clinical trials with time-to-event outcomes. *Statist. Med.* **33** (1), 31–45.
- REVIE, M., BEDFORD, T. & WALLS, L. 2010 Evaluation of elicitation methods to quantify Bayes linear models. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability* **224** (4), 322–332.
- RIPLEY, R. M., HARRIS, A. L. & TARASSENKO, L. 2004 Non-linear survival analysis using neural networks. *Statist. Med.* **23** (5), 825–842.

- ROBERTS, G. O. & SAHU, S. K. 1997 Updating schemes, correlation structure, blocking and parameterization for the Gibbs sampler. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **59** (2), 291–317.
- ROYSTON, P. & PARMAR, M. K. B. 2002 Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Statist. Med.* **21** (15), 2175–2197.
- RUE, H. & HELD, L. 2005 *Gaussian Markov Random Fields: Theory and Applications. Modeling in Computer Vision*. Boca Raton London, New York, Singapore: Chapman & Hall/CRC.
- SAHU, S. K., DEY, D. K., ASLANIDOU, H. & SINHA, D. 1997 A Weibull regression model with gamma frailties for multivariate survival data. *Lifetime Data Analysis* **3**, 123–137.
- SASIENI, P. 1992 Generalized additive models. T. J. Hastie and R. J. Tibshirani, Chapman and Hall, London, 1990. *Statist. Med.* **11** (7), 981–982.
- SCHMIDT, A. M. & RODRIGUEZ, M. A. 2011 Modelling multivariate counts varying continuously in space. *Bayesian Statistics* **9**, 611–638.
- SCOTT, S. L. 1999 Bayesian analysis of a two-state Markov modulated Poisson process. *Journal of Computational and Graphical Statistics* **8** (3), 662.
- SCOTT, S. L. 2002 Bayesian methods for hidden Markov models. *Journal of the American Statistical Association* **97** (457), 337–351.
- SEHN, L. H. 2005 Introduction of combined CHOP plus rituximab therapy dramatically improved outcome of diffuse large B-cell lymphoma in British Columbia. *Journal of Clinical Oncology* **23** (22), 5027–5033.
- SEHN, L. H., BERRY, B., CHHANABHAI, M., FITZGERALD, C., GILL, K., HOSKINS, P., KLASA, R., SAVAGE, K. J., SHENKIER, T. & SUTHERLAND, J. 2007 The revised international prognostic index (R-IPI) is a better predictor of outcome than the standard IPI for patients with diffuse large B-cell lymphoma treated with R-CHOP. *Blood* **109** (5), 1857–1861.
- SHIPP, M. A. AND HARRINGTON, D. P., ANDERSON, J. R. & ARMITAGE, J. O. 1993 A predictive model for aggressive non-Hodgkin's lymphoma, The international non-Hodgkin's lymphoma prognostic factors project. *New England Journal of Medicine* **329**, 987–994.

- SINHA, D. & DEY, D. K. 1997 Semiparametric Bayesian analysis of survival data. *Journal of the American Statistical Association* **92** (439), 1195.
- SOLAL-CELIGNY, P. 2004 Follicular lymphoma international prognostic index. *Blood* **104** (5), 1258–1265.
- SPIEGELHALTER, D., THOMAS, A., BEST, N. & GILKS, W. 1995 Bayesian inference using Gibbs sampling, Version 0.50. MRC Biostatistics Unit, Cambridge.
- SPIEGELHALTER, D., THOMAS, A., BEST, N. & LUNN, D. 2003 WinBUGS User Manual Version 1.4.
- SPIEGELHALTER, D. J., THOMAS, A. & BEST, N. G. 1999 WinBUGS Version 1.2 User Manual. *MRC Biostatistics Unit, Cambridge* .
- SPIEGELHALTER, D. J., THOMAS, A., BEST, N. G. & GILKS, W. R. 1996 BUGS: Bayesian inference using Gibbs sampling. *MRC Biostatistics Unit, Cambridge* **Version 0.5, (version ii)**.
- TANNER, M. A. & WONG, W. H. 1987 The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association* **82** (398), 528.
- THUNHURST, C. 1988 Book reviews : Health and deprivation: inequality and the North Peter Townsend, Peter Phillimore and Alistair Beattie Croom Helm, London. *Critical Social Policy* **8** (23), 129–133.
- TILLY, H. & DREYLING, M. 2009 Diffuse large B-cell non-Hodgkin’s lymphoma: ESMO clinical recommendations for diagnosis, treatment and follow-up. *Annals of Oncology* **20** (Supplement 4), iv110–iv112.
- VAUPEL, J. W., MANTON, K. G. & STALLARD, E. 1979 The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* **16** (3), 439.
- VICKERS, M. 1999 Wintrobe’s Clinical Haematology, 10th edn. . Edited by G. R. Lee, J. Foerster, J. Lukons, F. Paraskevar, J. P. Greer and G. M. Rodgers. Williams and Wilkins, Baltimore, 1999. Pp. 2764. f120. *British Journal of Haematology* **105** (2), 572–572.
- WEIBULL, W. 1951 A statistical distribution function of wide applicability. *Journal of Applied Mechanics* **18**, 293–297.
- WEST, M. 1992 Modelling time-varying hazards and covariate effects in survival analysis:. *Survival Analysis: State of the Art* **211**, 47–62.

- WEST, M. & HARRISON, J. 1997 *Bayesian forecasting and dynamic models*. New York: Springer.
- WILSON, K. J. & FARROW, M. 2010 Bayes linear kinematics in the analysis of failure rates and failure time distributions. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability* **224** (4), 309–321.
- WINKLER, R. L. 1967 The assessment of prior distributions in Bayesian analysis. *Journal of the American Statistical Association* **62**, 776–800.
- YAMAMOTO, J. F. & GOODMAN, M. T. 2007 Patterns of leukemia incidence in the United States by subtype and demographic characteristics. *Cancer Causes & Control* **19** (4), 379–390.
- YAN, X. 2013 Weighted K-nearest neighbor classification algorithm based on genetic algorithm. *TELKOMNIKA Indonesian Journal of Electrical Engineering* **11** (10).
- ZAPATA-VAZQUEZ, R. E., O'HAGAN, A. & SOARES BASTOS, L. 2014 Eliciting expert judgements about a set of proportions. *Journal of Applied Statistics* **41** (9), 1919–1933.
- ZHAO, X. 2010 Bayesian survival analysis for prognostic index development with many covariates and missing data. PhD thesis, School of Mathematics and Statistics, Newcastle University.
- ZHOU, Z., SEHN, L. H., RADEMAKER, A. W., GORDON, L. I., LACASCE, A. S., CROSBY-THOMPSON, A., VANDERPLAS, A., ZELENETZ, A. D., ABEL, G. A., RODRIGUEZ, M. A., NADEMANEE, A., KAMINSKI, M. S., CZUCZMAN, M. S., MILLENSON, M., NILAND, J., GASCOYNE, R. D., CONNORS, J. M., FRIEDBERG, J. W. & WINTER, J. N. 2014 An enhanced international prognostic index NCCN-IPI for patients with diffuse large B-cell lymphoma treated in the rituximab era. *Blood* **123**, 837–842.