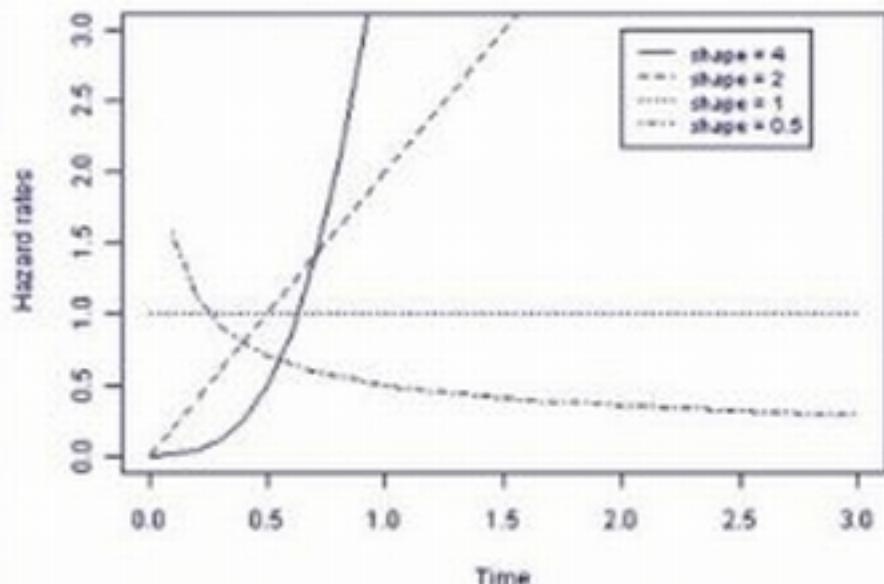


# MODELING SURVIVAL DATA USING FRAILTY MODELS

Plot of Hazard Rates of Weibull



**David D. Hanagal**



CRC Press  
Taylor & Francis Group

A CHAPMAN & HALL BOOK

# **MODELING SURVIVAL DATA USING FRAILTY MODELS**



# **MODELING SURVIVAL DATA USING FRAILTY MODELS**

**David D. Hanagal**



**CRC Press**  
Taylor & Francis Group  
Boca Raton London New York

---

CRC Press is an imprint of the  
Taylor & Francis Group an **informa** business  
A CHAPMAN & HALL BOOK

Chapman & Hall/CRC  
Taylor & Francis Group  
6000 Broken Sound Parkway NW, Suite 300  
Boca Raton, FL 33487-2742

© 2011 by Taylor and Francis Group, LLC  
Chapman & Hall/CRC is an imprint of Taylor & Francis Group, an Informa business

No claim to original U.S. Government works

Printed in the United States of America on acid-free paper  
10 9 8 7 6 5 4 3 2 1

International Standard Book Number: 978-1-4398-3667-5 (Hardback)

This book contains information obtained from authentic and highly regarded sources. Reasonable efforts have been made to publish reliable data and information, but the author and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The authors and publishers have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged please write and let us know so we may rectify in any future reprint.

Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access [www.copyright.com](http://www.copyright.com) (<http://www.copyright.com/>) or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

**Trademark Notice:** Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Visit the Taylor & Francis Web site at  
<http://www.taylorandfrancis.com>

and the CRC Press Web site at  
<http://www.crcpress.com>

---

**Dedicated to the Memory of My Loving Father**



---

# Contents

<b>List of Tables</b>	<b>xiii</b>
<b>List of Figures</b>	<b>xv</b>
<b>Preface</b>	<b>xvii</b>
<b>About the Author</b>	<b>xix</b>
<b>I Basic Concepts in Survival Analysis</b>	<b>1</b>
<b>1 Introduction to Survival Analysis</b>	<b>3</b>
1.1 Introduction . . . . .	3
1.2 Bone Marrow Transplantation (BMT) for Leukemia . . . . .	4
1.3 Remission Duration from a Clinical Trial for Acute Leukemia	5
1.4 Times of Infection of Kidney Dialysis Patients . . . . .	7
1.5 Kidney Infection Data . . . . .	8
1.6 Litters of Rats Data . . . . .	8
1.7 Kidney Dialysis (HLA) Patients Data . . . . .	8
1.8 Diabetic Retinopathy Data . . . . .	10
1.9 Myeloma Data . . . . .	12
1.10 Definitions and Notations . . . . .	13
1.10.1 Survival Function . . . . .	13
1.10.2 Failure (or Hazard) Rate . . . . .	13
1.11 Censoring . . . . .	16
1.11.1 Censored Type I Data . . . . .	16
1.11.2 Censored Type II Data . . . . .	17
1.11.3 Readout or Interval Censored Data . . . . .	17
1.11.4 Multicensored Data . . . . .	17
1.11.5 Separating out Failure Modes . . . . .	18
<b>2 Some Parametric Methods</b>	<b>19</b>
2.1 Introduction . . . . .	19
2.2 Exponential Distribution . . . . .	20
2.3 Weibull Distribution . . . . .	21
2.4 Extreme Value Distributions . . . . .	23
2.5 Lognormal . . . . .	25
2.6 Gamma . . . . .	26

2.7	Loglogistic . . . . .	29
2.8	Maximum Likelihood Estimation . . . . .	30
2.9	Parametric Regression Models . . . . .	35
2.9.1	Example 2.1 . . . . .	40
2.9.2	Example 2.2 . . . . .	41
<b>3</b>	<b>Nonparametric and Semiparametric Models</b>	<b>43</b>
3.1	Empirical Survival Function . . . . .	43
3.2	Graphical Plotting . . . . .	44
3.2.1	Probability Plotting . . . . .	44
3.2.2	Hazard and Cumulative Hazard Plotting . . . . .	47
3.2.3	Exponential and Weibull Hazard Plots . . . . .	48
3.3	Graphical Estimation . . . . .	50
3.4	Empirical Model Fitting: Distribution Free (Kaplan-Meier) Approach . . . . .	51
3.5	Comparison between Two Survival Functions . . . . .	59
3.6	Cox's Proportional Hazards Model . . . . .	62
<b>II</b>	<b>Univariate and Shared Frailty Models for Survival Data</b>	<b>69</b>
<b>4</b>	<b>The Frailty Concept</b>	<b>71</b>
4.1	Introduction . . . . .	71
4.2	The Definition of Shared Frailty . . . . .	73
4.3	The Implications of Frailty . . . . .	74
4.4	The Conditional Parametrization . . . . .	76
4.5	The Marginal Parametrization . . . . .	77
4.6	Frailty as a Model for Omitted Covariates . . . . .	78
4.7	Frailty as a Model of Stochastic Hazard . . . . .	78
4.8	Identifiability of Frailty Models . . . . .	79
<b>5</b>	<b>Various Frailty Models</b>	<b>81</b>
5.1	Introduction . . . . .	81
5.2	Gamma Frailty . . . . .	81
5.3	Positive Stable Frailty . . . . .	83
5.4	Power Variance Function Frailty . . . . .	85
5.5	Compound Poisson Frailty . . . . .	86
5.6	Compound Poisson Distribution with Random Scale . . . . .	89
5.7	Frailty Models in Hierarchical Likelihood . . . . .	92
5.8	Frailty Models in Mixture Distributions . . . . .	94
5.8.1	Gamma Frailty in Weibull Mixture . . . . .	95
5.8.2	Positive Stable Frailty in Weibull Mixture . . . . .	96
5.8.3	PVF Frailty in Weibull Mixture . . . . .	96
5.9	Piecewise Gamma Frailty Model . . . . .	97
5.9.1	Frailty Models and Dependence Function . . . . .	99
5.9.2	Example: Epilepsy Data . . . . .	102

<b>6 Estimation Methods for Shared Frailty Models</b>	<b>105</b>
6.1 Introduction . . . . .	105
6.2 Inference for the Shared Frailty Model . . . . .	106
6.3 The EM Algorithm . . . . .	108
6.4 The Gamma Frailty Model . . . . .	110
6.5 The Positive Stable Frailty Model . . . . .	111
6.6 The Lognormal Frailty Model . . . . .	113
6.6.1 Application to Seizure Data . . . . .	113
6.7 Modified EM (MEM) Algorithm for Gamma Frailty Models	114
6.8 Application . . . . .	116
6.9 Discussion . . . . .	117
<b>7 Analysis of Survival Data in Shared Frailty Models</b>	<b>119</b>
7.1 Introduction . . . . .	119
7.2 Analysis for Bone Marrow Transplantation (BMT) Data . . . . .	119
7.3 Analysis for Acute Leukemia Data . . . . .	122
7.4 Analysis for HLA Data . . . . .	125
7.5 Analysis for Kidney Infection Data . . . . .	129
7.6 Analysis of Litters of Rats . . . . .	131
7.7 Analysis for Diabetic Retinopathy Data . . . . .	133
<b>8 Tests of Hypotheses in Frailty Models</b>	<b>137</b>
8.1 Introduction . . . . .	137
8.2 Tests for Gamma Frailty Based on Likelihood Ratio and Score Tests . . . . .	138
8.2.1 The Model and the Main Results . . . . .	139
8.2.2 Analysis of Diabetic Retinopathy . . . . .	142
8.3 Logrank Tests for Testing $\beta = 0$ . . . . .	143
8.3.1 Notations and Review . . . . .	144
8.3.2 Parametric Tests for Uncensored Samples . . . . .	146
8.3.3 Nonparametric Tests for Uncensored Samples . . . . .	148
8.3.4 Effect of Censoring . . . . .	151
8.3.5 Some Numerical Examples . . . . .	154
8.4 Test for Heterogeneity in Kidney Infection Data . . . . .	155
8.4.1 Models and Methods . . . . .	157
<b>9 Shared Frailty in Bivariate Exponential and Weibull Models</b>	<b>161</b>
9.1 Introduction . . . . .	161
9.2 Bivariate Exponential Distributions . . . . .	162
9.2.1 Marshall-Olkin (M-O): . . . . .	162
9.2.2 Block-Basu (B-B): . . . . .	163
9.2.3 Freund . . . . .	163
9.2.4 Proschan-Sullo (P-S): . . . . .	164
9.3 Gamma Frailty in BVW Models . . . . .	165
9.3.1 Weibull Extension of BVE of Gumbel . . . . .	165

9.3.2	Weibull Extension of BVE of Marshall-Olkin . . . . .	169
9.3.3	Weibull Extension of BVE of Block-Basu . . . . .	170
9.3.4	Weibull Extension of BVE of Freund and Proschan-Sullo . . . . .	171
9.4	Positive Stable Frailty in BVW Models . . . . .	173
9.4.1	Weibull Extension of BVE of Gumbel . . . . .	173
9.4.2	Weibull Extension of BVE of Marshall-Olkin . . . . .	175
9.4.3	Weibull Extension of BVE of Block-Basu . . . . .	175
9.4.4	Weibull Extension of BVE of Freund and Proschan-Sullo . . . . .	176
9.5	Power Variance Function Frailty in BVW Models . . . . .	176
9.5.1	Weibull Extension of BVE Models . . . . .	176
9.6	Lognormal and Weibull Frailties in BVW Models . . . . .	178
9.6.1	Estimation of Parameters . . . . .	178
9.7	Compound Poisson Frailty in BVW Models . . . . .	180
9.8	Compound Poisson (with Random Scale) Frailty in BVW Models . . . . .	180
9.9	Estimation and Tests for Frailty under BVW Baseline . . . . .	181
<b>10</b>	<b>Frailty Models Based on Lévy Processes</b>	<b>185</b>
10.1	Introduction . . . . .	185
10.1.1	Biological Interpretation of Failure Rate . . . . .	186
10.1.2	A Model for Random Failure Rate Processes . . . . .	187
10.2	Lévy Processes and Subordinators . . . . .	189
10.2.1	Standard Compound Poisson Process . . . . .	190
10.2.2	Compound Poisson Process with General Jump Distribution . . . . .	191
10.2.3	Gamma Processes . . . . .	191
10.2.4	Stable Processes . . . . .	191
10.2.5	PVF Processes . . . . .	191
10.2.6	Special Cases . . . . .	192
10.3	Proportional Hazards Derived from Lévy Processes . . . . .	193
10.3.1	Example 10.1: Gamma Process . . . . .	193
10.3.2	Example 10.2: Compound Poisson Process . . . . .	194
10.4	Other Frailty Process Constructions . . . . .	195
10.5	Hierarchical Lévy Frailty Models . . . . .	196
10.5.1	Application to the Infant Mortality Data . . . . .	200
<b>III</b>	<b>Bivariate Frailty Models for Survival Data</b>	<b>203</b>
<b>11</b>	<b>Bivariate Frailty Models and Estimation Methods</b>	<b>205</b>
11.1	Introduction . . . . .	205
11.2	Bivariate Frailty Models and Laplace Transforms . . . . .	206
11.3	Proportional Hazard Model for Covariate Effects . . . . .	207
11.4	The Problem of Confounding . . . . .	207

11.5 A General Model of Covariate Dependence . . . . .	208
11.6 Pseudo-Frailty Model . . . . .	210
11.7 Likelihood Construction . . . . .	211
11.8 Semiparametric Representations . . . . .	212
11.9 Estimation Methods in Bivariate Frailty Models . . . . .	214
11.9.1 Two-Stage Estimation Method . . . . .	214
11.9.2 Basegroup Estimation Method . . . . .	214
11.9.3 Estimation Methods Based on the EM Algorithm . .	215
11.9.4 Profile Estimation for Frailty Models . . . . .	217
11.9.5 Profile Estimation for Transformation Models . .	219
11.9.6 Conclusions . . . . .	221
<b>12 Correlated Frailty Models</b> . . . . .	<b>223</b>
12.1 Introduction . . . . .	223
12.2 Correlated Gamma Frailty Model . . . . .	225
12.3 Correlated Power Variance Function Frailty Model . .	225
12.4 Genetic Analysis of Duration . . . . .	226
12.4.1 Correlation Coefficient . . . . .	227
12.4.2 Six Genetic Models of Frailty . . . . .	229
12.4.3 Danish Twins Survival Data . . . . .	230
12.4.4 Results: Genetics of Frailty and Longevity . . .	231
12.5 General Bivariate Frailty Model . . . . .	233
12.5.1 Gamma Model . . . . .	235
12.5.2 Lognormal Model . . . . .	236
12.5.3 Estimation Strategies . . . . .	237
12.6 Correlated Compound Poisson Frailty for the Bivariate Survival Lifetimes . . . . .	238
12.7 Applications . . . . .	240
<b>13 Additive Frailty Models</b> . . . . .	<b>243</b>
13.1 Introduction . . . . .	243
13.2 Modeling Multivariate Survival Data Using the Frailty Model . .	245
13.3 Correlated Frailty Model . . . . .	246
13.4 Relations to Other Frailty Models . . . . .	247
13.4.1 Shared Frailty Model . . . . .	248
13.4.2 Over-Dispersion Model . . . . .	248
13.4.3 Twin Model . . . . .	248
13.4.4 Litter Model . . . . .	249
13.4.5 Genetic Model . . . . .	249
13.4.6 Adoption Model . . . . .	250
13.4.7 Competing Risks . . . . .	250
13.4.8 Example . . . . .	251
13.5 Additive Genetic Gamma Frailty . . . . .	255
13.5.1 Example 13.1 . . . . .	255
13.5.2 Example 13.2 . . . . .	256

13.5.3 Example 13.3 . . . . .	257
13.5.4 Application in Danish Adoptive Register Data . . . . .	259
13.6 Additive Genetic Gamma Frailty for Linkage Analysis of Diseases . . . . .	263
13.6.1 Genetic Frailties Defined by Multiple Unlinked Disease Loci . . . . .	265
13.6.2 Expected Genetic Frailties over the Inheritance Vectors	266
13.6.3 Model for Age of Onset Data in a Family . . . . .	268
13.6.4 Conditional Hazards Ratio for Sib Pairs . . . . .	268
13.6.5 Estimation Methods and Test of Linkage . . . . .	270
13.6.6 Breast Cancer Data Example . . . . .	270
<b>14 Identifiability of Bivariate Frailty Models</b>	<b>273</b>
14.1 Introduction . . . . .	273
14.2 Identifiability of Bivariate Frailty Models . . . . .	275
14.3 Identifiability of Correlated Frailty Models . . . . .	276
14.4 Non-Identifiability of Frailty Models without Observed Covariates . . . . .	277
14.4.1 Bivariate Frailty Models with Infinite Mean . . . . .	277
14.4.2 Bivariate Frailty Models with Finite Mean . . . . .	278
14.5 Discussion . . . . .	279
<b>Appendix</b>	<b>281</b>
<b>Bibliography</b>	<b>291</b>
<b>Index</b>	<b>313</b>

---

## List of Tables

1.1	Bone marrow transplantation data . . . . .	6
1.2	Remission duration from a clinical trial for acute leukemia . . . . .	6
1.3	Times to infection of kidney dialysis patients . . . . .	7
1.4	Kidney infection data . . . . .	9
1.5	Litters of rats data . . . . .	10
1.6	Kidney dialysis patients data . . . . .	11
1.7	Diabetic retinopathy data . . . . .	12
1.8	Myeloma data . . . . .	13
3.1	Empirical CDF calculations . . . . .	46
3.2	Empirical cumulative hazard calculations . . . . .	48
6.1	Parameter estimates of $(\beta, \theta)$ with standard errors . . . . .	116
8.1	Diabetic retinopathy study: Parameter estimates (standard errors) . . . . .	142
8.2	Regression estimates (standard error) from Model I and Model II . . . . .	158
8.3	Regression estimates (standard error) with patient 21 removed . . . . .	158
9.1	Some properties of bivariate exponential distributions, ‘exp’ means exponential distribution . . . . .	164
12.1	Genetic and environmental components of frailty for Danish male twins survived to age 30 . . . . .	231
12.2	Genetic and environmental components of frailty for Danish female twins survived to age 30 . . . . .	232
13.1	Parameter estimates in correlated frailty model of death by infection . . . . .	253
13.2	Parameter estimates in correlated frailty model with a purely genetic effect . . . . .	253
13.3	Estimation of frailty parameter in the additive frailty model . . . . .	260
13.4	Estimation of frailty parameter in the additive frailty model of fathers (mothers) and children . . . . .	261
13.5	Estimation of frailty parameter in the correlated frailty model of fathers (mothers) and children . . . . .	262



---

## ***List of Figures***

1.1	Bathtub failure rate curve. . . . .	15
1.2	Readout or interval data diagram. . . . .	17
2.1	PDF and survival function graphs of exponential distribution.	20
2.2	Graphs of PDF of Weibull distribution with $\lambda = 1$ . . . . .	22
2.3	Failure rate curves of Weibull distribution with $\lambda = 1$ . . . . .	23
2.4	Weibull distribution: $\ln S(t)$ vs. time with $\lambda = 1$ . . . . .	24
2.5	Lognormal density curves with $\mu = 0$ . . . . .	26
2.6	PDF curves of the lognormal distribution with $\sigma = 1$ . . . . .	27
2.7	Hazard rate of the lognormal with different parameters. . . . .	28
2.8	Gamma density functions with $\lambda = 1$ . . . . .	29
2.9	Gamma density functions with $\gamma = 3$ . . . . .	30
2.10	Gamma failure rate functions with $\lambda = 1$ . . . . .	31
2.11	Density function of loglogistic distribution with $\alpha = 1$ . . . . .	32
2.12	Survival function of loglogistic distribution with $\alpha = 1$ . . . . .	33
2.13	Failure rate function of loglogistic distribution with $\alpha = 1$ . . . . .	34
3.1	Graphical Weibull fitting plot. . . . .	47
3.2	Graphical exponential fitting of the data. . . . .	49
3.3	Graphical Weibull fitting of the data. . . . .	49
3.4	K-M and Breslow(or N-A) estimates for kidney infection data set with 95% CI for K-M. . . . .	55
3.5	Graph of the survival function of three groups for BMT data set. . . . .	56
3.6	Graph of the $-\ln S(t)$ versus $t$ of three groups for BMT data set. . . . .	57
3.7	Graph of the $\ln[-\ln S(t)]$ versus $\ln t$ of three groups for BMT data set. . . . .	58
3.8	Graph of the $\ln[-\ln S(t)]$ versus $\ln t$ of three groups for BMT data set. . . . .	59
3.9	K-M curve with 95% CI for myeloma data. . . . .	67
13.1	Father, mother, and child genetic diagram. . . . .	256
13.2	Father, mother, and 2 children genetic diagram. . . . .	257



---

## Preface

There are several books on survival data, that is, data concerning the time to some event. In the standard case, the event is death, but the topic is much broader. This book, however, covers frailty models for the survival data. So far, there are only one or two books devoted to frailty models in spite of several new research developments and applications of frailty models. This book is an attempt to cover the recent topics that are not available in the existing books. The topics discussed in this book have wide applications and are useful for scientists, research students, and teachers. There are numerous research papers on this topic. I managed to collect most of these papers from the last four years. Readers may find one or two chapters on frailty models in existing survival analysis books and this topic is not discussed in depth. I take this opportunity to convey the recent methodology and applications of frailty models to all those who are working in this area. This book aims to present the basic concepts of survival analysis and frailty models ranging from fundamental to advanced topics. Eight data sets are discussed in this book. We carry out the analysis of these data sets using the R statistical package. This R statistical package is free open access to everyone. Every three to four months one can find the updated version of R statistical package on the Web site: <http://cran.r-project.org/bin/windows/base/>.

The first three chapters are devoted to basic concepts in survival analysis. Chapters 4 to 10 are based on shared frailty models and Chapters 11 to 14 are based on bivariate frailty models. Chapter 1 presents eight data sets on survival times with covariates and an introduction to survival analysis. Chapter 2 presents some important parametric distributions and their corresponding regression models with examples using R. Chapter 3 presents nonparametric Kaplan-Meier estimation and Cox's proportional hazard model, graphical plotting with examples using R. Chapter 4 gives the concept of frailty and Chapter 5 presents important frailty models. Chapter 6 presents different estimation procedures such as EM and modified EM algorithms. Chapter 7 gives data analysis of six data sets discussed in Chapter 1 using R. Chapter 8 presents logrank tests and cusum of chi-square tests for testing frailty. Chapter 9 gives shared frailty models in different bivariate exponential and bivariate Weibull distributions. Chapter 10 presents frailty models based on Lévy's processes. Chapter 11 presents different estimation procedures in bivariate frailty models and Chapter 12 gives correlated gamma frailty, correlated lognormal frailty and correlated power variance function frailty models.

and their estimation procedures. Chapter 13 presents additive frailty models with three examples and Chapter 14 gives identifiability of bivariate frailty and correlated frailty models. Some data sets are in the Appendix.

Finally, I am fully responsible for any errors remaining in this book. The views expressed in this book are those of the authors noted in the bibliography. Readers are encouraged to give suggestions and comments to further improve this book. They can be communicated to the author by email: david\_hanagal@yahoo.co.in.

## Acknowledgments

I thank the Council of Scientific and Industrial Research, New Delhi for providing financial assistance to write this book under the major research project number 25(0176)/09/EMR-II. I thank the reviewers for their constructive suggestions and comments. I thank H. L. Koul, Professor, Department of Statistics and Probability, Michigan State University, USA, where I began to work on frailty models in depth, which motivated me to write a book on frailty models and where I worked as a visiting professor from August 2006 to May 2007. I thank B. K. Kale, retired professor, Department Statistics, University of Pune, India for valuable suggestions. I also thank my wife, Anjali, for her support while writing this book.

David D. Hanagal

---

## ***About the Author***

David D. Hanagal is a professor of statistics at the University of Pune, Pune, India. He has 30 years of teaching experience and 22 years of research experience to date. He was elected a fellow of the Royal Statistical Society, London, UK in 2003. He is an editorial board member of international journals and has published 70 research publications in national and international journals. He has developed skills in writing several programs in SAS, R, MATLAB®, MINITAB, SPSS, and SPLUS. Hanagal worked as a visiting professor at several universities abroad (USA, Germany, and Mexico). He has presented several talks during national and international conferences all over the world. His research interests are statistical inference, selection problems, reliability, survival analysis, frailty models, Bayesian inference, stress-strength models, Monte Carlo methods, MCMC algorithms, bootstrapping, censoring schemes, distribution theory, multivariate models, characterization, and nonparametric inference.

## **Part I**

# **Basic Concepts in Survival Analysis**



# Chapter 1

---

## *Introduction to Survival Analysis*

1.1	Introduction .....	3
1.2	Bone Marrow Transplantation (BMT) for Leukemia .....	4
1.3	Remission Duration from a Clinical Trial for Acute Leukemia .....	5
1.4	Times of Infection of Kidney Dialysis Patients .....	7
1.5	Kidney Infection Data .....	8
1.6	Litters of Rats Data .....	8
1.7	Kidney Dialysis (HLA) Patients Data .....	8
1.8	Diabetic Retinopathy Data .....	9
1.9	Myeloma Data .....	11
1.10	Definitions and Notations .....	12
1.10.1	Survival Function .....	13
1.10.2	Failure (or Hazard) Rate .....	13
	Bivariate and Multivariate Survival Function .....	15
1.11	Censoring .....	16
1.11.1	Censored Type I Data .....	16
1.11.2	Censored Type II Data .....	17
1.11.3	Readout or Interval Censored Data .....	17
1.11.4	Multicensored Data .....	17
1.11.5	Separating out Failure Modes .....	18

---

### 1.1 Introduction

**Survival data** is a term used for describing data that measure the time to a certain event. In survival analysis, the event may be death, occurrence of disease (or complication), time to an epileptic seizure, time it takes for a patient to respond to a therapy, or time from response until disease relapse (i.e., disease returns). In demography, the event can be entering marriage. The event is a transition from one state to another. Death is a transition from state alive to the state dead. Occurrence of disease is a transition from the state of being healthy to a state of presence of disease. Marriage is the transition of being unmarried to being married. For epileptic seizure, an event is a transition from the seizure-free state to the state of active seizure.

Time to an event is a positive real valued variable having continuous distribution. It is necessary to define the starting time point, say 0, from which times are measured. When we measure age, the starting time point may be date of birth. For studying the occurrence of disease, the time scale is the known duration of the disease. Here the known duration is the time since di-

agnosis of the disease. For the drug trial, the starting time is the time of start of treatment.

The problem of analyzing time to event data arises in a number of applied fields, such as medicine, biology, public health, epidemiology, engineering, economics, and demography. Although the statistical tools we present are applicable to all these disciplines, our focus is on applying the statistical tools to biology and medicine. In this chapter, we present some examples drawn from these fields that are used throughout the text to illustrate the statistical techniques.

In biomedical applications the data are collected over a finite period of time and consequently the ‘time to event’ may not be observed for all the individuals in our study population (sample). This results in what is called ‘**censored**’ data. That is, the ‘time to event’ for those individuals who have not experienced the event under study is **censored** (by the end of study). It is also common that the amount of follow-up for the individuals in a sample vary from subject to subject. The combination of censoring and differential follow-up creates some unusual difficulties in the analysis of such data that cannot be handled properly by the standard statistical methods. Because of this, a research area in statistics has emerged which is called ‘**Survival Analysis**’.

A common feature of these data sets is that they contain either censored or truncated observations. Censored data arises when an individual’s life length is known to occur only in a certain period of time. Well-known censoring schemes are *right censoring*, where all that is known is that the individual is still alive at a given time; *left censoring*, when all that is known is that the individual has experienced the event of interest prior to the start of the study; or *interval censoring*, where the only information is that the event occurs within some interval. Truncation schemes are *left truncation*, where only individuals who survive a sufficient time are included in the sample and *right truncation*, where only individuals who have experienced the event by a specified time are included in the sample.

---

## 1.2 Bone Marrow Transplantation (BMT) for Leukemia

Data set on bone marrow transplantation for Leukemia of 137 patients at the Ohio State University Bone Marrow Transplantation Unit was studied by Avalos et al. (1993). The data set is also available in Klein and Moeschberger (2003). The data is presented in Table A.1 in the Appendix and partial data is given in [Table 1.1](#). Bone marrow transplant treatment is a standard method for acute Leukemia. Patients with three disease types are acute myelocytic leukemia (AML) with low and high risks and acute lymphoblastic leukemia (ALL). Prognosis for recovery may depend on risk factors known at the time of transplantation, such as patient’s age and sex (1-Male, 0-Female). Patients

may develop acute or chronic disease. Patient's age, sex, and disease group (1-ALL, 2-AML low risk, 3-AML high risk) were considered as covariates. The following is the description of [Table 1.1](#) with only six patients.

```

g--Disease Group
  1-ALL
  2-AML Low Risk
  3-AML High Risk
T1 -- Time to Death or on Study Time
T2 -- Disease Free Survival Time (Time to Relapse, Death or End of Study)
I1 -- Death Indicator
  1-Dead 0-Alive
I2 -- Relapse Indicator
  1-Relapsed, 0-Disease Free
I3--Disease Free Survival Indicator
  (1-Dead or Relapsed, 0-Alive Disease Free)
TA--Time to Acute Graft-Versus-Host Disease
IA--Acute GVHD Indicator
  (1-Developed Acute GVHD 0-Never Developed Acute GVHD)
TC--Time to Chronic Graft-Versus-Host Disease
IC--Chronic GVHD Indicator
  1-Developed Chronic GVHD 0-Never Developed Chronic GVHD
TP--Time to Return of Platelets to Normal Levels
IP--Platelet Recovery Indicator
  1-Platelets Returned to Normal, 0-Platelets Never Returned to Normal
Z1--Patient Age in Years
Z2--Donor Age in Years
Z3--Patient Sex
  1-Male, 0-Female
Z4--Donor Sex
  1-Male, 0-Female
Z5--Patient CMV Status
  1-CMV Positive, 0-CMV Negative
Z6--Donor CMV Status
  1-CMV Positive, 0-CMV Negative
Z7--Waiting Time to Transplant in Days
Z8--FAB
  1-FAB Grade 4 or 5 and AML, 0-Otherwise
Z9--Hospital
  1-The Ohio State University, 2-Alferd , 3-St. Vincent, 4-Hahnemann
Z10--MTX Used as a Graft-Versus-Host- Prophylactic 1-Yes 0-No

```

---

### 1.3 Remission Duration from a Clinical Trial for Acute Leukemia

Freireich et al. (1963) report the results of a clinical trial of a drug 6-mecaptopurine (6-MP) versus a placebo in 21 children with acute leukemia.

g	T1	T2	I1	I2	I3	TA	IA	TC	IC	TP	IP	Z1	Z2	Z3	Z4	Z5	Z6	Z7	Z8	Z9	Z10
1	2081	2081	0	0	0	67	1	121	1	13	1	26	33	1	0	1	1	98	0	1	0
1	1602	1602	0	0	0	1602	0	139	1	18	1	21	37	1	1	0	0	1720	0	1	0
1	1496	1496	0	0	0	1496	0	307	1	12	1	26	35	1	1	1	0	127	0	1	0
1	1462	1462	0	0	0	70	1	95	1	13	1	17	21	0	1	0	0	168	0	1	0
1	1433	1433	0	0	0	1433	0	236	1	12	1	32	36	1	1	1	1	93	0	1	0
1	1377	1377	0	0	0	1377	0	123	1	12	1	22	31	1	1	1	1	2187	0	1	0

TABLE 1.1: Bone marrow transplantation data

The trial was conducted at 11 American hospitals. Patients were selected who had a complete or partial remission of their leukemia induced by treatment with the drug prednisone. A complete or partial remission means that either most or all signs of disease has disappeared from the bone marrow. The trial was conducted by matching pairs of patients at a given hospital by remission status (complete or partial) and randomizing within the pair to either a 6-MP or placebo maintenance therapy. Patients were followed until their leukemia returned (relapse) or until the end of the study (in months). The data is reported in Table 1.2.

col 1: Pair

col 2: Remission status at randomization(S) (1=partial, 2=complete)

col 3: Time to relapse for placebo patients, months(TP)

col 4: Time to relapse for 6-MP patients, months(T6)

col 5: Relapse indicator (RI) (0=censored, 1=relapse) for 6-MP patients

Note: All placebo patients relapsed.

Pair	S	TP	T6	RI	Pair	S	TP	T6	RI
1	1	1	10	1	12	1	5	20	0
2	2	22	7	1	13	2	4	19	0
3	2	3	32	0	14	2	15	6	1
4	2	12	23	1	15	2	8	17	0
5	2	8	22	1	16	1	23	35	0
6	1	17	6	1	17	1	5	6	1
7	2	2	16	1	18	2	11	13	1
8	2	11	34	0	19	2	4	9	0
9	2	8	32	0	20	2	1	6	0
10	2	12	25	0	21	2	8	10	0
11	2	2	11	0					

TABLE 1.2: Remission duration from a clinical trial for acute leukemia

## 1.4 Times of Infection of Kidney Dialysis Patients

Nahman et al. (1992) reported the data on time to first exit-site infection (in months) in patients with renal insufficiency, 43 patients utilized a surgical placed catheter (Group 1), and 76 patients utilized a percutaneous placement of their catheter (Group 2). The data is reported in Table 1.3.

col 1: Time (in months) to Infection(T)

col 2: Infection Indicator(I) (0-No, 1-Yes)

col 3: Catheter Placement(P) (1-Surgically, 2-Percutaneously)

T	I	P	T	I	P	T	I	P	T	I	P
1.5	1	1	10.5	0	1	0.5	0	2	8.5	0	2
3.5	1	1	11.5	0	1	0.5	0	2	8.5	0	2
4.5	1	1	12.5	0	1	0.5	0	2	8.5	0	2
4.5	1	1	12.5	0	1	0.5	0	2	9.5	0	2
5.5	1	1	13.5	0	1	1.5	0	2	9.5	0	2
8.5	1	1	14.5	0	1	1.5	0	2	10.5	0	2
8.5	1	1	14.5	0	1	1.5	0	2	10.5	0	2
9.5	1	1	21.5	0	1	1.5	0	2	10.5	0	2
10.5	1	1	21.5	0	1	2.5	0	2	11.5	0	2
11.5	1	1	22.5	0	1	2.5	0	2	11.5	0	2
15.5	1	1	22.5	0	1	2.5	0	2	12.5	0	2
16.5	1	1	25.5	0	1	2.5	0	2	12.5	0	2
18.5	1	1	27.5	0	1	2.5	0	2	12.5	0	2
23.5	1	1	0.5	1	2	3.5	0	2	12.5	0	2
26.5	1	1	0.5	1	2	3.5	0	2	14.5	0	2
2.5	0	1	0.5	1	2	3.5	0	2	14.5	0	2
2.5	0	1	0.5	1	2	3.5	0	2	16.5	0	2
3.5	0	1	0.5	1	2	3.5	0	2	16.5	0	2
3.5	0	1	0.5	1	2	4.5	0	2	18.5	0	2
3.5	0	1	2.5	1	2	4.5	0	2	19.5	0	2
4.5	0	1	2.5	1	2	4.5	0	2	19.5	0	2
5.5	0	1	3.5	1	2	5.5	0	2	19.5	0	2
6.5	0	1	6.5	1	2	5.5	0	2	20.5	0	2
6.5	0	1	15.5	1	2	5.5	0	2	22.5	0	2
7.5	0	1	0.5	0	2	5.5	0	2	24.5	0	2
7.5	0	1	0.5	0	2	5.5	0	2	25.5	0	2
7.5	0	1	0.5	0	2	6.5	0	2	26.5	0	2
7.5	0	1	0.5	0	2	7.5	0	2	26.5	0	2
8.5	0	1	0.5	0	2	7.5	0	2	28.5	0	2
9.5	0	1	0.5	0	2	7.5	0	2			

TABLE 1.3: Times to infection of kidney dialysis patients

## 1.5 Kidney Infection Data

McGilchrist and Aisbett(1991) reported kidney infection data for the first and second recurrence times (in days) of infections of 38 kidney patients following insertion of a catheter until it has to be removed owing to infection. The catheter may have to be removed for reasons other than infection, and we regard this as censoring. [Table 1.4](#) shows the first and second recurrence times with recurrence indicator variable (0-censored, 1-recurrence) and covariates age, sex (0-male, 1-female), and three indicator variables GN, AN, and PKD.

## 1.6 Litters of Rats Data

Mantel and et al. (1979) reported the data on litters of rats. There are three rats per litter, one for treatment and two for control. [Table 1.5](#) contains four columns, the first column is the litter number (even litter numbers are male rats, odd litter numbers are female), the second column is treatment indicator, the third column is follow up time and the last column is status (0-tumor, 1-censored due to animal death).

## 1.7 Kidney Dialysis (HLA) Patients Data

Batchelor and Hackett (1970) reported 16 cases of badly burned patients. Because different patients had varying number of grafts in the study, it will be referred to as the ‘unbalanced data set’. Each patient was given a number of skin grafts from variety of donors, and the time to graft rejection was recorded. The covariate was the quality of kidney dialysis (HAL) matching, indicated by 1 for good and 0 for poor. The censoring indicator is 1 for censored and 0 when graft was rejected. [Table 1.6](#) shows these data.

Pat	Time1	Ind1	Time2	Ind2	Age	Sex	GN	AN	PKD
1	8	1	16	1	28	0	0	0	0
2	23	1	13	0	48	1	1	0	0
3	22	1	28	1	32	0	0	0	0
4	447	1	318	1	31.5	1	0	0	0
5	30	1	12	1	10	0	0	0	0
6	24	1	245	1	16.5	1	0	0	0
7	7	1	9	1	51	0	1	0	0
8	511	1	30	1	55.5	1	1	0	0
9	53	1	196	1	69	1	0	1	0
10	15	1	154	1	51.5	0	1	0	0
11	7	1	333	1	44	1	0	1	0
12	141	1	8	0	34	1	0	0	0
13	96	1	38	1	35	1	0	1	0
14	149	0	70	0	42	1	0	1	0
15	536	1	25	0	17	1	0	0	0
16	17	1	4	0	60	0	0	1	0
17	185	1	177	1	60	1	0	0	0
18	292	1	114	1	43.5	1	0	0	0
19	22	0	159	0	53	1	1	0	0
20	15	1	108	0	44	1	0	0	0
21	152	1	562	1	46.5	0	0	0	1
22	402	1	24	0	30	1	0	0	0
23	13	1	66	1	62.5	1	0	1	0
24	39	1	46	0	42.5	1	0	1	0
25	12	1	40	1	43	0	0	1	0
26	113	0	201	1	57.5	1	0	1	0
27	132	1	156	1	10	1	1	0	0
28	34	1	30	1	52	1	0	1	0
29	2	1	25	1	53	0	1	0	0
30	130	1	26	1	54	1	1	0	0
31	27	1	58	1	56	1	0	1	0
32	5	0	43	1	50.5	1	0	1	0
33	152	1	30	1	57	1	0	0	1
34	190	1	5	0	44.5	1	1	0	0
35	119	1	8	1	22	1	0	0	0
36	54	0	16	0	42	1	0	0	0
37	6	0	78	1	52	1	0	0	1
38	63	1	8	0	60	0	0	0	1

TABLE 1.4: Kidney infection data

L	I	T	S	L	I	T	S	L	I	T	S	L	I	T	S
1	1	101	0	13	0	50	1	26	0	104	0	39	1	76	0
1	0	49	1	14	1	103	1	26	0	104	0	39	0	84	1
1	0	104	0	14	0	69	0	27	1	78	0	39	0	78	1
2	1	104	0	14	0	91	0	27	0	104	0	40	1	80	1
2	0	102	0	15	1	93	0	27	0	104	0	40	0	81	1
2	0	104	0	15	0	104	0	28	1	104	0	40	0	76	0
3	1	104	0	15	0	103	0	28	0	81	1	41	1	72	1
3	0	104	0	16	1	85	0	28	0	64	1	41	0	95	0
3	0	104	0	16	0	72	0	29	1	86	1	41	0	104	0
4	1	77	0	16	0	104	0	29	0	55	1	42	1	73	1
4	0	97	0	17	1	104	0	29	0	94	0	42	0	104	0
4	0	79	0	17	0	63	0	30	1	34	1	42	0	66	1
5	1	89	0	17	0	104	0	30	0	104	0	43	1	92	1
5	0	104	0	18	1	104	0	30	0	54	1	43	0	104	0
5	0	104	0	18	0	104	0	31	1	76	0	43	0	102	1
6	1	88	1	18	0	74	0	31	0	87	0	44	1	104	0
6	0	96	1	19	1	81	0	31	0	74	0	44	0	98	0
6	0	104	0	19	0	104	0	32	1	103	1	44	0	73	0
7	1	104	1	19	0	69	0	32	0	73	1	45	1	55	0
7	0	94	0	20	1	67	1	32	0	84	1	45	0	104	0
7	0	77	1	20	0	104	0	33	1	102	1	45	0	104	0
8	1	96	1	20	0	68	1	33	0	104	0	46	1	49	0
8	0	104	0	21	1	104	0	33	0	80	0	46	0	83	0
8	0	104	0	21	0	104	0	34	1	80	1	46	0	77	0
9	1	82	0	21	0	104	0	34	0	104	0	47	1	89	1
9	0	77	0	22	1	104	0	34	0	73	0	47	0	104	0
9	0	104	0	22	0	104	0	35	1	45	1	47	0	104	0
10	1	70	1	22	0	104	0	35	0	79	0	48	1	88	0
10	0	104	0	23	1	104	0	35	0	104	0	48	0	79	0
10	0	77	0	23	0	83	0	36	1	94	1	48	0	99	0
11	1	89	1	23	0	40	1	36	0	104	0	49	1	103	1
11	0	91	0	24	1	87	0	36	0	104	0	49	0	91	0
11	0	90	0	24	0	104	0	37	1	104	0	49	0	104	0
12	1	91	0	24	0	104	0	37	0	104	0	50	1	104	0
12	0	70	0	25	1	104	0	37	0	104	0	50	0	104	0
12	0	92	0	25	0	104	0	38	1	104	0	50	0	79	1
13	1	39	1	25	0	104	0	38	0	101	1				
13	0	45	0	26	1	89	0	38	0	94	0				

TABLE 1.5: Litters of rats data

## 1.8 Diabetic Retinopathy Data

Huster et al. (1989) reported the diabetic retinopathy data. The 197 patients in this data set were a 50% random sample of the patients with ‘high-

Pat	Time	Cen	HAL	Pat	Time	Cen	HAL
1	37	1	1	9	77	1	1
1	29	1	0	9	63	1	1
2	19	1	1	9	43	1	0
2	13	1	0	9	29	1	1
3	57	0	1	10	29	1	1
3	57	0	1	10	18	1	0
3	15	1	0	10	15	1	0
4	93	1	1	11	60	0	1
4	26	1	0	11	38	1	0
5	16	1	1	12	19	1	0
5	11	1	0	13	24	1	1
6	21	1	1	14	18	1	0
6	15	1	0	14	18	1	0
7	26	1	0	15	19	1	0
7	20	1	1	15	19	1	0
8	19	1	0	16	28	0	0
8	18	1	1	16	28	0	0

TABLE 1.6: Kidney dialysis patients data

risk' diabetic retinopathy as defined by the diabetic retinopathy Study (DRS). Each patient had one eye randomized to laser treatment and the other eye received no treatment. For each eye, the event of interest was the time from initiation of treatment to the time when visual acuity dropped below 5/200 two visits in a row (call it 'blindness'). Thus there is a built-in lag time of approximately 6 months (visits were every 3 months). Survival times in this data set are therefore the actual time to blindness in months, minus the minimum possible time to event (6.5 months). Censoring was caused by death, dropout, or end of the study. The full data is in Table A.2 is given in Appendix and the partial data is given in [Table 1.7](#) with six patients.

```

1: Subject id(SI)
2: laser type(LT): 1=xenon, 2=argon
3: treated eye(TE): 1=right 2=left
4: age at diagnosis of diabetes(Age):
5: type of diabetes(TD): 1= juvenile (age at dx $<$ 20), 2=adult
6: Outcome for the treated eye(OT): risk group: 6-12
7: status(ST): 0=censored, 1=blindness
8: follow-up time for treated eye(TT):
9: Outcome for the untreated eye(OU): risk group: 6-12
10: status(SU): 0=censored, 1=blindness
11: follow-up time for untreated eye(TU):
The risk group variable was used to define the 'high risk' samples.

```

SI	LT	TE	Age	TD	OT	ST	TT	OU	SU	TU
5	2	2	28	2	9	0	46.23	9	0	46.23
14	2	1	12	1	8	0	42.5	6	1	31.3
16	1	1	9	1	11	0	42.27	11	0	42.27
25	2	2	9	1	11	0	20.6	11	0	20.6
29	1	2	13	1	9	0	38.77	10	1	0.3
46	1	1	12	1	9	0	65.23	9	1	54.27

TABLE 1.7: Diabetic retinopathy data

## 1.9 Myeloma Data

Krall, Uthoff, and Harley (1975) analyzed data from a study on multiple myeloma in which researchers treated 65 patients with alkylating agents. Of those patients, 48 died during the study and 17 survived. In the data set Myeloma, the variable Time represents the survival time in months from diagnosis. The variable Status consists of two values, 0 and 1, indicating whether the patient was alive or dead, respectively, at the end of the study. If the value of Status is 0, the corresponding value of Time is censored. The variables thought to be related to survival are levels of LogBUN (log(Blood Urea Nitrogen) at diagnosis), HGB (hemoglobin at diagnosis), Platelet (platelets at diagnosis: 0=abnormal, 1=normal), Age (age at diagnosis in years), LogWBC (log(WBC) at diagnosis), Frac (fractures at diagnosis: 0=none, 1=present), LogPBM (log percentage of plasma cells in bone marrow), Protein (proteinuria at diagnosis), and SCalc (serum calcium at diagnosis). Interest lies in identifying important prognostic factors from these nine explanatory variables. The complete data is presented in Table A.3 in Appendix and partial data is presented in [Table 1.8](#) with six patients.

```

Time(T)
Status(S)
LogBUN(LBUN)
HGB
Platelet(P)
Age
LogWBC(LW)
Frac(F)
LogPBM(LPBM)
Protein(Pr)
SCalc(SC)

```

T	S	LBUN	HGB	P	Age	LW	F	LPBM	Pr	SC
1.25	1	2.2175	9.4	1	67	3.6628	1	1.9542	12	10
1.25	1	1.9395	12.0	1	38	3.9868	1	1.9542	20	18
2.00	1	1.5185	9.8	1	81	3.8751	1	2.0000	2	15
2.00	1	1.7482	11.3	0	75	3.8062	1	1.2553	0	12
2.00	1	1.3010	5.1	0	57	3.7243	1	2.0000	3	9
3.00	1	1.5441	6.7	1	46	4.4757	0	1.9345	12	10

TABLE 1.8: Myeloma data

## 1.10 Definitions and Notations

### 1.10.1 Survival Function

The basic quantity employed to describe time-to-event phenomenon is the **Survival Function**  $S(t)$ , and it is defined as:

$$S(t) = P[T > t] = \text{the probability an individual survives beyond time } t.$$

Since a unit either fails, or survives, and one of these two mutually exclusive alternatives must occur, we have

$$S(t) = 1 - F(t), \quad F(t) = 1 - S(t),$$

where  $F(t)$  is the cumulative distribution function(CDF). If  $T$  is a continuous random variable, then  $S(t)$  is a continuous, strictly decreasing function. The survival function is the integral of the probability density function (pdf),  $f(t)$ , that is

$$S(t) = \int_t^{\infty} f(x)dx \tag{1.1}$$

Thus,

$$f(t) = -\frac{dS(t)}{dt}. \tag{1.2}$$

### 1.10.2 Failure (or Hazard) Rate

The failure rate is defined as the instantaneous rate of failure (experiencing the event) for the survivors to time  $t$  during the next instant of time. It is a rate per individual of time. The next instant the failure rate may change and the individuals that have already failed play no further role since only the survivors count.

The failure rate (or hazard rate) is denoted by  $h(t)$  and is defined by the following equation

$$\begin{aligned} h(t) &= \lim_{h \rightarrow 0} \frac{P[t \leq T \leq t+h | T \geq t]}{h} \\ &= \frac{f(t)}{S(t)} = \text{the instantaneous (conditional) failure rate.} \end{aligned} \quad (1.3)$$

The failure rate is sometimes called a “conditional failure rate” since the denominator  $S(t)$  (i.e., the population survivors) converts the expression into a conditional rate, given survival past time  $t$ .

Since  $h(t)$  is also equal to the negative of the derivative of  $\ln\{S(t)\}$ , we have the useful identity:

$$S(t) = \exp \left\{ - \int_0^t h(t) dt \right\}. \quad (1.4)$$

If we let

$$H(t) = \int_0^t h(t) dt \quad (1.5)$$

be the **Cumulative Hazard Function**, we then have  $S(t) = e^{-H(t)}$ . Two other useful identities that follow from these formulas are:

$$h(t) = -\frac{d \ln S(t)}{dt} \quad (1.6)$$

$$H(t) = -\ln S(t). \quad (1.7)$$

### Bathtub curve:

People have calculated empirical population failure rates as units age over time and repeatedly obtained a graph such as shown in [Figure 1.1](#). Because of the shape of this failure rate curve, it has become widely known as the “Bathtub” curve.

The initial region that begins at time zero when a baby is born, the mortality rate is high and gradually decreases after one year. This region is known as the **Early Failure Period** (also referred to as **Infant Mortality Period**, from the actuarial origins of the first bathtub curve plots).

Next, the failure rate levels off and remains roughly constant for (hopefully) the majority of the useful life of an individual. This long period of a level failure rate is known as the **Intrinsic Failure Period** (also called the **Stable Failure Period**) and the constant failure rate level is called the **Intrinsic Failure Rate**.

Finally, an individual reaches his/her retirement age, the failure rate begins to increase because of old age. This is the **Wearout Failure Period**.

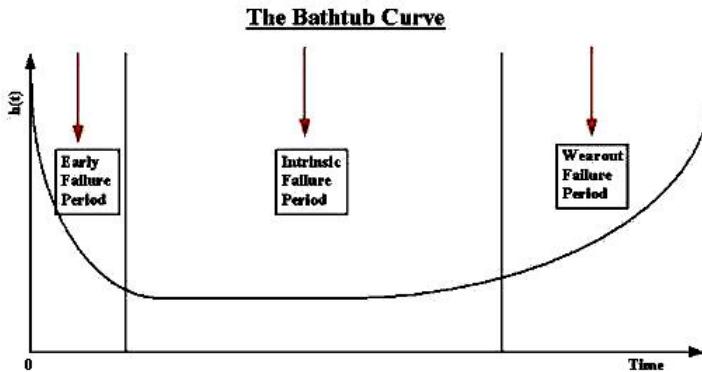


FIGURE 1.1: Bathtub failure rate curve.

It is also sometimes useful to define an average failure rate over any interval  $(T_1, T_2)$  that ‘averages’ the failure rate over that interval. This rate, denoted by average failure rate,  $AFR(T_1, T_2)$ , is a single number that can be used as a specification or target for the population failure rate over that interval.

The formulas for calculating AFR’s are:

$$\begin{aligned} AFR(t_1, t_2) &= \frac{\left(\int_{t_1}^{t_2} h(t)dt\right)}{t_2 - t_1} = \frac{H(t_2) - H(t_1)}{t_2 - t_1} = \frac{\ln S(t_1) - \ln S(t_2)}{t_2 - t_1} \\ AFR(0, t) &= AFR(t) = \frac{H(t)}{t} = \frac{-\ln S(t)}{t}. \end{aligned}$$

## Bivariate and Multivariate Survival Function

The bivariate survival function of the lifetimes  $(T_1, T_2)$  is given by

$$S(t_1, t_2) = P[T_1 > t_1, T_2 > t_2] = \exp[-H(t_1, t_2)], \quad (1.8)$$

where  $H(t_1, t_2)$  is the bivariate integrated hazard function of  $(T_1, T_2)$  which can be written in term of bivariate survival function as

$$H(t_1, t_2) = -\ln S(t_1, t_2). \quad (1.9)$$

The multivariate survival function of the lifetimes  $(T_1, \dots, T_k)$

$$S(t_1, \dots, t_k) = P[T_1 > t_1, \dots, T_k > t_k] = \exp[-H(t_1, \dots, t_k)], \quad (1.10)$$

where  $H(t_1, \dots, t_k)$  is the multivariate integrated hazard function of  $(T_1, \dots, T_k)$  which can be written in term of multivariate survival function as

$$H(t_1, \dots, t_k) = -\ln S(t_1, \dots, t_k). \quad (1.11)$$

The multivariate hazard rate of  $(T_1, \dots, T_k)$  is defined by

$$\begin{aligned} h(t_1, \dots, t_k) &= \frac{f(t_1, \dots, t_k)}{S(t_1, \dots, t_k)} \\ &= \frac{(-1)^k \frac{\partial^k S(t_1, \dots, t_k)}{\partial t_1, \dots, \partial t_k}}{S(t_1, \dots, t_k)}. \end{aligned} \quad (1.12)$$


---

## 1.11 Censoring

One peculiar feature, often present in time to event data, is known as censoring, which, broadly speaking, occurs when some lifetimes are known to occur at some point of time or within certain intervals. The meaning of censoring is already discussed in the introduction of this chapter. For example,  $n$  patients under some specific disease (cancer or AIDS) are chosen for the treatment, some of them die and the remaining will be survived from the disease at the end of study period. The patients which are survived are censored at that particular time or end of study period.

When fitting models and estimating failure rates from survival data, the precision of the estimates (as measured by the width of the confidence intervals) tends to vary inversely with the square root of the number of failures observed - not the number of individuals on test or the length of the test. In other words, a test where 5 fail out of a total of 10 on a test gives more information than a test with 1000 units but only 2 failures.

Since the number of failures  $r$  is critical, and not the sample size  $n$  on a test, it becomes increasingly difficult to assess the failure rates of highly censored data.

### 1.11.1 Censored Type I Data

Consider a situation in which we have  $n$  cancer patients taken from a population. In the typical scenario, we have a fixed time  $T$  to observe patients whether they survive or fail. The data obtained are called **censored type I** data.

Suppose we observe  $r$  deaths (where  $r$  can be any number from 0 to  $n$ ) out of  $n$  patients admitted in a hospital under some specific disease. Let  $T$  (in days) be the time when the treatment is terminated. The (exact) death times are  $t_1, t_2, \dots, t_r$  and there are  $(n - r)$  individuals that survived the entire  $T$ -days without failing. Note that  $T$  is fixed in advance and  $r$  is random, since we don't know how many deaths will occur until  $T$  days. Note also that we assume the exact times of failure are recorded when there are failures.

This type of censoring is also called “right censored” data since the times of failure to the right (i.e., larger than  $T$ ) are missing.

### 1.11.2 Censored Type II Data

Another (much less common) way to test is to decide in advance that you want to see exactly  $r$  failure times and then test until they occur. For example, you might put 100 patients on test and decide you want to see at least half of them fail. Then  $r = 50$ , but  $T$  is unknown until the 50th fail occurs. This is called **censored type II** data.

We observe  $t_1, t_2, \dots, t_r$ , where  $r$  is specified in advance. The test ends at time  $T = t_r$ , and  $(n - r)$  units have survived. Again we assume it is possible to observe the exact time of death for dead individuals.

Type II censoring has the significant advantage that you know in advance how many failure times your test will yield - this helps enormously when planning adequate tests. However, an open-ended random test time is generally impractical from a management point of view and this type of testing is rarely seen.

### 1.11.3 Readout or Interval Censored Data

Sometimes exact times of failure are not known; only an interval of time in which the failure occurred is recorded. This kind of data is called **Readout or interval censored** data and the situation is shown in the Figure 1.2. Let  $T_i - T_{i-1}$ ,  $i = 1, 2, \dots, k$ ,  $T_0 = 0$  are the  $k$  time intervals and  $r_i$  be the number of deaths in the time interval  $T_i - T_{i-1}$  and  $n - \sum r_i$  patients are survived (censored) at time  $T_k$  out of  $n$  patients exposed to some particular disease. Here we do not know the exact time of deaths of patients but we know the patient died during the time interval. Figure 1.2 clearly shows the during each time interval how many patients died.

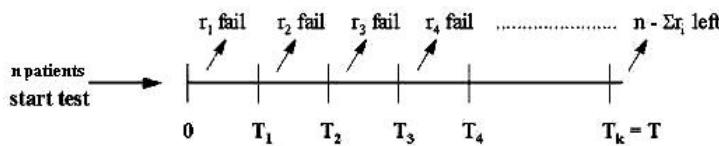


FIGURE 1.2: Readout or interval data diagram.

### 1.11.4 Multicensored Data

In the most general case, every individual observed yields exactly one of the following three types of information:

- a run-time if an individual did not fail while under observation

- an exact failure time
- an interval of time during which an individual failed.

The individual may all have different run-times and/or readout intervals. Many statistical methods can be used to fit models and estimate failure rates, even with censored data. In later chapters we will discuss the Kaplan-Meier approach, Probability Plotting, Hazard Plotting, Graphical Estimation, and Maximum Likelihood Estimation.

### **1.11.5 Separating out Failure Modes**

Note that when a data set consists of failure times that can be sorted into several different failure modes, it is possible (and often necessary) to analyze and model each mode separately. Consider all failures due to modes other than the one being analyzed as censoring times, with the censored run-time equal to the time it failed due to the different (independent) failure mode.

# **Chapter 2**

---

## ***Some Parametric Methods***

2.1	Introduction .....	19
2.2	Exponential Distribution .....	20
2.3	Weibull Distribution .....	21
2.4	Extreme Value Distributions .....	22
2.5	Lognormal .....	25
2.6	Gamma .....	26
2.7	Loglogistic .....	28
2.8	Maximum Likelihood Estimation .....	30
2.9	Parametric Regression Models .....	35
2.9.1	Example 2.1 .....	39
2.9.2	Example 2.2 .....	41

---

### **2.1 Introduction**

There are a handful of parametric models that have successfully served as population models for failure time individuals. Sometimes there are probabilistic arguments based on the physics of the failure mode that tend to justify the choice of model. Some times the model is used solely because of its empirical success in fitting actual failure data.

Six parametric models will be described in this section:

1. Exponential
2. Weibull
3. Extreme Value
4. Lognormal
5. Gamma
6. Loglogistic

## 2.2 Exponential Distribution

The exponential model, with only one unknown parameter, is the simplest of all life distribution models. The key equations for the exponential distribution are shown below:

$$\text{CDF: } F(t) = 1 - e^{-\lambda t}, \quad 0 < t < \infty$$

$$\text{SURVIVAL: } S(t) = e^{-\lambda t}$$

$$\text{PDF: } f(t) = \lambda e^{-\lambda t}$$

$$\text{MEAN: } \frac{1}{\lambda}$$

$$\text{MEDIAN: } \frac{\ln 2}{\lambda} \simeq \frac{.693}{\lambda}$$

$$\text{VARIANCE: } \frac{1}{\lambda^2}$$

$$\text{HAZARD RATE: } h(t) = \lambda.$$

Note that the failure rate reduces to the constant  $\lambda$  for any time. The exponential distribution is the only distribution to have a constant failure rate. Also, another name for the exponential mean is the **mean time to fail** or **MTTF** and we have  $\text{MTTF} = 1/\lambda$ . It has one most important property, that is, loss of memory property (LMP). Exponential is the only continuous distribution which has LMP.

The cumulative hazard function for the exponential is just the integral of the failure rate or  $H(t) = \lambda t$ .

Figure 2.1 presents the probability density function and survival function of the exponential distribution with parameter  $\lambda$  taking values 1 and 0.5.

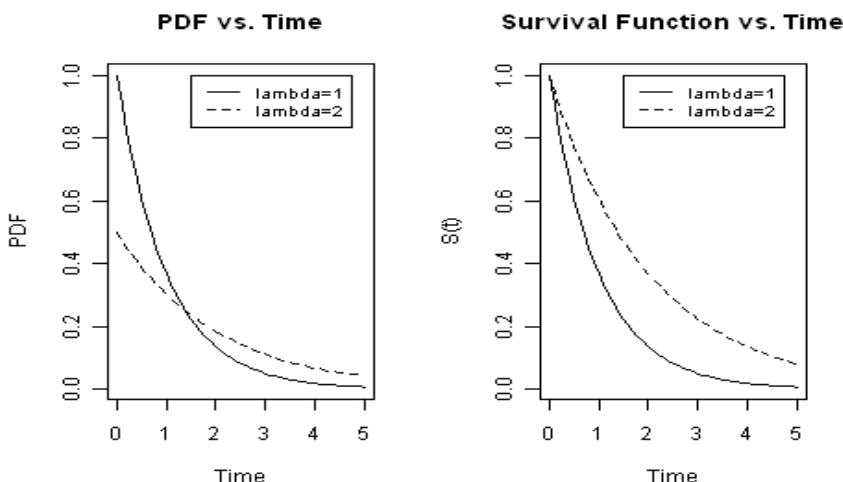


FIGURE 2.1: PDF and survival function graphs of exponential distribution.

## 2.3 Weibull Distribution

The Weibull is a very flexible life distribution model with two parameters. The CDF, survival, PDF and other key formulas of Weibull distribution are, respectively given by,

$$\text{CDF} : F(t) = 1 - e^{-(\frac{t}{\lambda})^\gamma}, \quad 0 < t < \infty$$

$$\text{SURVIVAL} : S(t) = e^{-(\frac{t}{\lambda})^\gamma}$$

$$\text{PDF} : f(t) = \frac{\gamma}{\lambda} \left(\frac{t}{\lambda}\right)^\gamma e^{-(\frac{t}{\lambda})^\gamma}$$

$$\text{HAZARD RATE} : \frac{\gamma}{\lambda} \left(\frac{t}{\lambda}\right)^{\gamma-1}$$

$$\text{MEAN} : \lambda \Gamma\left(1 + \frac{1}{\gamma}\right)$$

$$\text{MEDIAN} : \lambda (\ln 2)^{1/\gamma}$$

$$\text{VARIANCE} : \lambda^2 \Gamma\left(1 + \frac{2}{\gamma}\right) - \left[\lambda \Gamma\left(1 + \frac{1}{\gamma}\right)\right]^2$$

with  $\lambda$  the scale parameter (the **Characteristic Life**),  $\gamma$  the **Shape Parameter**, and  $\Gamma$  is the Gamma function with  $\Gamma(N) = (N - 1)!$  for integer  $N$ .

The cumulative hazard function for the Weibull is the integral of the failure rate or

$$H(t) = \left(\frac{t}{\lambda}\right)^\gamma.$$

A more general 3-parameter form of the Weibull includes an additional **waiting time** parameter  $\mu$  (sometimes called a **shift** or **location** parameter). The formulas for the 3-parameter Weibull are easily obtained from the above formulas by replacing  $t$  by  $(t - \mu)$  wherever  $t$  appears. No failure can occur before  $\mu$  hours, so the time scale starts at  $\mu$ , and not 0. If a shift parameter  $\mu$  is known (based, perhaps, on the physics of the failure mode), then all you have to do is subtract  $\mu$  from all the observed failure times and/or readout times and analyze the resulting shifted data with a 2-parameter Weibull.

**Special Case :** When  $\gamma = 1$ , the Weibull reduces to the Exponential Model, with  $1/\lambda =$  the **mean time to fail (MTTF)**.

Depending on the value of the shape parameter  $\gamma$ , the Weibull model can empirically fit a wide range of data histogram shapes. [Figure 2.2](#) gives the the Weibull density with scale parameter  $\lambda = 1$  and several different values of the shape parameter  $\gamma$ .

From a failure rate model viewpoint, the Weibull is a natural extension of the constant failure rate exponential model since the Weibull has a polynomial failure rate with exponent  $\{\gamma - 1\}$ . This makes all the failure rate curves shown in the following [Figure 2.3](#) when  $\lambda = 1$ . When  $\gamma = 1$ , the failure rate remains constant as time increases; this is the exponential case. The failure rate increases when  $\gamma > 1$  and decreases when  $\gamma < 1$  as time increases. Thus,

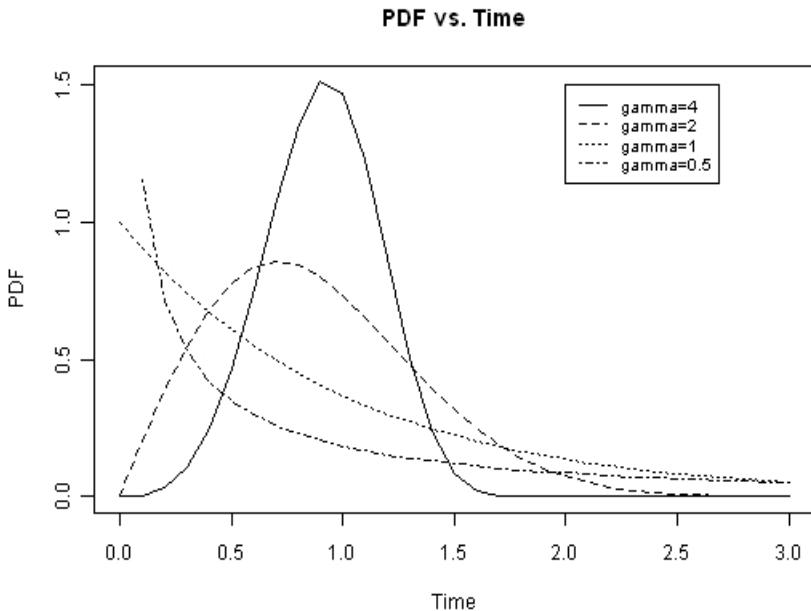


FIGURE 2.2: Graphs of PDF of Weibull distribution with  $\lambda = 1$ .

the Weibull distribution is used to model the survival distribution of a population with increasing, decreasing, or constant risk. Examples of increasing and decreasing failure rates are respectively, patients with lung cancer and patients who undergo successful major surgery.

For the survival curve, it is simple to plot the logarithm of  $S(t)$ ,

$$\ln S(t) = -\left(\frac{t}{\lambda}\right)^\gamma.$$

[Figure 2.4](#) gives  $\ln S(t)$  for  $\lambda = 1$  and  $\gamma = 1, > 1, < 1$ . When  $\gamma = 1$  is a straight line with negative slope. When  $\gamma < 1$ , negative aging,  $\ln S(t)$  decreases very slowly from 0 and then approaches a constant value. When  $\gamma > 1$ , positive aging,  $\ln S(t)$  decreases sharply from 0 as time increases. The above equation can also be written as

$$\ln(-\ln S(t)) = -\gamma \ln \lambda + \gamma \ln t.$$

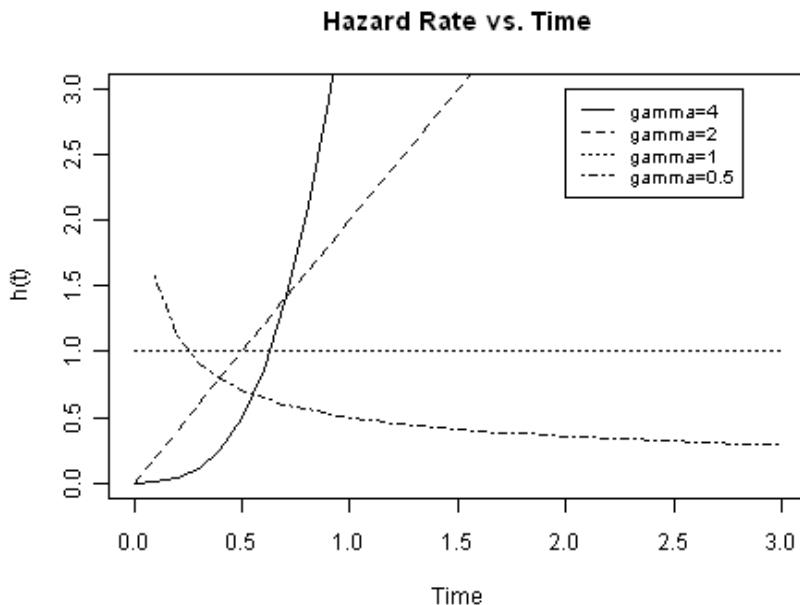


FIGURE 2.3: Failure rate curves of Weibull distribution with  $\lambda = 1$ .

## 2.4 Extreme Value Distributions

The Weibull distribution and the extreme value distribution have a useful mathematical relationship. If  $t_1, t_2, \dots, t_n$  are a sample of random times of fail from a Weibull distribution, then  $\ln t_1, \ln t_2, \dots, \ln t_n$  are random observations from the extreme value distribution. In other words, the natural log of a Weibull random time is an extreme value random observation.

If the Weibull has shape parameter  $\gamma$  and characteristic life  $\lambda$ , then the extreme value distribution (after taking natural logarithms) has  $\mu = \ln \lambda$ ,  $\beta = 1/\gamma$ .

Because of this relationship, computer programs designed for the extreme value distribution can be used to analyze Weibull data. The situation exactly parallels using normal distribution programs to analyze lognormal data, after first taking natural logarithms of the data points.

Extreme value distributions are the limiting distributions for the minimum or the maximum of a very large collection of random observations from the same arbitrary distribution. For any well-behaved initial distribution (i.e.,  $F(t)$ )

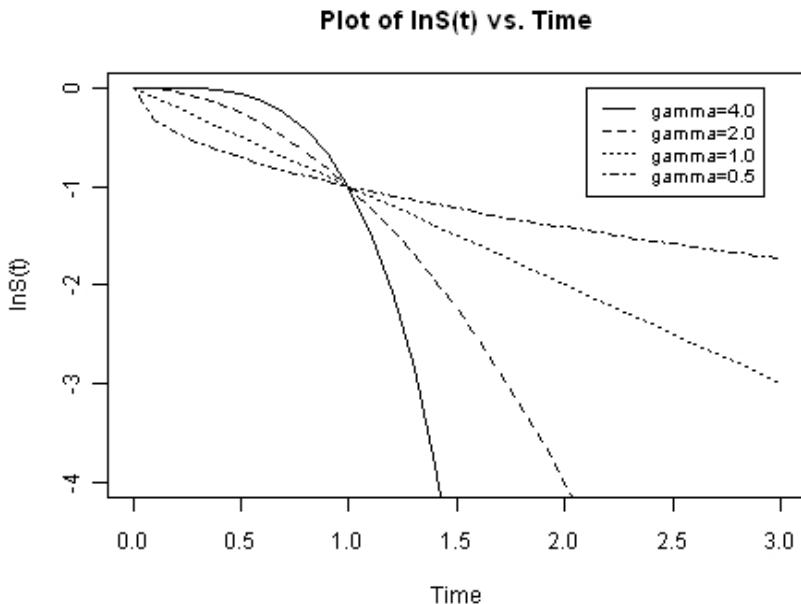


FIGURE 2.4: Weibull distribution:  $\ln S(t)$  vs. time with  $\lambda = 1$ .

is continuous and has an inverse), only a few models are needed, depending on whether you are interested in the maximum or the minimum, and also if the observations are bounded above or below.

The distribution often referred to as the Extreme Value Distribution (Type I) is the limiting distribution of the minimum of a large number of unbounded identically distributed random variables. The PDF and CDF are given by:

$$\begin{aligned} f(t) &= \frac{1}{\beta} e^{\frac{t-\mu}{\beta}} e^{-e^{\frac{t-\mu}{\beta}}}, \quad -\infty < t < \infty, \beta > 0 \\ F(t) &= 1 - e^{-e^{\frac{t-\mu}{\beta}}}, \quad -\infty < t < \infty, \beta > 0. \end{aligned}$$

In any modeling application for which the variable of interest is the minimum of many random factors, all of which can take positive or negative values, try the extreme value distribution as a likely candidate model. For lifetime distribution modeling, since failure times are bounded below by zero, the Weibull distribution is a better choice.

## 2.5 Lognormal

The lognormal life distribution, like the Weibull, is a very flexible model that can empirically fit many types of failure data. The two parameter form has parameters  $\sigma$  = the **shape** parameter and  $T_{50}$  = the **median** (a **scale** parameter).

**NOTE :** If time to failure,  $t_f$ , has a lognormal distribution, then the (natural) logarithm of time to failure has a normal distribution with mean  $\mu = \ln T_{50}$  and standard deviation  $\sigma$ . This makes lognormal data convenient to work with; just take natural logarithms of all the failure times and censoring times and analyze the resulting normal data. Later on, convert back to real time and lognormal parameters using  $\sigma$  as the lognormal shape and  $T_{50} = e^\mu$  as the (median) scale parameter. The density, cumulative density, survival, hazard, mean, and variance of lognormal distribution are, respectively,

$$\text{PDF: } f(t) = \frac{1}{\sigma t \sqrt{2\pi}} e^{-(\frac{1}{2\sigma^2})(\ln t - \ln T_{50})^2}, \quad 0 < t < \infty$$

$$\text{CDF : } F(t) = \int_0^t \frac{1}{\sigma x \sqrt{2\pi}} e^{-(\frac{1}{2\sigma^2})(\ln x - \ln T_{50})^2} dx = \Phi\left(\frac{\ln t - \ln T_{50}}{\sigma}\right)$$

with  $\Phi(z)$  denoting the standard Normal CDF

$$\text{SURVIVAL : } S(T) = 1 - F(t)$$

$$\text{FAILURE RATE : } h(t) = \frac{f(t)}{S(t)} = \frac{(1/t\sigma\sqrt{2\pi}) \exp[-(\ln t - \ln T_{50})^2/2\sigma^2]}{1 - G[(\ln t - \ln T_{50})/\sigma]}$$

$$\text{MEAN : } T_{50}e^{\sigma^2/2}$$

$$\text{MEDIAN : } T_{50}$$

$$\text{VARIANCE : } T_{50}^2 e^{\sigma^2} (e^{\sigma^2} - 1).$$

**NOTE :** A more general 3-parameter form of the lognormal includes an additional **waiting** time parameter  $\theta$  (sometimes called a **shift** or **location** parameter). The formulas for the 3-parameter lognormal are easily obtained from the above formulas by replacing  $t$  by  $(t - \theta)$  wherever  $t$  appears. No failure can occur before  $\theta$  hours, so the time scale starts at  $\theta$  and not 0. If a shift parameter  $\theta$  is known, then all you have to do is subtract  $\theta$  from all the observed failure times and/or readout times and analyze the resulting shifted data with a 2-parameter lognormal.

Examples of lognormal density curves are shown in Figures 2.5 and 2.6. Figure 2.5 gives the lognormal frequency curves for  $\mu = 0, \sigma = 0.5, 1, 2$ , from which the idea of the flexibility of the distribution may be obtained. It is obvious that the distribution is positively skewed and that the greater the value of  $\sigma^2$ , the greater the skewness. Figure 2.6 shows the frequency curves for  $\sigma = 1, \mu = 0, 0.5, 1.5$ . Note that lognormal shapes for small sigmas are very similar to Weibull shapes when the shape parameter  $\gamma$  is large and large sigmas give plots similar to small Weibull  $\gamma$ 's. Both distributions are very flexible and it is often difficult to choose which to use based on empirical fits to small samples of (possibly censored) data.

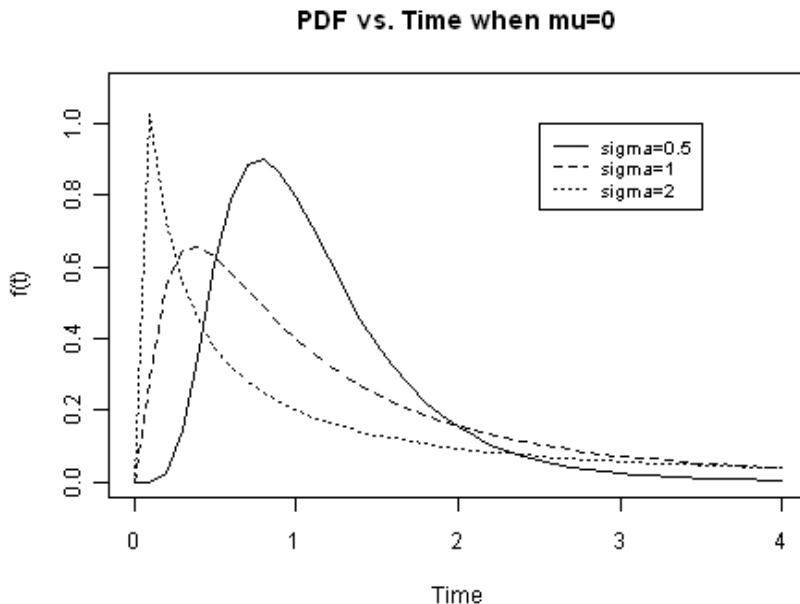


FIGURE 2.5: Lognormal density curves with  $\mu = 0$ .

The failure (hazard) rate functions of lognormal distribution for different combinations of  $\mu$  and  $\sigma$  are plotted in Figure 2.7. The three hazard curves in Figure 2.7 are corresponding to the parameters  $\mu = 0.4, \sigma = 0.4$ ,  $\mu = 0.6, \sigma = 0.6$ , and  $\mu = 1, \sigma = 1$  respectively.

As shown in the preceding plots, the lognormal PDF and failure rate shapes are flexible enough to make the lognormal a very useful empirical model. In addition, the relationship to the normal (just take natural logarithms of all the data and time points and you have “normal” data) makes it easy to work with mathematically, with many good software analysis programs available to treat normal data.

## 2.6 Gamma

There are different ways of writing (parameterizing) the gamma distribution that are common in the literature. In addition, different authors use different symbols for the shape and scale parameters. Below we define the gamma, with  $\gamma$ , the ‘shape’ parameter, and  $\lambda$ , the ‘scale’ parameter. This

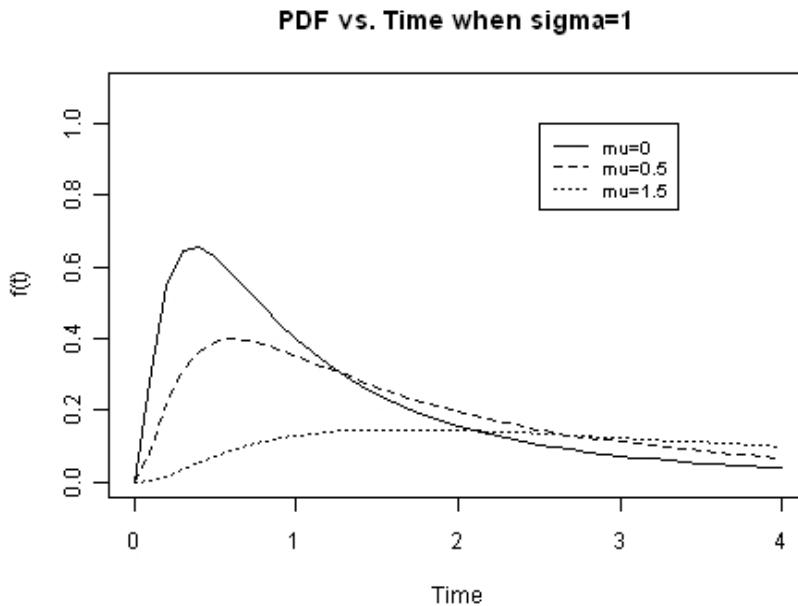


FIGURE 2.6: PDF curves of the lognormal distribution with  $\sigma = 1$ .

choice of parameters  $(\gamma, \lambda)$  will be the most convenient for later applications of the gamma. The density, cumulative density, survival, hazard, mean, and variance of gamma distribution are, respectively,

$$\text{PDF: } f(t, \gamma, \lambda) = \frac{\lambda^\gamma}{\Gamma(\gamma)} t^{\gamma-1} e^{-\lambda t}, \quad 0 < t < \infty$$

$$\text{CDF: } F(t) = \int_0^t f(t) dt$$

$$\text{SURVIVAL: } S(t) = 1 - F(t)$$

$$\text{HAZARD RATE : } h(t) = \frac{f(t)}{S(t)}$$

$$\text{MEAN: } \frac{\gamma}{\lambda}$$

$$\text{VARIANCE: } \frac{\gamma}{\lambda^2}.$$

**NOTE :** When  $\gamma = 1$ , the gamma reduces to an exponential distribution with  $\lambda$ .

Another well-known statistical distribution, the Chi-Square, is also a special case of the gamma. A Chi-Square distribution with  $n$  degrees of freedom is the same as a gamma with  $\gamma = n/2$  and  $\lambda = .5$ .

Figures 2.8 and 2.9 give examples of gamma density functions with various values of  $\gamma$  and  $\lambda$ , respectively. It is seen that varying  $\gamma$  changes the shape of the distribution while varying  $\lambda$  changes only the scaling. Consequently,  $\gamma$  and  $\lambda$  are shape and scale parameters, respectively. When  $\gamma > 1$ , there is a single peak at  $t = (\gamma - 1)/\lambda$ .

When  $0 < \gamma < 1$ , there is negative aging and the failure rate decreases

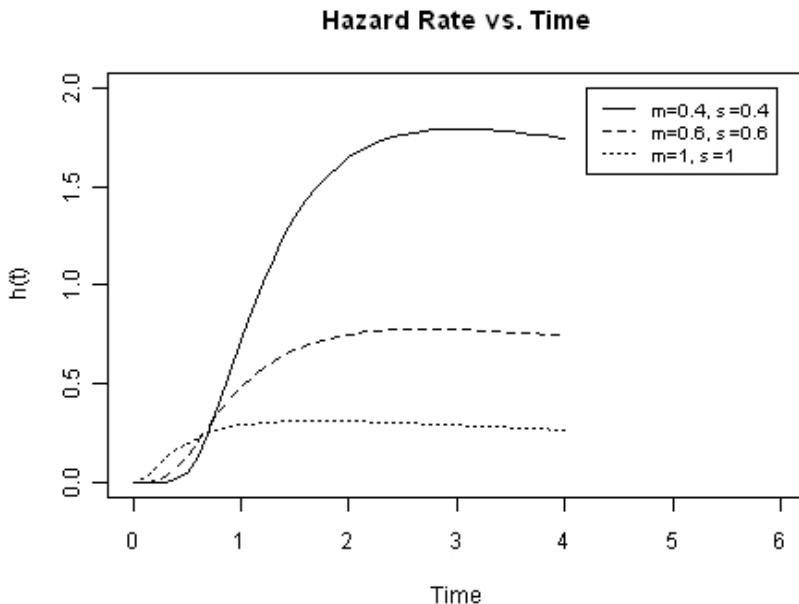


FIGURE 2.7: Hazard rate of the lognormal with different parameters.

monotonically from infinity to  $\lambda$  as time increases from 0 to infinity. When  $\gamma > 1$ , there is positive aging and the failure rate increases monotonically from 0 to  $\lambda$  as time increases from 0 to infinity. When  $\gamma = 1$ , the failure rate equals  $\lambda$ , a constant, as in the exponential case. Figure 2.10 shows the gamma failure rate function for  $\lambda = 1$  and  $\gamma = 0.5, 1, 2, 4$ . Thus, the gamma distribution describes a different type of survival pattern where the failure rate is decreasing or increasing to a constant value as time approaches to infinity.

The gamma is a flexible life distribution model that may offer a good fit to some sets of failure data. It is not, however, widely used as a life distribution model for common failure mechanisms.

**Note :** when  $\gamma$  is a positive integer, the gamma is sometimes called an **Erlang distribution**. The Erlang distribution is used frequently in queuing theory applications.

A common use of the gamma model occurs in Bayesian reliability applications. When a system follows an HPP (exponential) model with a constant repair rate  $\lambda$ , and it is desired to make use of prior information about possible values of  $\lambda$ , a gamma Bayesian prior for  $\lambda$  is a convenient and popular choice.

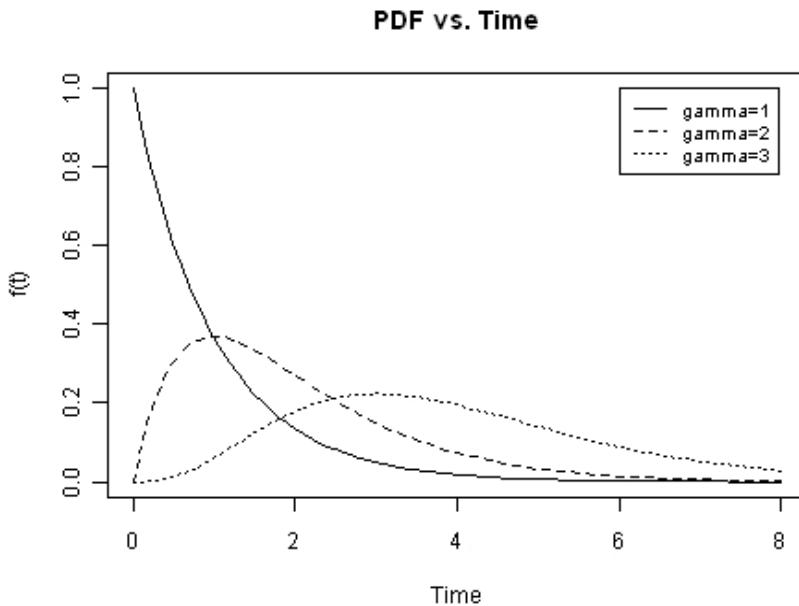


FIGURE 2.8: Gamma density functions with  $\lambda = 1$ .

## 2.7 Loglogistic

The survival time  $T$  has a loglogistic distribution if  $\ln(T)$  has a logistic distribution. The density, survival, hazard, and cumulative hazard functions of the loglogistic distribution are, respectively,

$$\text{PDF: } f(t, a, b) = \frac{\alpha\gamma t^{\gamma-1}}{(1+\alpha t^\gamma)^2}, \quad 0 < t < \infty$$

$$\text{CDF: } F(t) = \frac{\alpha t^\gamma}{1+\alpha t^\gamma}$$

$$\text{SURVIVAL: } S(t) = \frac{1}{1+\alpha t^\gamma}$$

$$\text{HAZARD RATE : } h(t) = \frac{\alpha\gamma t^{\gamma-1}}{1+\alpha t^\gamma}$$

$$\text{CUMULATIVE HAZARD: } \ln(1 + \alpha t^\gamma)$$

$$t \geq 0, \quad \alpha > 0, \quad \gamma > 0.$$

The loglogistic distribution is characterized by two parameters  $\alpha$ , and  $\gamma$ . The median of the loglogistic distribution is  $\alpha^{-1/\gamma}$ . Figures 2.11, 2.12, and 2.13 show the loglogistic density, hazard, and survival functions with  $\alpha = 1$  and various values of  $\gamma = 2.0, 1$ , and  $0.67$ , respectively.

When  $\gamma > 1$ , the loglogistic hazard has the value 0 at time 0, increases to a peak at  $t = (\gamma - 1)^{1/\gamma}/\alpha^{1/\gamma}$ , and then declines, which is similar to the

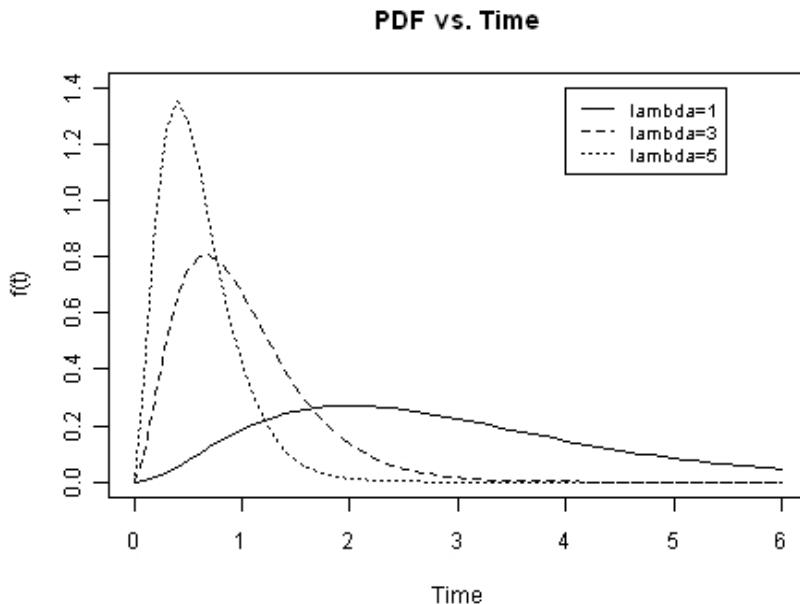


FIGURE 2.9: Gamma density functions with  $\gamma = 3$ .

lognormal hazard. When  $\gamma = 1$ , the hazard starts at  $\alpha^{1/\gamma}$  and then declines monotonically. When  $\gamma < 1$ , the hazard starts at infinity and then declines toward 0 as  $t$  approaches infinity. Thus, the loglogistic distribution may be used to describe a first increasing and then decreasing hazard or a monotonically decreasing hazard.

## 2.8 Maximum Likelihood Estimation

Maximum likelihood estimation begins with writing a mathematical expression known as the **Likelihood Function** of the sample data. Loosely speaking, the likelihood of a set of data is the probability of obtaining that particular set of data, given the chosen probability distribution model. This expression contains the unknown model parameters. The values of these parameters that maximize the sample likelihood are known as the **Maximum Likelihood Estimates** or **MLE's**.

Maximum likelihood estimation is a maximization procedure. It applies to every form of censored or multicensored data, and it is even possible to

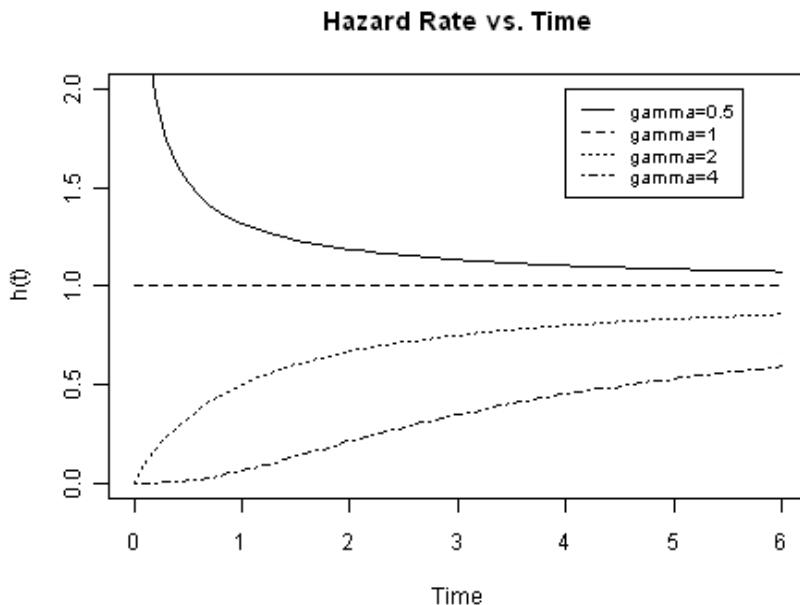


FIGURE 2.10: Gamma failure rate functions with  $\lambda = 1$ .

use the technique across several stress cells and estimate acceleration model parameters at the same time as life distribution parameters. Moreover, MLE's and likelihood functions generally have very desirable large sample properties:

- they become unbiased minimum variance estimators as the sample size increases
- they have approximate normal distributions and approximate sample variances that can be calculated and used to generate confidence bounds
- likelihood functions can be used to test hypotheses about models and parameters.

There are only two drawbacks to MLE's, but they are important ones:

- With small numbers of failures (less than 5, and sometimes less than 10 is small), MLE's can be heavily biased and the large sample optimality properties do not apply.
- Calculating MLE's often requires specialized software for solving complex non-linear equations. This is less of a problem as time goes by, as more statistical packages are upgrading to contain MLE analysis capability every year.

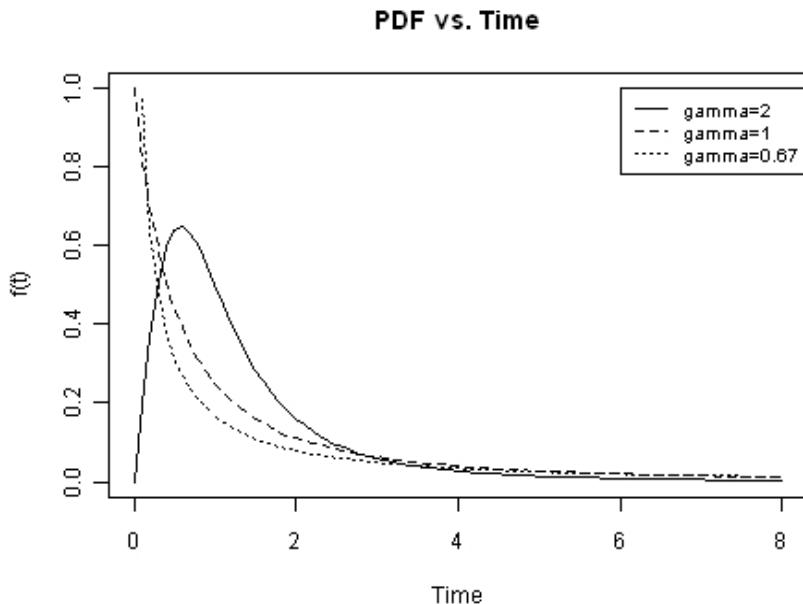


FIGURE 2.11: Density function of loglogistic distribution with  $\alpha = 1$ .

### Likelihood Function Examples for Survival Data:

Let  $f(t)$  be the PDF and  $F(t)$  the CDF for the chosen life distribution model. Note that these are functions of  $t$  and the unknown parameters of the model. Assuming the independence between censoring and failure time distributions, the likelihood function for censored data is:

$$L = C \prod_{i=1}^n [f(t_i)]^{\delta_i} [1 - F(t_i)]^{1-\delta_i}$$

where  $\delta_i = 1$ , if  $i$ -th observation is failed, and 0, otherwise (censored).  $C$  denoting a constant that plays no role when solving for the MLE's. Note that with no censoring, the likelihood reduces to just the product of the densities, each evaluated at a failure time. The likelihood function for Type I Censored data is:

$$L = C \left( \prod_{i=1}^r f(t_i) \right) (1 - F(T))^{n-r}$$

For Type II Censored Data, just replace T above by the random end of test time  $t_r$ .

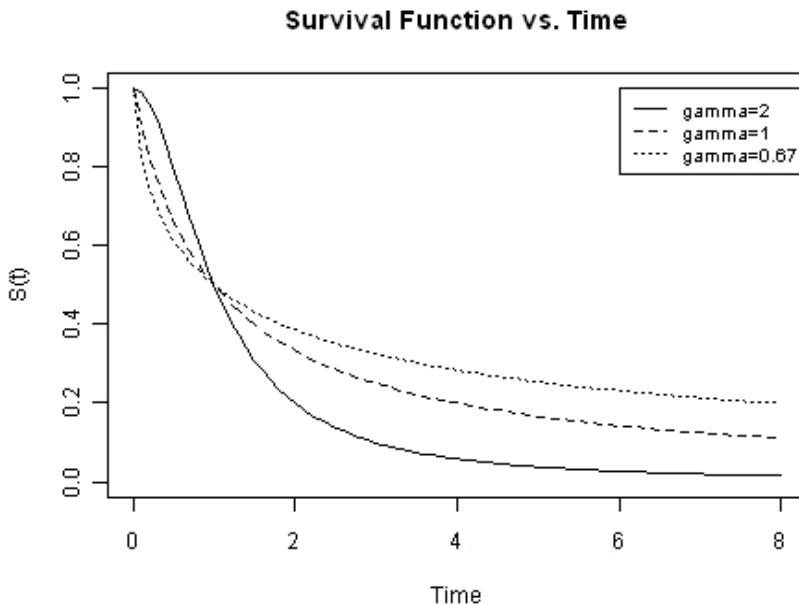


FIGURE 2.12: Survival function of loglogistic distribution with  $\alpha = 1$ .

The likelihood function for readout data is:

$$L = C \left( \prod_{i=1}^k [F(T_i) - F(T_{i-1})]^{r_i} \right) (1 - F(T_k))^{n - \sum_{i=1}^k r_i}$$

with  $F(T_0)$  defined to be 0.

In general, any multicensored data set likelihood will be a constant times a product of terms, one for each unit in the sample, that look like either  $f(t_i)$ ,  $[F(T_i) - F(T_{i-1})]$ , or  $[1 - F(t_i)]$ , depending on whether the unit was an exact time failure at time  $t_i$ , failed between two readouts  $T_{i-1}$  and  $T_i$ , or survived to time  $t_i$  and was not observed any longer.

The general mathematical technique for solving for MLE's involves setting partial derivatives of  $\ln L$  (the derivatives are taken with respect to the unknown parameters) equal to zero and solving the resulting (usually non-linear) equations. The equation for the exponential model can easily be solved.

#### MLE's for the Exponential Model (Type I Censoring):

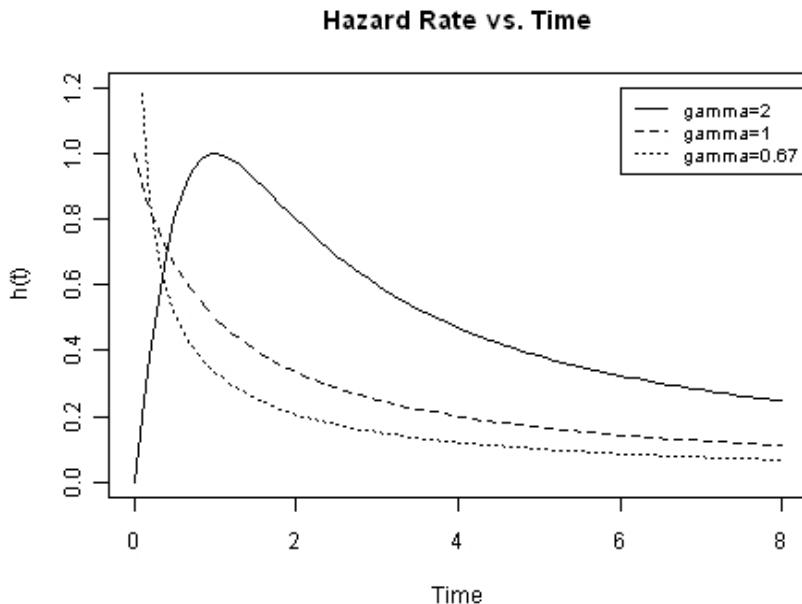


FIGURE 2.13: Failure rate function of loglogistic distribution with  $\alpha = 1$ .

$$\begin{aligned}
 L &= C\lambda^r e^{-\lambda \sum_{i=1}^r t_i} (e^{-\lambda(n-r)T}) \\
 \ln L &= \ln C + r \ln \lambda - \lambda \sum_{i=1}^r t_i - \lambda(n-r)T \\
 \frac{\partial \ln L}{\partial \lambda} &= \frac{r}{\lambda} - \sum_{i=1}^r t_i - (n-r)T = 0 \\
 \hat{\lambda} &= \frac{r}{\sum_{i=1}^r t_i + (n-r)T}.
 \end{aligned}$$

**NOTE :** The MLE of the failure rate in the exponential case turns out to be the total number of failures observed divided by the total time on test. For the MLE of the MTBF, take the reciprocal of this or use the total unit test hours divided by the total observed failures.

### Conclusions:

MLE analysis is an accurate and easy way to estimate life distribution

parameters, provided that a good software analysis package is available. The package should also calculate confidence bounds and loglikelihood values. SAS has this capability, as do several other commercial statistical packages.

---

## 2.9 Parametric Regression Models

Before discussing parametric regression models for survival data, let us introduce the accelerated failure time (AFT) model. Denote the survival functions of two populations by  $S_1(t)$  and  $S_2(t)$ , respectively. The AFT model is given by

$$S_1(t) = S_2(ct), \quad \text{for all } t \geq 0, \quad (2.1)$$

where  $c$  is a constant. This model implies that aging rate of population 1 is  $c$  times as much as that of population 2. For example, if  $S_1(t)$  is the survival function of rat population (in most of the clinical trials, rats are used before doing drug experiments on human beings) and  $S_2(t)$  is the survival function for the human population, then by convention a year for a rat is equivalent to 20 years for a human implies  $c = 20$ , and  $S_1(t) = S_2(20t)$ . So the probability that a rat can survive 3 years or beyond is the same as the probability that a human subject can survive 60 years or beyond.

Let  $\mu_i$  be the mean survival time for the population  $i$  and  $\varphi_i$  be the population quantile such that  $S_i(\varphi_i) = \theta$  for some  $\theta \in (0, 1)$ . Then

$$\begin{aligned} \mu_2 &= \int_0^\infty S_2(t)dt \\ &= c \int_0^\infty S_2(cu)du \quad (t = cu) \\ &= c \int_0^\infty S_1(u)du \\ &= c\mu_1 \end{aligned}$$

and

$$S_2(\varphi_2) = \theta = S_1(\varphi_1) = S_2(c\varphi_1).$$

Then we have

$$\varphi_2 = c\varphi_1.$$

Under the AFT model (2.1), one can note that from the above equations, the expected survival time, median survival time of population 2 all are  $C$  times as much as those of population 1.

Suppose we have a sample of size  $n$  from a target population. For subject  $i$  ( $i = 1, 2, \dots, n$ ), we have observed values of covariates  $y_{i1}, y_{i2}, \dots, y_{ip}$  and

possibly censored survival time  $T_i$ . The procedure **Proc Lifereg** in SAS fits models to data specified by the following equations.

$$\ln(T_i) = \beta_0 + \beta_1 y_{i1} + \dots + \beta_p y_{ip} + \sigma \epsilon_i, \quad (2.2)$$

where  $\beta_0, \dots, \beta_p$  are the regression coefficients of interest,  $\sigma$  is a scale parameter and  $\epsilon_i$  are the random errors terms, usually assumed to be independent and identically distributed (iid) with some density function  $F(\epsilon)$ . Equation (2.2) is similar to a linear regression model for the log-transformed response variable  $\ln(T_i)$  with  $\epsilon_i$  are iid from  $N(0,1)$ .

Let us increase the covariate  $y_k$  to  $y_k + 1$  and denote by  $T_1$  and  $T_2$  the corresponding survival times for the two populations with covariate values  $y_k$  and  $y_k + 1$  (with other covariate values fixed). Then  $T_1$  and  $T_2$  can be expressed as

$$T_1 = e^{\beta_0 + \beta_1 y_{i1} + \dots + \beta_k y_k + \dots + \beta_p y_{ip} + \sigma \epsilon_1} = c_1 e^{\sigma \epsilon_1}$$

$$T_2 = e^{\beta_0 + \beta_1 y_{i1} + \dots + \beta_k (y_k + 1) + \dots + \beta_p y_{ip} + \sigma \epsilon_2} = c_2 e^{\sigma \epsilon_2}$$

where  $c_1$  and  $c_2$  are related by  $c_2 = c_1 e^{\beta_k}$ . The corresponding survival functions are

$$S_1(t) = P[T_1 \geq t] = P[c_1 e^{\sigma \epsilon_1} \geq t] = P[e^{\sigma \epsilon_1} \geq c_1^{-1} t]$$

$$S_2(t) = P[T_2 \geq t] = P[c_2 e^{\sigma \epsilon_2} \geq t] = P[e^{\sigma \epsilon_2} \geq c_2^{-1} t]$$

Since  $\epsilon_1$  and  $\epsilon_2$  have the same distribution, and  $c_2 = c_1 e^{\beta_k}$ , we have

$$S_2(e^{\beta_k} t) = P[e^{\sigma \epsilon_2} \geq c_2^{-1} e^{\beta_k} t] = P[e^{\sigma \epsilon_2} \geq c_1^{-1} e^{-\beta_k} e^{\beta_k} t]$$

$$= P[c_1 e^{\sigma \epsilon_1} \geq t] = P[c_1 e^{\sigma \epsilon_1} \geq t] = S_1(t).$$

Therefore, we have accelerated failure time model between populations 1 (covariate value =  $y_k$ ) and 2 (covariate value =  $y_k + 1$ ) with  $c = e^{\beta_k}$ . So if we increase the covariate value of  $y_k$  by one unit while holding other covariate values unchanged, the corresponding average survival time  $\mu_2$  and  $\mu_1$  will be related by

$$\mu_2 = e^{\beta_k} \mu_1.$$

If  $\beta_k$  is small, then

$$\frac{\mu_2 - \mu_1}{\mu_1} = e^{\beta_k} - 1 \approx \beta_k.$$

Similarly we have for the population quantile  $\varphi_i$

$$\frac{\varphi_2 - \varphi_1}{\varphi_1} = e^{\beta_k} - 1 \approx \beta_k.$$

Therefore, when  $\beta_k$  is small, it can be interpreted as the percentage increase if  $\beta_k > 0$  or percentage decrease if  $\beta_k < 0$  in the average survival time and/or median survival time when we increase the covariate value of  $y_k$  by one unit. Thus the greater the value of the covariate with positive  $\beta_k$  is more beneficial

in improving survival time for the target population. This interpretation of  $\beta_k$  is very similar to that in a linear regression.

We can assume different distribution for the error term  $\epsilon_i$  in the model (2.2). For example, we can assume  $\epsilon_i$  follows iid  $N(0,1)$ . This assumption is equivalent to assuming  $T_i$  has lognormal distribution (conditional on the covariates  $y$ 's). In this section, we introduce exponential, Weibull, and lognormal parametric regression models for  $T_i$  (equivalently for  $\epsilon_i$ ) and the remaining accelerated parametric regression models can be done in a similar way.

### Exponential Model:

The simplest model is the exponential model where  $T$  at  $y = 0$  (usually referred as baseline) has exponential distribution with constant hazard  $\exp(-\beta_0)$ . This is equivalent to assuming that  $\sigma = 1$  and  $\epsilon$  has a standard extreme value distribution given by

$$f(\epsilon) = \exp(\epsilon - e^\epsilon).$$

From this argument, it is easy to see that the distribution of  $T$  at any covariate vector  $y$  is exponential with constant hazard (independent of  $t$ )

$$h(t|y) = \exp(-\beta_0 - \beta_1 y_1 - \dots - \beta_p y_p).$$

If we increase the value of covariate  $y_k$  ( $k = 1, \dots, p$ ) by one unit from  $y_k$  to  $y_k + 1$  while holding other covariate values fixed, then the ratio of the corresponding hazards is equal to

$$\frac{\lambda(t|y_k + 1)}{\lambda(t|y_k)} = e^{-\beta_k}.$$

Thus  $e^{-\beta_k}$  can be interpreted as the hazard ratio corresponding to one unit increase in the covariate  $y_k$ , or equivalently,  $\beta_k$  can be interpreted as the decrease in log-hazard as the value of covariate  $y_k$  increases by one unit (while other covariate values being held fixed).

### Weibull Model:

In this case, the distribution of  $\sigma\epsilon$  is an extreme value distribution with scale parameter  $\sigma$ . The survival function of  $T$  at covariate value  $y = (1, y_1, \dots, y_p)'$  is

$$S(t|y) = \exp \left\{ - \left[ t e^{-y' \beta} \right]^{\frac{1}{\sigma}} \right\},$$

where  $\beta = (\beta_0, \dots, \beta_p)'$  is a vector of regression coefficients. Equivalently, in terms of log hazard function

$$\ln h(t|y) = \left( \frac{1}{\sigma} - 1 \right) \ln t - \ln \sigma - y' \beta / \sigma$$

the above expression can be written as

$$\ln h(t|y) = (\alpha - 1) \ln t + \beta_0^* + y_1 \beta_1^* + \dots + y_p \beta_p^*$$

where  $\alpha = 1/\sigma$ ,  $\beta_0^* = -\ln \sigma - \beta_0/\sigma$ , and  $\beta_j^* = -\beta_j/\sigma$  for  $j = 1, \dots, p$ . We also get proportional hazard model and the coefficient  $\beta_k^*$  ( $k = 1, \dots, p$ ) also has the interpretation that it is the increase in log-hazard when the value of covariate  $y_k$  increase by one unit while other covariate values being held fixed. The function

$$\lambda_0(t) = t^{\alpha-1} e^{\beta_0^*} = \alpha t^{\alpha-1} e^{-\alpha \sigma}$$

is the baseline hazard ( i.e., when  $\mathbf{y} = 0$ ).

### Lognormal Model:

The lognormal model assumes that  $\epsilon$  follow  $N(0,1)$ . Let  $h_0(t)$  be the hazard function of  $T$  when  $\beta = 0$ . Then  $h_0(t)$  can be written as

$$h_0(t) = \frac{\phi(\frac{\ln t}{\sigma})}{[1 - \Phi(\frac{\ln t}{\sigma})] \sigma t},$$

where  $\phi(t)$  is the pdf and  $\Phi(t)$  is the CDF of the standard normal distribution. Then log-hazard function of  $T$  at any covariate value  $y$  can be expressed as

$$\ln h_0(t|y) = \ln h_0(te^{y' \beta}) - y' \beta.$$

It is clear from the above expression that the above model is not a proportional hazard model. The survival function

$$S(t|\mathbf{y}) = \beta_0^* + \beta_1^* y_1 + \dots + \beta_p^* y_p - \alpha \ln(t),$$

or equivalently

$$S(t|\mathbf{y}) = \Phi[\beta_0^* + \beta_1^* y_1 + \dots + \beta_p^* y_p - \alpha \ln(t)],$$

where  $\alpha = 1/\sigma$  and  $\beta_i/\sigma$  for  $i = 0, 1, \dots, p$ . This is a probit regression model with intercept depending on  $t$ .

### Loglogistic Model:

The loglogistic model assumes that the error term  $\epsilon$  has a standard logistic distribution

$$f(\epsilon) = \frac{e^\epsilon}{(1 + e^\epsilon)^2}.$$

The hazard function of  $T$  at any covariate value  $\mathbf{y}$  has a closed form

$$\lambda(t|\mathbf{y}) = \frac{\alpha t^{\alpha-1} e^{-\mathbf{y}' \beta / \sigma}}{1 + t^\alpha e^{-\mathbf{y}' \beta / \sigma}},$$

where  $\alpha = 1/\sigma$ .

The random variable  $T$  has survival function at covariate value  $\mathbf{y}$

$$S(t|\mathbf{y}) = \frac{1}{1 + (te^{-\mathbf{y}'\beta})^{1/\sigma}}.$$

After some algebra, the above equation leads to

$$\ln \left[ \frac{S(t|\mathbf{y})}{1 - S(t|\mathbf{y})} \right] = \beta_0^* + \beta_1^* y_1 + \dots + \beta_p^* y_p - \alpha \ln(t),$$

where  $\alpha = 1/\sigma$  and  $\beta_i/\sigma$  for  $i = 0, 1, \dots, p$ . This is nothing but a logistic regression model with the intercept depending on  $t$ . Since  $S(|\mathbf{y})$  is the probability of surviving to time  $t$  for any given time  $t$ , the ratio  $S(t|\mathbf{y})/(1 - S(t|\mathbf{y}))$  is often called the **odds** of surviving to time  $t$ . Therefore, with one unit increase in  $y_k$  while other covariate being fixed, the **odds ratio** is given by

$$\frac{S(t|\mathbf{y}+1)/(1 - S(t|\mathbf{y}+1))}{S(t|\mathbf{y})/(1 - S(t|\mathbf{y}))} = e^{\beta_k^*} \quad \text{for all } t \geq 0,$$

which is constant over time. Therefore, we have a proportional odds models. Hence  $\beta_k^*$  can be interpreted as the log odds ratio (for surviving) with one unit increase in  $y_k$  and  $-\beta^*$  is the log odds ratio of dying before time  $t$  with one unit increase in  $y_k$ . At the times when the event of failure is rare (such as the early phase of a study),  $-\beta_k^*$  can also be approximately interpreted as the log relative risk of dying. The loglogistic model is the only one that is both AFT model and a proportional odds model.

Obviously,  $\epsilon$  has the following cumulative distribution

$$F(u) = \frac{e^u}{1 + e^u}, \quad u \in (-\infty, \infty),$$

whose inverse function

$$\text{logit}(\pi) = \ln \left( \frac{\pi}{1 - \pi} \right), \quad \psi \in (0, 1),$$

is often called the logit function.

### Gamma Model:

For a given set of covariates  $(y_1, y_2, \dots, y_p)$ , let  $\lambda = e^{\beta_0 + y_1\beta_1 + \dots + y_p\beta_p} = e^{\mathbf{y}'\beta}$ . Then  $\log(T) = \mathbf{y}'\beta + \sigma\epsilon$  implies  $T = r^{\mathbf{y}'\beta}[e^\epsilon]^\sigma = \lambda T_0^\sigma$ .  $T_0$  is distributed as standard gamma and  $\log T_0$  is distributed as standard log gamma.

### 2.9.1 Example 2.1

The Myeloma data discussed in Section 1.9 is revisited here in this example. We obtain parametric Weibull regression model for assessing the nine covariates on the survival time using R-programm.

```
myeloma=read.table(file="myeloma.txt",header=T)
library(survival)
fit=survreg(Surv(T,S)~LBUN+HGB+P+Age+LW+F+LPBM+Pr+SC,myeloma,
dist="weibull")
summary(fit)
```

The output of the R-programm is

```
Call:
survreg(formula = Surv(T, S) ~ LBUN + HGB + P + Age + LW + F +
    LPBM + Pr + SC, data = myeloma, dist = "weibull")
             Value Std. Error      z      p
(Intercept) 7.12248   2.7422  2.597 0.00939
LBUN        -1.53251   0.5412 -2.832 0.00463
HGB         0.09698   0.0621  1.561 0.11855
P            0.26546   0.4557  0.583 0.56017
Age          0.01032   0.0169  0.611 0.54106
LW           -0.40081   0.5989 -0.669 0.50331
F            -0.33236   0.3526 -0.943 0.34589
LPBM        -0.38709   0.4290 -0.902 0.36695
Pr           -0.00923   0.0228 -0.404 0.68617
SC           -0.09985   0.0898 -1.112 0.26611
Log(scale)   -0.14255   0.1070 -1.333 0.18261

Scale= 0.867

Weibull distribution
Loglik(model)= -206.1  Loglik(intercept only)= -215.1
                 Chisq= 17.95 on 9 degrees of freedom, p= 0.036
Number of Newton-Raphson Iterations: 5
n= 65
```

From the output of R-programm it is observed that the test for the regression parameters equal to zero is rejected with chi-square value 17.95 for 9 df and p-value is 0.036. log(BUN) is the most effective variable which is related to the survival of patients. Similar analysis can be done using other baseline distributions like lognormal, and loglogistic. The maximum of loglik (intercept only) among these distribution gives the best fit for the parametric model. The loglikelihood for Weibull baseline is -215.1, for lognormal baseline is -215.3 and for loglogistic baseline is -216. Weibull baseline has the maximum loglikelihood.

The SAS programm commands can be written to analyze this data is as follows. We will not give the output of SAS because of large output and similar results.

```

data Myeloma;
infile 'myeloma.txt';
input LBUN HGB P Age LW FLPBM Pr SC;
label Time='Survival Time'
      Status='0=Alive 1=Dead';
run;
proc phreg data=Myeloma;
model Time*Status(0)=LBUN HGB P Age LW FLPBM Pr
SC/dist=lognormal;
run;

```

### 2.9.2 Example 2.2

The data on times of infection of kidney dialysis discussed in Section 1.4 is revisited here in this example. We obtain parametric Weibull regression model for assessing the catheter placement on the survival time using R-programm.

```

> dialysis=read.table(file="C:/David/Book2/Data/dialysis.txt",
  header=T)
> fit1=survreg(Surv(time,ind)~place,data=dialysis,
  dist='weibull')
> summary(fit1)

Call:
survreg(formula = Surv(time, ind) ~ place, data = dialysis,
       dist = "weibull")

            Value Std. Error     z      p
(Intercept) 2.974      0.682 4.36 1.28e-05
place        0.623      0.469 1.33 1.84e-01
Log(scale)   0.129      0.167 0.77 4.41e-01

Scale= 1.14

Weibull distribution
Loglik(model)= -122    Loglik(intercept only)= -122.9
                  Chisq= 1.93 on 1 degrees of freedom, p= 0.16
Number of Newton-Raphson Iterations: 7
n= 119

```

From the output of R-programm it is observed that the test for the regression parameters equal to zero is not rejected with chi-square value 1.93 for 1 df and p-value is 0.16. Catheter placement is not significant covariate related to the survival of patients. Similar analysis can be done using other baseline distributions like lognormal, and loglogistic. The maximum of loglik (intercept only) among these distribution gives the best fit for the parametric model. The loglikelihood for Weibull baseline is -123.9, for lognormal base-

line is -123.6 and for loglogistic baseline is -123.3. Weibull baseline has the maximum loglikelihood.

# Chapter 3

---

## Nonparametric and Semiparametric Models

3.1	Empirical Survival Function .....	43
3.2	Graphical Plotting .....	44
3.2.1	Probability Plotting .....	44
3.2.2	Hazard and Cumulative Hazard Plotting .....	47
3.2.3	Exponential and Weibull Hazard Plots .....	48
3.3	Graphical Estimation .....	50
3.4	Empirical Model Fitting: Distribution Free (Kaplan-Meier) Approach ..	51
3.5	Comparison between Two Survival Functions .....	59
3.6	Cox's Proportional Hazards Model .....	62

Survival data are conveniently summarized through estimates of the survival function and hazard function. Methods of estimating these functions from a sample of survival data are said to be *nonparametric* or *distribution-free*, since they do not require specific assumptions to be made about the underlying distribution of the survival times. An initial step in the analysis of survival data is to present numerical or graphical summaries of the survival times for individuals in a particular group. Such summaries may be of interest in their own right, or as a precursor to a more detailed analysis of the data. Once the estimated survival function has been found, the median and other percentiles of the distribution of survival times can be estimated. When the survival times of two groups of patients are being compared, an informal comparison of the survival experience of each group of individuals can be made using the estimated survival functions. However, there are more formal procedures that enable two groups of survival data to be compared. Nonparametric procedure for comparing two or more groups of survival times is the *logrank test* which is the most powerful test against the alternatives that the hazard functions are proportional.

---

### 3.1 Empirical Survival Function

Suppose we have a single sample of survival times, where none of the observations are censored. The survivor function  $S(t)$ , is the probability that

an individual survives for a time greater than or equal to  $t$ . This function can be estimated by the **empirical survival function**, given by

$$\hat{S}(t) = \frac{\text{Number of individuals with survival times } \geq t}{\text{Number of individuals in the data set}}.$$

Note that the empirical survival function is equal to unity for values of  $t$  before the first death time, and zero after the final death time. The estimated survival function  $\hat{S}(t)$  is assumed to be constant between two adjacent times, and so a plot of  $\hat{S}(t)$  against  $t$  is a step function. The function decreases immediately after each observed survival time.

---

## 3.2 Graphical Plotting

Graphical plots of survival data are quick, useful visual tests of whether a particular model is consistent with the observed data. The basic idea behind virtually all graphical plotting techniques is the following:

If the survival data consist of (possibly multicensored) failure data from a population then the models are life distribution models such as the exponential, Weibull, or lognormal.

The kinds of plots we will consider for failure data are:

- Probability (CDF) plots
- Hazard and Cum Hazard plots

**NOTE:** Many of the plots discussed in this chapter can also be used to obtain quick estimates of model parameters. This will be covered in later sections. While there may be other, more accurate ways of estimating parameters, simple graphical estimates can be very handy, especially when other techniques require software programs that are also readily available.

### 3.2.1 Probability Plotting

Probability plots are simple visual ways of summarizing survival data by plotting CDF estimates vs time.

**Plotting Positions:** Censored Data (Type I or Type II)

At the time  $t_i$  of the  $i$ -th failure, we need an estimate of the CDF (or the Cum. Population Percent Failure). The simplest and most obvious estimate is just  $100 \times i/n$  (with a total of  $n$  units on test). This, however, is generally

an overestimate (i.e. biased). Various texts recommend corrections such as  $100 \times (i - .5)/n$  or  $100 \times i/(n + 1)$ . Here, we recommend what are known as (approximate) **median rank** estimates:

Corresponding to the time  $t_i$  of the  $i$ -th failure, use a CDF or Percentile estimate of  $100 \times (i - .3)/(n + .4)$ .

### Plotting Positions: Readout Data

Let the readout times be  $T_1, T_2, \dots, T_k$  and let the corresponding new failures recorded at each readout be  $r_1, r_2, \dots, r_k$ . Again, there are  $n$  individuals on test. Corresponding to the readout time  $T_j$ , use a CDF or percentile estimate of

$$\frac{100 \times \sum_{i=1}^j r_i}{n}.$$

### Plotting Positions: Multicensored Data

The calculations are more complicated for multicensored data. K-M estimates (described in the [Section 3.7](#)) can be used to obtain plotting positions at every failure time. The more precise Modified K-M Estimates are recommended. They reduce to the censored type I or the censored type II median rank estimates when the data consist of only failures, without any removals except possibly at the end of the study period.

The general idea is to take the model CDF equation and write it in such a way that a function of  $F(t)$  is a linear equation of a function of  $t$ . This will be clear after a few examples. In the formulas that follow, “ln” always means “natural logarithm”, while “log” always means “base 10 logarithm”.

a) **Exponential Model:** Take the exponential survival function and rewrite it as

$$-\ln S(t) = \lambda t \quad \text{or, equivalently.}$$

If  $-\ln S(t)$  is linear in  $t$  with slope  $\lambda$ , then exponential model is a good fit.

b) **Weibull Model:** Take the Weibull survival function and rewrite it as

$$-\ln(-\ln S(t)) = \gamma \ln t - \gamma \ln \alpha.$$

If  $\ln(-\ln S(t))$  is linear in  $\ln t$  with slope  $\gamma$  and intercept  $-\ln \alpha$ , then Weibull model is a good fit.

c) **Lognormal Model:** Take the lognormal survival function and rewrite it as

$$\ln t = \sigma \Phi^{-1}\{1 - S(t)\} + \ln T_{50},$$

(1) Fail # = i	(2) Time of Fail (x)	(3) $F(t_i)$ estimate ( $i - 0.3$ ) / 20.4	(4) $\ln\{1/(1 - F(t_i))\}$ (y)
1	54	.034	.035
2	187	.083	.087
3	216	.132	.142
4	240	.181	.200
5	244	.230	.262
6	335	.279	.328
7	361	.328	.398
8	373	.377	.474
9	375	.426	.556
10	386	.475	.645

TABLE 3.1: Empirical CDF calculations

where  $\Phi^{-1}$  denoting the inverse function for the standard normal distribution. If  $\ln t$  is linear in  $\Phi^{-1}\{1 - S(t)\}$  with slope  $\sigma$  and intercept  $\ln T_{50}$ , then lognormal model is a good fit.

d) **Extreme Value Distribution (Type I - for minimum):** Take the survival function of extreme value distribution and rewrite it as

$$\ln \{-\ln S(t)\} = (t - \mu)/\beta.$$

If  $\ln \{-\ln S(t)\}$  is linear in  $t$  with slope  $1/\beta$  and intercept  $-\mu/\beta$ , then extreme value model is a good fit.

**Example 3.1:** To generate Weibull random failure times, we generate 20 Weibull failure times with a shape parameter of  $\gamma = 1.5$  and  $\alpha = 500$ . Assuming a test time of  $T = 500$  hours, only 10 of these failure times would have been observed. They are, to the nearest hour: 54, 187, 216, 240, 244, 335, 361, 373, 375, and 386. First we will compute plotting position survival function estimates based on these failure times, and then a probability plot is plotted. Table 3.1 gives the empirical CDF and  $-\ln S(t)$  for this data.

Figure 3.1 shows Weibull plot of  $\ln \ln(1/(1 - F(t)))$  versus  $\ln t$  for the data given in Example 3.1.

Note that the configuration of points appears to have some curvature. This is mostly due to the very first point on the plot (the earliest time of failure). The first few points on a probability plot have more variability than points in the central range and less attention should be paid to them when visually testing for “straightness”.

This would give a slope estimate of 1.46, which is close to the 1.5 value used in the simulation.

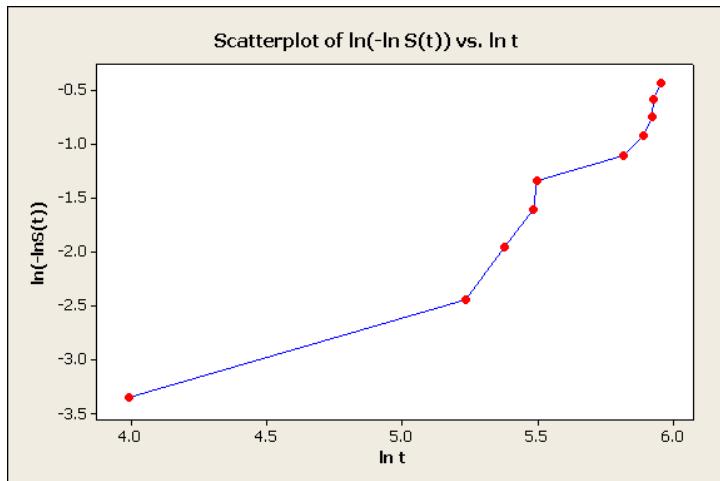


FIGURE 3.1: Graphical Weibull fitting plot.

The intercept is  $-4.114$  and setting this equal to  $-\gamma \log \alpha$  we estimate  $\alpha = 657$  (the “true” value used in the simulation was 500).

### 3.2.2 Hazard and Cumulative Hazard Plotting

Since probability plots are generally more useful, we will only give a brief description of hazard plotting. We describe how to make cumulative probability plots in the following steps.

1. Order the failure times and running times for each of the  $n$  units on test in ascending order from 1 to  $n$ . The order is called the rank of an individual. Calculate the reverse rank for each individual (reverse rank =  $n - \text{rank} + 1$ ).
2. Calculate a Hazard “value” for every failed individual (do this only for the failed individuals). The Hazard value for the failed individual with reverse rank  $k$  is just  $1/k$ .
3. Calculate the cumulative hazard values for each failed individual. The cumulative hazard value corresponding to a particular failed individual is the sum of all the hazard values for failed individuals with ranks up to and including that failed individual.
4. Plot the time of fail vs the cumulative hazard value as covered below for the exponential and the Weibull model.

**Example 3.2:** Ten cancer patients in a clinical trial were under observation up to 250 days. Six patients dead at 37, 73, 132, 195, 222, and 248 days. Four

patients discontinued the treatment and admitted to another hospital at the following times: 50, 100, 200, and 250 days. Cumulative hazard values were computed in the following Table 3.2.

(1) Time of Event	(2) 1=Failure 0 = Censored	(3) Rank	(4) Reverse Rank	(5) Hazard Value (2)/(4)	(6) Cum. Hazard Value
37	1	1	10	1/10	.10
50	0	2	9		
73	1	3	8	1/8	.225
100	0	4	7		
132	1	5	6	1/6	.391
195	1	6	5	1/5	.591
200	0	7	4		
222	1	8	3	1/3	.924
248	1	9	2	1/2	1.424
250	0	10	1		

TABLE 3.2: Empirical cumulative hazard calculations

Next ignore the rows with no cumulative hazard value and plot column (1) vs column (6) of Table 3.2.

### 3.2.3 Exponential and Weibull Hazard Plots

The cumulative hazard for the exponential is just  $H(t) = \lambda t$ , which is linear in  $t$  with a 0 intercept. So a simple linear graph paper plot of  $y = \text{col} (6)$  vs  $x = \text{col} (1)$  should line up as approximately a straight line going through the origin with slope  $\lambda$  if the exponential model is appropriate. [Figure 3.2](#) gives the plot of  $H(t)$  versus  $t$  for the data given in Example 3.2.

The cumulative hazard for the Weibull is  $H(t) = (t/\alpha)^\gamma$ , so a plot of  $\ln H(t) = \ln \text{col}(6)$  vs  $\ln t = \ln \text{col}(1)$  resemble a straight line with slope  $\gamma$ , if the Weibull model is appropriate. [Figure 3.3](#) gives the probability plot for Weibull distribution for the data given in Example 3.2.

The Weibull fit looks better, although the slope estimate is 1.27, which is not far from an exponential model slope of 1. Of course, with a sample of just 10, and only 6 failures, it is difficult to pick a model from the data alone.

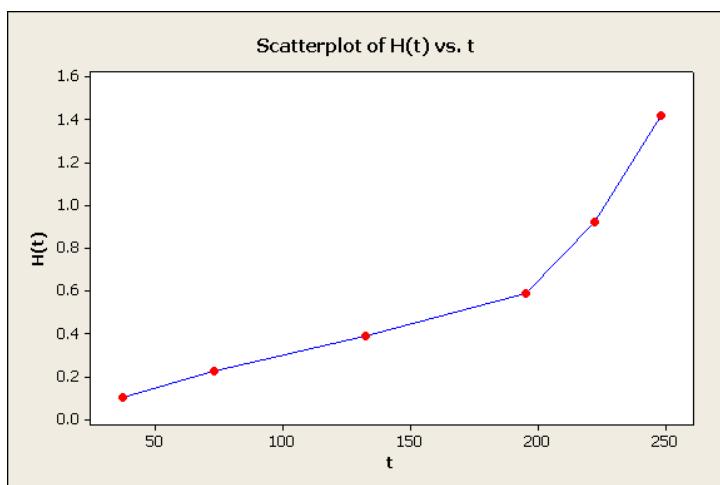


FIGURE 3.2: Graphical exponential fitting of the data.

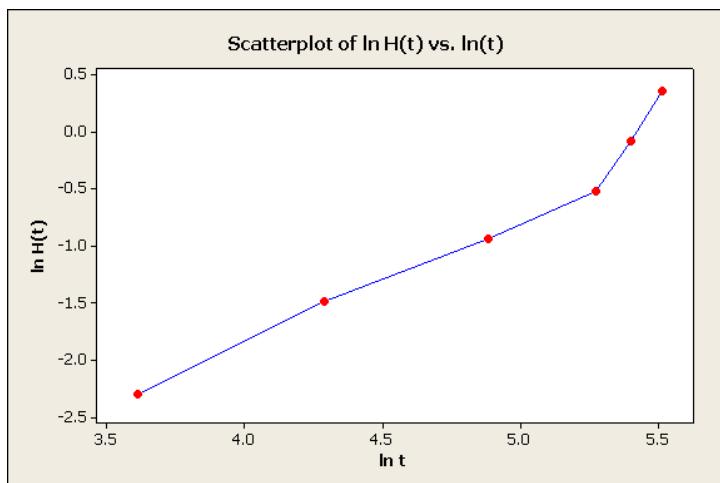


FIGURE 3.3: Graphical Weibull fitting of the data.

### 3.3 Graphical Estimation

Once you have calculated plotting positions from your failure data, and put the points on the appropriate graph paper for your chosen model, parameter estimation follows easily. But along with the mechanics of graphical estimation, be aware of both the advantages and the disadvantages of graphical estimation methods.

#### Graphical Estimation Mechanics

If you draw a line through the points, and the paper is commercially designed probability paper, there are usually simple rules to find estimates of the slope (or shape parameter) and the scale parameter. On lognormal paper with time on the  $x$ -axis and cumulative percent on the  $y$ -axis, draw horizontal lines from the 34th and the 50th percentiles across to the line, and drop vertical lines to the time axis from these intersection points. The time corresponding to the 50th percentile is the  $T_{50}$  estimate. Divide  $T_{50}$  by the time corresponding to the 34th percentile (this is called  $T_{34}$ ). The natural logarithm of that ratio is the estimate of sigma, or the slope of the line ( $\sigma = \ln(T_{50}/T_{34})$ ).

On commercial Weibull probability paper there is often a horizontal line through the 62.3 percentile point. That estimation line intersects the line through the points at a time that is the estimate of the characteristic life parameter  $\alpha$ . In order to estimate the line slope (or the shape parameter  $\gamma$ ), some papers have a special point on them called an estimation point. You drop a line from the estimation point perpendicular to the fitted line and look at where it passes through a special **estimation scale**. The estimate of  $\gamma$  is read off the estimation scale where the line crosses it.

Other papers may have variations on the methods described above. To remove the subjectivity of drawing a line through the points, a least squares (regression) fit can be performed.

#### Advantages of Graphical Methods of Estimation

- Graphical methods are quick and easy to use and make visual sense.
- Calculations can be done with little or no special software needed.
- Visual test of model (i.e., how well the points line up) is an additional benefit.

## Disadvantages of Graphical Methods of Estimation

The statistical properties of graphical estimates (i.e., how precise are they on the average) are not good for

- they are biased
- even with large samples, they do not become minimum variance (i.e., most precise) estimates
- graphical methods do not give confidence intervals for the parameters (intervals generated by a regression program for this kind of data are incorrect)
- formal statistical tests about model fit or parameter values cannot be performed with graphical methods.

As we have seen in the last chapter, maximum likelihood estimates overcome all these disadvantages - at least for survival data sets with a reasonably large number of failures - at a cost of losing all the advantages listed above for graphical estimation.

---

### 3.4 Empirical Model Fitting: Distribution Free (Kaplan-Meier) Approach

Kaplan and Meier (1958) proposed an estimator called as Kaplan-Meier (K-M) Product Limit estimator which provides quick, simple estimates of the survival function or the CDF based on failure data that may even be multicensored. No underlying model (such as Weibull or lognormal) is assumed; K-M estimation is an empirical (non-parametric) procedure. Exact times of failure are required.

#### Calculating Kaplan-Meier Estimates

The steps for calculating K-M estimates are the following:

1. Order the actual failure times from  $t_1$  through  $t_r$ , where there are  $r$  failures
2. Corresponding to each  $t_i$ , associate the number  $n_i$ , with  $n_i =$  the number of operating units just before the the  $i$ -th failure occurred at time  $t_i$
3. Estimate  $S(t_1)$  by  $(n_1 - 1)/n_1$

4. Estimate  $S(t_i)$  by  $S(t_{i-1})(n_i - 1)/n_i$
5. Estimate the CDF  $F(t_i)$  by  $1 - S(t_i)$ .

Note that unfailed individuals taken off test (i.e., censored) only count up to the last actual failure time before they were censored. They are included in the  $n_i$  counts up to and including that failure time, but not after.

**Example 3.3:** A simple example will illustrate the K-M procedure. Assume that 20 patients with some disease in a clinical study are on life test and 6 deaths occur at the following times: 10, 32, 56, 98, 122, and 181 days. There were four patients discontinued from the clinical study at the following times: 50, 100, 125, and 150 days. The remaining 10 patients were survived at the end of the clinical study, i.e., on 200th day. The K-M estimates for this life test are:

$$\begin{aligned}S(10) &= 19/20 \\S(32) &= 19/20 \times 18/19 \\S(56) &= 19/20 \times 18/19 \times 16/17 \\S(98) &= 19/20 \times 18/19 \times 16/17 \times 15/16 \\S(122) &= 19/20 \times 18/19 \times 16/17 \times 15/16 \times 13/14 \\S(181) &= 19/20 \times 18/19 \times 16/17 \times 15/16 \times 13/14 \times 10/11.\end{aligned}$$

## A General Expression for K-M Estimates

A general expression for the K-M estimates can be written. Assume that we have  $n$  individuals on test and order the observed lifetimes for these  $n$  individuals from  $t_1$  to  $t_n$ . Some of these are actual failure times and some are running times for individuals taken off test before they die. Suppose there  $r$  deaths have occurred, and the ordered death times are  $t_{(1)}, \dots, t_{(r)}$ , where  $r \leq n$ . The number individuals who are alive just before time  $t_{(j)}$ , including those who are about to die at this time, will be denoted  $n_j$ ,  $j = 1, 2, \dots, r$ , and  $d_j$  will denote the number who die at this time. The probability that an individual dies during the interval from  $t_{(j)} - \delta$  to  $t_{(j)}$  is estimated by  $d_j/n_j$  where  $\delta$  is an infinitesimal time interval. The corresponding estimated probability of survival through that interval is then  $(n_j - d_j)/n_j$ . The probability of surviving through the interval from  $t_{(k)}$  to  $t_{(k+1)}$ , and all preceding intervals, and leads to the Kaplan-Meier estimate of the survival function, which is given by

$$\hat{S}(t) = \prod_{j=1}^k \left( \frac{n_j - d_j}{n_j} \right),$$

for  $t_{(k)} \leq t < t_{(k+1)}$ ,  $k = 1, 2, \dots, r$ . A plot of Kaplan-Meier estimate of the

survival function is a step function, in which the estimated survival probabilities are constant between adjacent death times and decrease at each death time.

The estimated variance of the estimate of  $S(t)$  and is given by

$$\text{var}\{\hat{S}(t) \approx [\hat{S}(t)]^2 \sum_{j=1}^k \frac{d_j}{n_j(n_j - d_j)}.$$

The standard error of the K-M estimate of survival function is

$$\text{se}(\hat{S}(t) \approx \hat{S}(t) \left\{ \sum_{j=1}^k \frac{d_j}{n_j(n_j - d_j)} \right\}^{\frac{1}{2}}$$

for  $t_{(k)} \leq t < t_{(k+1)}$ .

### Nelson-Aalen Estimator

The most common estimate of integrated hazard is the Nelson-Aalen estimate which was first introduced by Nelson (1969) based on counting process and is given by

$$\hat{H}(t) = \sum_{i:t_i \leq t} \frac{\bar{N}(t_i) - \bar{N}(t_i-)}{\bar{Y}(t_i)}, \quad (3.1)$$

where  $\bar{Y}(t_i) = \sum_j Y_j(t_i)$ ,  $\bar{N}(t_i) = \sum_j N_j(t_i)$ .

The  $N_j(t)$  is the counting process, that is, the number of observed events in  $[0,t]$  for the unit  $j$  and  $Y(t)$  is a predictable process, a process whose value at time  $t$  is known infinitesimally before  $t$ . The Nelson-Aalen estimator is essentially a method of moment estimator. Its variance is estimated consistently by

$$\text{var}[\hat{H}(t)] = \sum_{i:t_i \leq t} \frac{\bar{N}(t_i) - \bar{N}(t_i-)}{\bar{Y}^2(t_i)}, \quad (3.2)$$

Breslow (1972) suggested a nonparametric estimate for survival function and is given by  $\hat{S}_B(t) = \exp[-\hat{H}(t)]$ . Fleming and Harrington (1984) showed the close relationship between the Breslow and Kaplan-Meier estimators, and compared them numerically for several sample sizes and censoring percentages. Let  $d\hat{H}(t_i) = [\bar{N}(t_i) - \bar{N}(t_i-)]/\bar{Y}(t_i)$ , the increment in the Nelson-Aalen (N-A) estimator at the  $i$ -th failure. Then Breslow estimator can be written as

$$\hat{S}_B(t) = \prod_{j:t_j \leq t} e^{-d\hat{H}(t_j)}$$

and the Kaplan-Meier estimate is

$$\hat{S}_{KM}(t) = \prod_{j:t_j \leq t} [1 - d\hat{H}(t_j)].$$

Since  $e^{-x} \approx 1 - x$  for small  $x$ , the two estimators are quite similar when the increments  $d\hat{H}$  are small, that is, when there are many subjects still at risk. The two estimates are asymptotically equivalent, since as  $n \rightarrow \infty$  the individual increments get arbitrarily small. Since  $e^{-x} \geq 1-x$ ,  $\hat{S}_B(t) \geq \hat{S}_{KM}(t)$  in finite samples. If the largest time  $T$  in a data set is a death,  $\hat{S}_{KM}(t) = 0$ , but  $\hat{S}_B(t)$  is positive.

### Example 3.4:

The data on kidney infection is presented in Section 1.5 in Chapter 1. We plot survival function for the time to first infection or censoring at the point of insertion of the catheter based K-M estimate and Breslow (or N-A) estimate. We also obtain 95% confidence intervals for K-M estimates. We can very easily observe from the survival function of the graph that Breslow (or N-A) estimates are larger than K-M estimates. The R-program commands to draw this graph are as follows.

```
jpeg(file="D:/David/Book2/surv01.jpg",height=5,width=6.5,
units="in",quality=100, bg="white", par(mfrow=c(1,1)), res=72)
kidney=read.table(file="D:/David/Book2/kidney.txt", header=T)
afit=survfit(Surv(T1,I1)^~1, kidney, type="fleming-harrington")
kfit=survfit(Surv(T1,I1)^~1, kidney, type="kaplan-meier")
plot(kfit, mark.time=F, main="Survival Function vs Time",
xlim=c(0,600), ylim=c(0,1), xlab="Survival Time(days)",
ylab="Survival Function", pch=" ")
lines(afit$time, afit$surv, type="s", lty=2)
legend(0,0.3, c("K-M", "N-A"), lty=c(1,2), cex=0.8)
dev.off()
```

The graph of survival functions based on K-M and Breslow(or N-A) estimates are displayed in [Figure 3.4](#) and it is observed from the graph that Breslow estimates are slightly larger than than K-M estimates.

### Example 3.5:

The data on bone marrow transplantation (BMT) is presented in Section 1.2 of Chapter 1. Patient's age, sex, and disease group (1-ALL, 2-AML low risk, 3-AML high risk) were considered as covariates. We compare the graph of survival function based on K-M and Breslow (or N-A) estimates in three different groups (1-ALL, 2-AML low risk, 3-AML high risk) using R-program. The R-program commands are as follows.

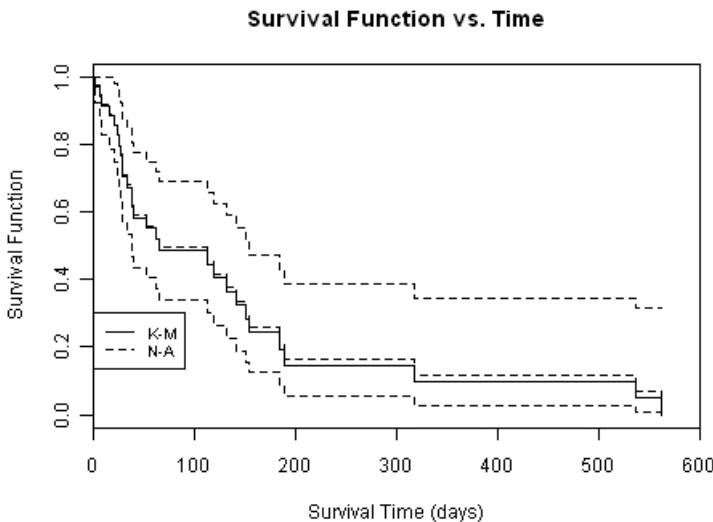


FIGURE 3.4: K-M and Breslow(or N-A) estimates for kidney infection data set with 95% CI for K-M.

```

jpeg(file="surv0.jpg",height=5,width=6.5,units="in",quality=100,
bg="white",par(mfrow=c(1,1)),res=72)
bone1=read.table(file="D:/David/SurvivalAnalysis/BMT1.txt",
header=T)
fit=survfit(Surv(T2,I2)~g,conf.type=c("plain"),bone1)
fit1=fit[g=1]
fit2=fit[g=2]
fit3=fit[g=3]
plot(0,0,main="Survival Function vs Time",xlim=c(0,600),
ylim=c(0,1),xlab="Survival Time(days)",
ylab="Survival Function",pch=" ")
lines(fit1$time,fit1$surv,type="s",lty=1)
lines(fit2$time,fit2$surv,type="s",lty=2)
lines(fit3$time,fit3$surv,type="s",lty=3)
legend(0,0.3,c("Group1=ALL","Group2=Low-AML","Group3=High-AML"),
lty=c(1,2,3),cex=0.8)
dev.off()

```

The graph of K-M survival function for three groups are displayed in the [Figure 3.5](#).

We obtain graph of  $-\ln S(t)$  versus  $t$  for the three groups (1-ALL, 2-AML low risk, 3-AML high risk) using R-programm. The R-programm commands are as follows.

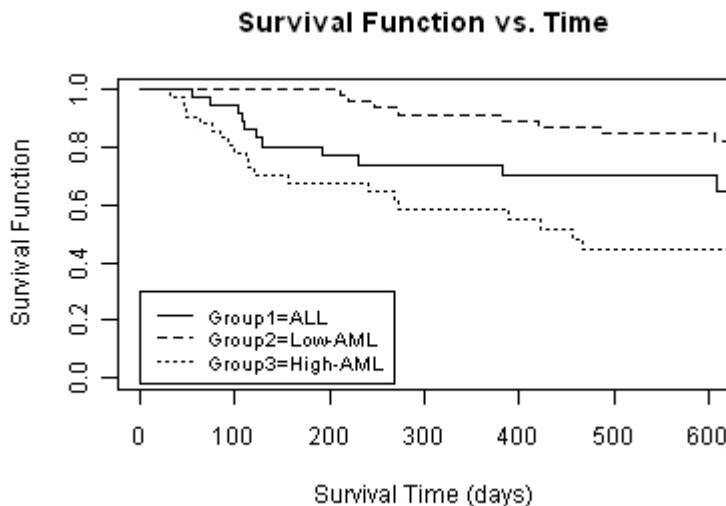


FIGURE 3.5: Graph of the survival function of three groups for BMT data set.

```
jpeg(file="surv1.jpg",height=5,width=6.5,units="in",quality=100,
bg="white",par(mfrow=c(1,1)),res=72)
bone1=read.table(file="D:/David/SurvivalAnalysis/BMT1.txt",
header=T)
fit=survfit(Surv(T2,I2)~g,conf.type=c("plain"),bone1)
fit1=fit[g=1]
fit2=fit[g=2]
fit3=fit[g=3]
plot(0,0,main="-Log(S(t)) vs Time",xlim=c(0,600),ylim=c(0,1),
xlab="Survival Time(days)",ylab="-Log(S(t))",pch=" ")
lines(fit1$time,-log(fit1$surv),type="s",lty=1)
lines(fit2$time,-log(fit2$surv),type="s",lty=2)
lines(fit3$time,-log(fit3$surv),type="s",lty=3)
legend(1,1,c("Group1=ALL","Group2=Low-AML",
"Group3=High-AML"),lty=c(1,2,3),cex=0.8)
dev.off()
```

The graph of  $-\ln S(t)$  versus  $t$  for three groups are displayed in the [Figure 3.6](#).

We obtain graph of  $\ln[-\ln S(t)]$  versus  $\ln t$  for the three groups (1-ALL, 2-AML low risk, 3-AML high risk) using R-programm. The R-programm commands are as follows.

```
jpeg(file="surv2.jpg",height=5,width=6.5,units="in",quality=100,
```

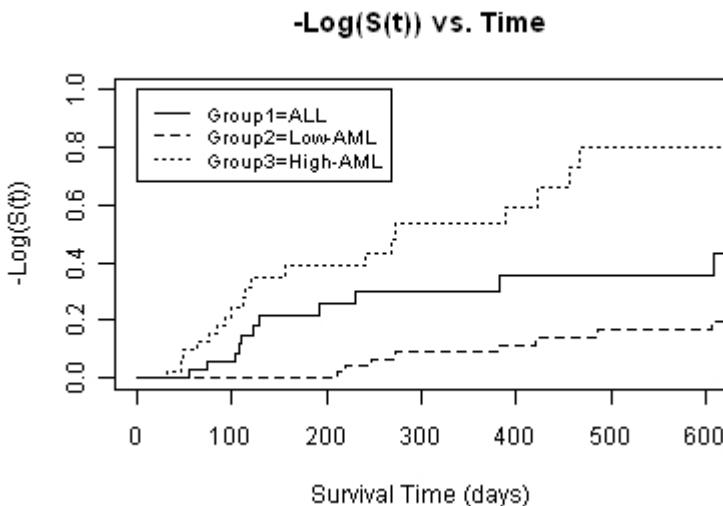


FIGURE 3.6: Graph of the  $-\ln S(t)$  versus  $t$  of three groups for BMT data set.

```

bg="white",par(mfrow=c(1,1)),res=72)
bone1=read.table(file="D:/David/SurvivalAnalysis/BMT1.txt",
header=T)
fit=survfit(Surv(T2,I2)~g,conf.type=c("plain"),bone1)
fit1=fit[g=1]
fit2=fit[g=2]
fit3=fit[g=3]
plot(0,0,main="Log(-Log(S(t))) vs log of Time",xlim=c(3.5,7),
ylim=c(-4,0.5),xlab="Log of Survival Time(days)",
ylab="Log(-Log(S(t)))",pch=" ")
lines(log(fit1$time),log(-log(fit1$surv)),type="s",lty=1)
lines(log(fit2$time),log(-log(fit2$surv)),type="s",lty=2)
lines(log(fit3$time),log(-log(fit3$surv)),type="s",lty=3)
legend(3.5,0.5,c("Group1=ALL","Group2=Low-AML","Group3=High-AML"),
lty=c(1,2,3),cex=0.8)
dev.off()

```

The graph of  $\ln[-\ln S(t)]$  versus  $\ln t$  for three groups are displayed in the Figure 3.7.

One can compare the non-parametric K-M estimate of survival curves with survival curves of parametric models, say, exponential and Weibull distributions. Figure 3.8 gives the graph of K-M survival curve and 95% confidence intervals of K-M estimate with exponential and Weibull survival curves using R-programm and SAS program together. Exponential survival curve is very close to K-M survival curve. The R-programm is as follows.

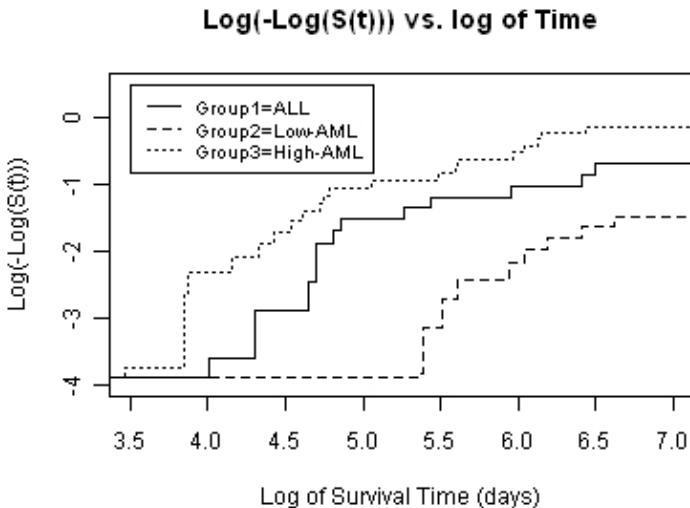


FIGURE 3.7: Graph of the  $\ln[-\ln S(t)]$  versus  $\ln t$  of three groups for BMT data set.

```

plot(0,0,xlim=c(0,600),ylim=c(0,1),xlab="Survival time(days)",
ylab="Survival probability",pch=" ")
bone=read.table(file="BMT1.txt",header=T)
fit=survfit(Surv(T2,I2),conf.type=c("plain"),bone)
lines(fit,lty=1)
x=seq(0,600, by=5)
lam=exp(-7.163) # value 7.163 is obtained from SAS program
# using proc lifereg
sf=exp(-lam*x)
lines(x,sf,lty=2)
gam=1/1.72 # value 1.72 is obtained from SAS program
# using proc lifereg
lamb=exp(-7.294/1.72)
sf1=exp(-lamb*x^gam)
lines(x,sf1,lty=3)
legend(400,1,c("KM estimate","Exponential fit","Weibull fit"),
lty=c(1,2,3),cex=0.8)

```

The SAS programm is used to calculate  $\lambda = \exp(-intercept)$  in the exponential distribution. In the same way, SAS programm is used to calculate  $\lambda = \exp(-intercept/scale)$  and  $\gamma = 1/scale$  in the Weibull distribution. The following are the SAS commands.

```

proc lifereg data=BMT1;
  model T2*I2(0)=/dist=exponential;
  run;
proc lifereg data=BMT1;
  model T2*I2(0)=/dist=weibull;
  run;

```

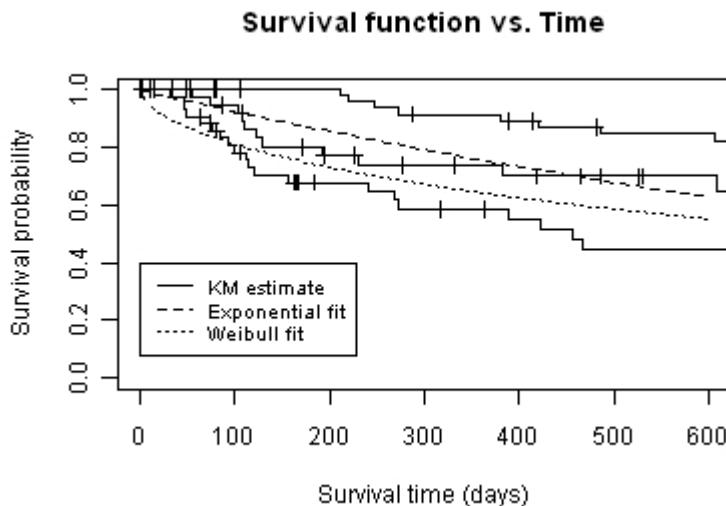


FIGURE 3.8: Graph of the  $\ln[-\ln S(t)]$  versus  $\ln t$  of three groups for BMT data set.

### 3.5 Comparison between Two Survival Functions

Let us consider the comparison between two survival functions  $S_1(t)$  and  $S_2(T)$ . The test is developed for the hypothesis

$$H_0 : S_1(t) = S_2(t)$$

that the two survival functions are the same. Treating the observations from both distributions as coming from a single population and by defining a dummy regressor variable  $y$  takes on the value 0 or 1 according whether an observation comes from first or second distribution. The hazard functions for

the two survival functions are  $h_1(t) = h_0(t)$  and  $h_2(t) = h_0(t)e^\beta$ , and the two distributions are identical if and only if  $\beta = 0$ , i.e.,

$$S_2(t) = S_1(t)^{\exp(\beta)}$$

so that by testing  $\beta = 0$ , one is testing the hypothesis

$$H_0 : S_2(t) = S_1(t) \quad \text{vs} \quad H_1 : S_2(t) = S_1(t)^\delta, \quad \delta \neq 1$$

where  $\delta = \exp(\beta)$ .

### Log Rank Test:

Let  $n_1$  and  $n_2$  be the number on individuals in the group 1 and 2, respectively and  $n = n_1 + n_2$ . Let  $n_{1i}$  and  $n_{2i}$  be the number of individuals at risk just prior to  $t_{(i)}$  from the treatments 1 and 2 and  $d_{1i}$  and  $d_{2i}$  be the number of deaths at  $t_{(i)}$  among the individuals in group 1 and group 2 and  $d_{1i} + d_{2i} = d_i$ ,  $n_{1i} + n_{2i} = n_i$ . The log rank statistic is given by

$$X_{LR} = \frac{[\sum_{i=1}^r (d_{1i} - e_{1i})]^2}{\sum_{i=1}^r V_{1i}}$$

where

$$e_{1i} = n_{1i}d_i/n_i, \quad V_{1i} = \frac{n_{1i}n_{2i}d_i(n_i - d_i)}{n_i^2(n_i - 1)}, \quad i = 1, \dots, r$$

$X_{LR}$  is the log rank statistic which is distributed approximately central chi-square distribution with one degree of freedom when the null hypothesis is true and the sample size is moderate.

### Wilcoxon Test:

The Wilcoxon test, sometimes known as the **Breslow test**, is also used to test the null hypothesis that there is no difference in the survival functions for two groups of survival data. The Wilcoxon test is based on the test statistic

$$X_W = \frac{[\sum_{i=1}^r n_i(d_{1i} - e_{1i})]^2}{\sum_{i=1}^r n_i^2 V_{1i}}$$

which has central chi-square distribution with one degree of freedom when the null hypothesis is true. The Wilcoxon test is conducted in the same manner as the log rank test.

The difference between log rank and Wilcoxon tests is that in the Wilcoxon test, each difference  $(d_{1i} - e_{1i})$  is weighted by  $n_i$ , the total number of individuals at risk at time  $t_{(i)}$ . The effect of this is to give less weight to differences between  $d_{1i}$  and  $e_{1i}$  at those times when the total number of individuals who

are still alive is small, that is, at the longest survival times. This statistic is therefore less sensitive than the log rank statistic to deviations of  $d_{1i}$  from  $e_{1i}$  in the tail of the distribution of the survival times.

The log rank test is the more suitable when the alternative to the null hypothesis of no difference between two groups of survival times is that the hazard of death at any given time for an individual in one group is proportional to the hazard at that time for a similar individual in the other group. This is the assumption of proportional hazards, which underlies a number of methods for analyzing survival data. For other types of departure from the null hypothesis, the Wilcoxon test is more appropriate than the log rank test for comparing the two survival functions. When the hazard functions are proportional, the survival functions for the two groups of survival data do not cross one another.

### Example 3.6:

The data on bone marrow transplantation is presented in Section 1.4 of Chapter 1 and also in Example 3.6 of this chapter. Patient's age, sex, and disease group (1-ALL, 2-AML low risk, 3-AML high risk) were considered as covariates. We obtain test for testing equality of survival functions for the three groups (1-ALL, 2-AML low risk, 3-AML high risk) using R-programm based on log rank and Wilcoxon tests. We also obtain estimates and 95% confidence intervals of the estimates of survival functions. The R-programm commands are as follows.

```
bone=read.table(file="BMT1.txt",header=T)
fit=survfit(Surv(T2,I2),conf.type=c("plain"),bone)
summary(fit)
survdiff(Surv(T2,I2)~g,bone,rho=0)
survdiff(Surv(T2,I2)~g,bone,rho=1)
```

The output of these R-programm commands is as follows.

```
Call: survfit(formula=Surv(T2, I2), data=bone,
  conf.type = c("plain"))
```

time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
32	133	1	0.992	0.00749	0.978		1.000	
47	131	2	0.977	0.01294	0.952		1.000	
48	129	1	0.970	0.01489	0.941		0.999	
55	126	1	0.962	0.01665	0.929		0.995	
64	124	1	0.954	0.01823	0.919		0.990	
.....								

609	62	1	0.658	0.04523	0.569	0.747
625	61	1	0.647	0.04576	0.558	0.737
662	59	1	0.636	0.04628	0.546	0.727
748	56	1	0.625	0.04683	0.533	0.717

Call:

```
survdiff(formula = Surv(T2, I2) ~ g, data = bone, rho = 0)
```

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
g=1	38	12	11.2	0.0625	0.0854
g=2	54	9	20.2	6.1851	12.0778
g=3	45	21	10.7	10.0122	13.5301

Chisq= 16.5 on 2 degrees of freedom, p= 0.000263

Call:

```
survdiff(formula = Surv(T2, I2) ~ g, data = bone, rho = 1)
```

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
g=1	38	10.08	9.36	0.0555	0.0896
g=2	54	6.72	16.47	5.7688	13.1752
g=3	45	18.04	9.02	9.0355	14.4025

Chisq= 17.8 on 2 degrees of freedom, p= 0.000138

This output shows that there is difference in observed and expected observations in the group 2 and 3, i.e., AML low risk and AML high risk. Both logrank test and Wilcoxon test are rejected for testing the hypothesis of the equality survival functions in three groups. The chi-square with 2 df is 17.8 and p-value is 0.000138. There is very high significance difference between the survival curves of three groups.

### 3.6 Cox's Proportional Hazards Model

The proportional hazards model, proposed by Cox (1972, 1975), has been used primarily in medical testing analysis, to model the effect of secondary variables on survival. Its strength lies in its ability to model and test many inferences about survival without making any specific assumptions about the form of the life distribution model.

This section will give only a brief description of the proportional hazards model. It has wide applications in medical and biology.

#### Proportional Hazards Model Assumption

Let  $y = \{y_1, y_2, \dots, y_k\}$  be a vector of one or more **explanatory** variables

believed to affect lifetime. These variables may be continuous (dosage level of a particular drug in medical studies) or they may be indicator variables with the value 1 if a given factor or condition is present, and 0 otherwise.

Let the hazard rate for a nominal (or baseline) set  $y^0 = (y_1^0, y_2^0, \dots, y_k^0)$  of these variables be given by  $h_0(t)$ , with  $h_0(t)$  denoting legitimate hazard function (failure rate) for some unspecified life distribution model.

The proportional hazards model assumes that we can write the changed hazard function for a new value of  $y$  as

$$h_y(t) = g(y)h_0(t).$$

In other words, changing  $y$ , the explanatory variable vector, results in a new hazard function that is proportional to the nominal hazard function, and the proportionality constant is a function of  $y$ ,  $g(y)$ , independent of the time variable  $t$ .

A common and useful form for  $g(y)$  is the equation:  $g(y) = e^{\beta y}$  for one variable,  $g(y_1, y_2) = e^{\beta_1 y_1 + \beta_2 y_2}$  for two variables, etc.

## Properties and Applications of the Proportional Hazards Model

1. The proportional hazards model is equivalent to the acceleration factor concept if and only if the life distribution model is a Weibull (which includes the exponential model, as a special case). For a Weibull with shape parameter  $\gamma$ , and an acceleration factor AF between nominal use fail time  $t_0$  and high stress fail time  $t_s$  (with  $t_0 = AFt_s$ ) we have  $g(y) = AF^\gamma$ . In other words,  $h_s(t) = AF^\gamma h_0(t)$ .
2. Under a log-linear model assumption for  $g(y)$ , without any further assumptions about the life distribution model, it is possible to analyze experimental data and compute maximum likelihood estimates and use likelihood ratio tests to determine which explanatory variables are highly significant.

The general proportional hazard model for the  $i$ -th individual is

$$h_i(t) = h_0(t) \exp(\beta_1 y_{1i} + \beta_2 y_{2i} + \dots + \beta_p y_{pi}).$$

This model can be reexpressed in the following form

$$\ln \left\{ \frac{h_i(t)}{h_0(t)} \right\} = \beta_1 y_{1i} + \beta_2 y_{2i} + \dots + \beta_p y_{pi}.$$

Notice that there is no constant term in the linear component of the proportional hazards model.

There are two unknown components in the proportional hazards model,

one the regression parameter and the second the baseline hazard function  $h_0(t)$ . One can write the baseline survival function as

$$S_0(t) = \exp\left(-\int_0^t h_0(u)du\right) = \exp[-H_0(t)]$$

where  $H_0(t)$  is the baseline cumulative hazard function. The survival function of  $T$  given  $\mathbf{y}$  is

$$S(t|\mathbf{y}) = \exp\left(-\int_0^t h_0(u|\mathbf{y})du\right) = [S_0(t)]^{\exp(\beta'\mathbf{y})}.$$

We wish to estimate  $\beta$  and  $S_0(t)$ , or  $h_0(t)$ , from data that are possibly censored. One approach would be to attempt to maximize the likelihood function for the observed data simultaneously with respect to  $\beta$  and  $h_0(t)$ . A more attractive approach is that given by Cox (1972), in which a likelihood that does not depend upon  $h_0(t)$  is obtained for  $\beta$ . This can be maximized to give an estimate  $\hat{\beta}$  and to provide tests for  $\beta$  in the absence of knowledge of  $h_0(t)$ . Once  $\beta$  has been estimated,  $h_0(t)$  (or  $S_0(t)$ ) can be estimated by K-M product limit estimate. This approach is taken here.

Suppose that data are available for  $n$  individuals, among whom there are  $r$  distinct death times and  $n - r$  right-censored survival times. We will for the moment assume that only one individual dies at each death time, so that there are no ties in the data. The  $r$  ordered death times will be denoted by  $t_{(1)} < t_{(2)} < \dots < t_{(r)}$ . The  $\beta$  coefficients in the proportional hazards model can be estimated using the *method of maximum likelihood*. To operate this method, we first obtain the likelihood of the sample data.

Cox (1972) showed that the partial likelihood function for the proportional hazards model and is given by

$$L(\beta) = \prod_{j=1}^r \frac{\exp(\beta'y_{(j)})}{\sum_{l \in R(t_{(j)})} \exp(\beta'y_l)},$$

where  $y_{(j)}$  is the vector of covariates for the individual who dies at the  $j$ -th ordered death time,  $t_{(j)}$ . The summation in the denominator of this likelihood function is the sum of the  $\exp(\beta'y)$  over all individuals who are at risk at time  $t_{(j)}$ . Notice that the product is taken over the individuals for whom death times have been recorded. Individuals for whom the survival times are censored do not contribute to the numerator of the log-likelihood function, but they do enter into the simulation over the risk sets at death that occur before a censored time. Moreover, the likelihood function depends only on the ranking of the death times, since this determines the risk set at each death time. When there are more than one deaths at each time, Peto (1972) and Breslow (1974) have obtained the partial likelihood function  $L(\beta)$  which is given by

$$L(\beta) = \prod_{j=1}^r \frac{\exp(\beta'S_j)}{\left(\sum_{l \in R(t_{(j)})} \exp(\beta'y_l)\right)^{d_j}},$$

where  $d_j$  is the number of lifetimes equal to  $t_{(j)}$  and  $S_j$  is the sum of the regression vectors  $\mathbf{y}$  for these  $d_j$  individuals. That is, if  $D_j$  represents the set of individuals who die at  $t_{(j)}$ , then  $d_j = |D_j|$  and  $S_j = \sum_{l \in D_j} \mathbf{y}_l$ . When there are no ties, all  $d_j = 1$ . The log-likelihood is

$$\ln L(\beta) = \sum_{j=1}^r S_j \beta - \sum_{j=1}^r d_j \ln \left( \sum_{l \in R(t_{(j)})} e^{\beta' \mathbf{y}_j} \right)$$

and the first derivatives of  $\log L(\beta)$  are

$$\frac{\partial \ln L}{\partial \beta_m} = \sum_{j=1}^r \left( S_{jm} - \frac{d_j \sum_{l \in R(t_{(j)})} y_{lm} e^{\beta' \mathbf{y}_j}}{\sum_{l \in R(t_{(j)})} e^{\beta' \mathbf{y}_j}} \right), \quad m = 1, \dots, p$$

where  $S_{jm}$  is the  $r$ -th component of  $S_j = (S_{j1}, \dots, S_{jp})$ . The matrix  $I$  containing minus the second order partial derivatives of  $\log L(\beta)$  has entries

$$\begin{aligned} I_{mn} &= \frac{-\partial^2 \ln L(\beta)}{\partial \beta_m \partial \beta_n}, \quad m, n = 1, \dots, p. \\ &= \sum_{j=1}^r d_j \left[ \sum_{l \in R(t_{(j)})} y_{lm} y_{ln} e^{\beta' \mathbf{y}_j} / \sum_{l \in R(t_{(j)})} e^{\beta' \mathbf{y}_j} \right. \\ &\quad \left. - \left( \sum_{l \in R(t_{(j)})} y_{lm} e^{\beta' \mathbf{y}_j} \right) \left( \sum_{l \in R(t_{(j)})} y_{ln} e^{\beta' \mathbf{y}_j} \right) / \left( \sum_{l \in R(t_{(j)})} e^{\beta' \mathbf{y}_j} \right)^2 \right]. \end{aligned} \quad (3.3)$$

The inverse of matrix  $I$  is the variance-covariance matrix and it is used in the Newton-Raphson procedure to obtain the maximum likelihood estimates of  $\beta$ .

### Example 3.7:

The Myeloma data discussed in Section 1.9 of Chapter 1 is revisited here in this example. We take Cox's proportional hazards model for assessing the nine covariates on the survival time using R-programm commands.

```
> myeloma=read.table(file="myeloma.txt",header=T)
> library(survival)
> fit <- coxph(Surv(T,S)~LBUN+HGB+P+Age+LW+F+LPBM+Pr+SC,myeloma)
> plot( survfit( fit ),xlab="Time (in months)",
+ ylab="Survival Function")
> summary(fit)
```

The output of the R-programm is as follows.

```

Call:
coxph(formula = Surv(T, S) ~ LBUN + HGB + P + Age + LW + F +
LPBM + Pr + SC, data = myeloma)

n= 65

      coef exp(coef) se(coef)     z      p
LBUN   1.8556   6.395   0.6563  2.827 0.0047
HGB   -0.1263   0.881   0.0721 -1.751 0.0800
P    -0.2549   0.775   0.5119 -0.498 0.6200
Age  -0.0131   0.987   0.0196 -0.668 0.5000
LW    0.3539   1.425   0.7158  0.494 0.6200
F     0.3423   1.408   0.4072  0.841 0.4000
LPBM  0.3816   1.465   0.4874  0.783 0.4300
Pr    0.0130   1.013   0.0261  0.498 0.6200
SC    0.1298   1.139   0.1050  1.236 0.2200

      exp(coef) exp(-coef) lower .95 upper .95
LBUN      6.395      0.156     1.767    23.15
HGB       0.881      1.135     0.765     1.02
P        0.775      1.290     0.284     2.11
Age      0.987      1.013     0.950     1.03
LW       1.425      0.702     0.350     5.79
F        1.408      0.710     0.634     3.13
LPBM     1.465      0.683     0.563     3.81
Pr       1.013      0.987     0.963     1.07
SC       1.139      0.878     0.927     1.40

Rsquare= 0.237  (max possible= 0.991 )
Likelihood ratio test= 17.6 on 9 df,  p=0.0399
Wald test            = 17.9 on 9 df,  p=0.0361
Score (logrank) test = 19.0 on 9 df,  p=0.0255

```

Here also LBUN is the most effective variable which is related to the survival of patients as we have seen in the parametric regression model in Chapter 2. The global test for testing all the regression parameters equal to zero is rejected at 5% level of significance. 95% confidence interval for  $\exp(-\beta_i)$ ,  $i = 1, \dots, 9$  are displayed in the output of R-programm.

[Figure 3.9](#) shows the graph of survival function versus time (in days) for the Myeloma data along with 95% confidence interval.

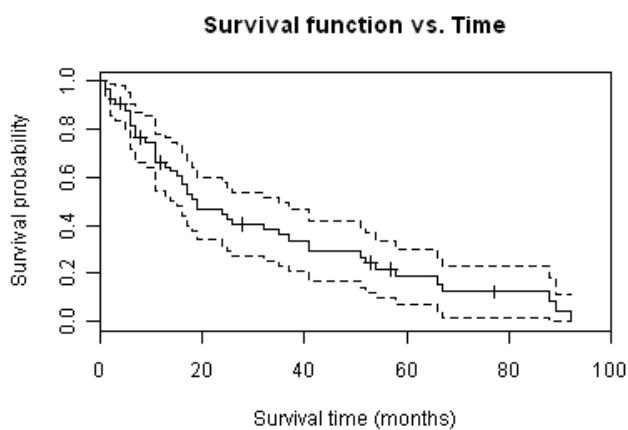


FIGURE 3.9: K-M curve with 95% CI for myeloma data.



# Part II

## Univariate and Shared Frailty Models for Survival Data



# **Chapter 4**

---

## ***The Frailty Concept***

---

4.1	Introduction .....	71
4.2	The Definition of Shared Frailty .....	73
4.3	The Implications of Frailty .....	74
4.4	The Conditional Parametrization .....	76
4.5	The Marginal Parametrization .....	77
4.6	Frailty as a Model for Omitted Covariates .....	78
4.7	Frailty as a Model of Stochastic Hazard .....	78
4.8	Identifiability of Frailty Models .....	79

---

### **4.1 Introduction**

An individual said to be frail if he or she is much more susceptible (exposed or infected) to adverse events than others. Vaupel et al. (1979) introduced the term frailty to indicate that different individuals are at risks even though on the surface they may appear to be quite similar with respect to measurable attributes such as age, gender, weight, etc. They used the term frailty to represent an unobservable random effect shared by subjects with similar (unmeasured) risks in the analysis of mortality rates. A random effect describes excess risk or frailty for distinct categories, such as individuals or families, over and above any measured covariates. Thus random effect, or frailty models, have been introduced into the statistical literature in an attempt to account for the existence of unmeasured attributes such as genotype that do introduce heterogeneity into a study population. It is recognized that individuals in the same family are more similar than the individuals in different families because they share similar genes and similar environment. Thus frailty or random effect models try to account for correlations within groups. Such groups arise naturally in those studies involving two or more failure times on the same individual subject. These failure times may be times to recurrence of the same event or times to occurrence of different types of events. Examples include the time sequences of asthmatic attacks, infection episodes, tumor diagnosis, tumor recurrences, or bleeding incidents in individual subjects (Prentice et al., 1981).

Frailties are useful in modeling correlations in multivariate survival and event history data. Examples include recurrent events such as epileptic seizures or depressive episodes, where an individual's frailty influences the occurrence of events, and community trials, where the different events within each com-

munity share a common frailty (or shared frailty), shared by each individual within the community, each unit belongs to precisely one category, and frailties of different categories are independent. More complex models are possible. Frailties can be nested; individuals within a family may share a common frailty, while families within communities share another common frailty. Frailties can be correlated, as in studies of pedigrees.

A common approach to the analysis of survival data is to assume a homogeneous population of individuals with the same covariate structure. However, it is clear that individuals identical in many respects such as age, sex, and treatment may differ in unmeasured ways, if only because of genotypical differences.

It is easy to see that it is important to consider the effect of ignoring frailty in any study where the existence of such heterogeneity may be present. Consider the study of survival times for a population consisting of one high-risk subgroup and one low-risk subgroup. Early in the study, one would estimate the population hazard to be high as the high-risk individuals start to die off rather quickly; after a period of time the remaining population consists primarily of low-risk individuals and this would give rise to a low hazard rate. Thus the population hazard function would appear to decrease over time, even if the individuals in the same subpopulation had the same hazard. More formally, if the high and low risk subpopulation were of the same size and had exponentially distributed failure times with hazards of 3 and 1 respectively, the population hazard at time  $t$  would be

$$\begin{aligned} h(t) &= \frac{f(t)}{S(t)} \\ &= \frac{3 \exp(-3t) + \exp(-t)}{\exp(-3t) + \exp(-t)} \\ &= 1 + \frac{1}{1 + \exp(2t)} \end{aligned}$$

with hazard decreasing from 3 to 1 over time.

Hougaard (1995) and Aalen (1988) provide illuminating discussions of more complex scenarios. In one example where the relative risk of failure is high in one group, the population relative risk could even drop below over time due to selection pressures.

Keyfitz and Littman (1979) developed a procedure to estimate life expectancy in a heterogeneous population and showed that ignoring heterogeneity results in overestimates of life expectancy. When Lancaster (1990) modeled the unemployment rate, he found an underestimation of the covariate effects when frailty was ignored in the model. In their view, Pickles and Crouchley (1995) suggest that ignoring such heterogeneity generally attenuates parameter estimates towards zero.

If one's experimental material involves grouped or clustered data at either

the individual or population level, one should consider the use of frailty (random effect) models. Examples would include situations where students are nested within classroom, animals are nested within litter, individuals related by family membership or shared exposure to unmeasured environmental exposures, multiple measurements on the same individual, etc. Sometimes frailty effects are modeled in a hope to account for measurement error.

More formally, a heterogeneous population can be sometimes be modeled as a mixture problem with an underlying random variable called frailty. For example, suppose  $T$  is the failure time of a subject (for example, the time to infection for a kidney patient using portable dialysis), then the probability density of  $T$  might be modeled conditional on  $Z$ , an unobserved non-measurable random variable, called frailty, which is intended to allow for individual variation. This representation can be symbolized by  $f(t; y, z)$ . The additional variable  $Y$  represents measurable covariates which are thought to be related to the failure time. Under this representation, the failure-time distribution can be considered to be continuous mixture induced by the frailty  $Z$ .

Recent research has addressed the problem of heterogeneity. Hougaard (1986) suggested the power variance function (PVF) distribution which includes gamma, inverse gaussian, positive stable distributions as frailty model. Hedeker et al. (1996) discussed a frailty regression model for the analysis of correlated grouped time survival data. Frailty models have been applied to the analysis of event history data, including the study of age at time of death for individuals in terms of population (Zelterman, 1992), unemployment duration (McCall, 1994), pregnancy in women (Aalen, 1987) and migration (Lindstrom, 1996).

## 4.2 The Definition of Shared Frailty

In a frailty model, it is assumed that the hazard function can be separated into multiplicative components: frailty, the baseline hazard function, and the linear predictor. For data classified by genotype group, the hazard function for individual  $j$  in group  $i$  would be  $z_i h_0(t) \exp(y'_{ij}\beta)$ , where  $z_i$  is the common frailty of each individual in group  $i$  which is shared by all the individuals in group  $i$  and we call it as shared frailty,  $h_0(t)$  gives the baseline hazard function and  $\exp(y'_{ij}\beta)$  is the exponential of the linear predictor. The variability of  $z_i$  determines the degree of heterogeneity among the groups and its distribution is described by the probability density function  $g(z)$ , where  $g(z)$  is interpreted as the distribution of genotypic frailty in the population. In empirical applications, the observed survival data are used to estimate the parameters of the distribution of frailty  $g(z)$  and to actually predict the individual frailties.

Frailty is a measure of relative risk because the greater an individual's frailty, with regard to some cause of death, the greater the individual's sus-

ceptibility to the cause of death. For example, if an individual with frailty 1 is called a ‘standard’ individual, then an individual with frailty 2 is having double hazard as compared to standard individual for any given time  $t$  and covariates. This definition of frailty assumes that each individual is born at a certain level of relative frailty and stays at this level all its life (Vaupel et al., 1979). Individuals in the same cluster usually share the same unobserved frailty (Shih and Louis, 1995).

Extensive research has been devoted to the frailty issue in survival analysis and generalized linear model (GLIM). Recently, investigators have recognized that ignoring individual heterogeneity may lead to inaccurate conclusions. Models for heterogeneity have been proposed by Vaupel et al. (1979), who introduced frailty as an unobserved quantity in population mortality. Oakes (1989) proposed frailty models for bivariate survival times and introduced several possible frailty models. Flinn and Heckman (1982) also introduced heterogeneity into their model for analyzing individual event histories. They believed that improper modeling of heterogeneity will result in biased estimates since the covariates in the model fail to explain the true effect of the covariates on a response variable. Keyfitz and Littman (1979) showed that ignoring heterogeneity will lead to an incorrect calculation of the life expectancy from known death rates. A similar conclusion was reached by Vaupel et al. (1979) using a continuous mixture model in which an unobserved non-negative random frailty represents all individual differences in endowment for longevity.

For reasons of convenience, analysts often choose parametric representations of frailty models that are mathematically tractable. Hougaard (1986) used several distributions for frailty including gamma, inverse gaussian, positive stable distributions and claimed that these two distributions are relevant and mathematically tractable as a frailty distribution for heterogeneous populations. Flinn and Heckman (1982) used a lognormal distribution for frailty, whereas Vaupel et al. (1979) assumed that frailty is distributed across individuals as a gamma distribution. Frailty models have been used by demographers, economists, epidemiologists, and biostatisticians to denote proneness to disease, accidents, and other events because there are persistent differences in susceptibility among individuals.

---

### 4.3 The Implications of Frailty

In hazard rate analysis, it is assumed that any differences in failure rates among individuals are picked up by the covariate structure which typically is assumed to act multiplicatively on a baseline hazard. This assumption can be relaxed by allowing for different baseline hazards in different strata and by allowing for time-dependent covariates. Even this structure assumes that individuals with the same covariate strata value has the same hazard rate.

It is useful and important to examine this assumption that all heterogeneity has been picked up by the statistical model (Trussell and Richards, 1985). The issue of ignoring heterogeneity a result of omitting important (possibly unmeasurable) variables in the model, was investigated in a real-world application by Flinn and Heckman (1982). When they incorporated frailty into the analysis of event history data, they found that there was no effect of the covariates on the response variable (duration). On the other hand, when frailty was omitted from the model, some covariates had an effect. Thus ignoring such heterogeneity may have a drastic effect on the fitted statistical model with obvious consequences in its application in real world settings. Another example of the consequences of ignoring frailty is an underestimation of the covariate effects on the risk of bladder cancer (Babiker and Cuzick, 1994).

The shared gamma frailty models was suggested by Clayton (1978) for the analysis of the correlation between clustered survival times in genetic epidemiology. An advantage is that without covariates its mathematical properties are convenient for estimation. (Oakes, 1982, 1986). However, when adjusting for environment risk factors the analysis of the clustering is more difficult (Parner, 1998). Until recently, a lack of theory and reliable software had prevented widespread use of the model.

In a frailty model, it is absolutely necessary to be able to include explanatory variables. The reason is that the frailty describes the influence of common unknown factors. If some common covariates are included in the model, the variation owing to unknown covariates should be reduced. Common covariates are common for all members of the group.

For monozygotic twins, examples are sex and any other genetically based covariate. Both monozygotic and dizygotic twins share date of birth and common pre-birth environment. By measuring some potentially important covariates, we can examine the influence of the covariates, and we can examine, whether they explain the dependence, that is, whether the frailty has no effect (or more correctly, no variation), when the covariate is included in the model.

It is not possible in practice to include all relevant covariates. For example, we might know that some given factor is important, but if we do not know the value of the factor for each individual, we cannot include the variable in the analysis. For example, it is known that excretion of small amounts of albumin in the urine is a diagnostic marker for increased mortality, not only for diabetic patients, but also for general population. However we are unable to include this variable, unless we actually obtain urine and analyze samples for each individual under study. It is furthermore possible that we are not aware that there exist variables that we ought to include. For example, this could be a genetic factor, as we do not know all possible genes having influence on survival. This consideration is true for all regression models, not only survival models. If it is known that some factor is important, it makes sense to try to obtain the individual values, but if it is not possible, the standard is to ignore the presence of such variables. In general terms, we let the heterogeneity go

into the error term. This will, of course, lead to an increase in the variability of the response compared to the case, when the variables are included. In the survival data case, however, the increased variability implies a change in the form of the hazard function, as will be illustrated by some more detailed calculations.

In statistics one is frequently interested in extracting information about unobserved quantities from observed quantities. The relationship between the two is given by a statistical model. A typical multilevel model buildup consists of a specification of a *measurement model* (i.e. the distribution of the observed quantities given the unobserved) and some description of the unobserved quantities, sometimes called *random effects*. In the context of survival studies of related individuals examples of such unobserved quantities may be genotypes, susceptibility to some disease or condition (i.e. the age-specific relative risk of death), genetic transmission pattern in a family tree, whereas the outcome it typically a survival time.

A class of random effect models which proved particularly useful in survival studies of related individuals is a class of *frailty* models which are based on the proportional hazards property.

---

## 4.4 The Conditional Parametrization

Let  $T$  be a survival time with an absolutely continuous distribution. A non-negative random variable  $Z$  is called *frailty* (Vaupel et al., 1979) if the conditional hazard function given  $Z$  has the form:

$$h(t|Z) = Zh_0(t) \quad (4.1)$$

where  $h_0(t)$  is called *baseline hazard* function. The conditional survival function is in this case given by

$$S(t|Z) = e^{-ZH(t)} \quad (4.2)$$

where  $H(t) = \int_0^t h_0(u)du$  is the cumulative baseline hazard. The marginal survival function  $S(t)$  may be obtained by taking the expectation

$$S(t) = E[S(t|Z)] = E[e^{-ZH(t)}]. \quad (4.3)$$

As it was pointed out by Hougaard (1984), the expectation (4.3) may be written as

$$S(t) = L(H(t)) \quad (4.4)$$

where  $L$  is the Laplace transform of the frailty distribution. In a similar way, conditionally on  $Z$ , the bivariate survival function is

$$S(t_1, t_2|Z) = \exp[-Z\{H_1(t_1) + H_2(t_2)\}], \quad (4.5)$$

where  $H_i(t_i) = \int_0^{t_i} h_i(u)du$ ,  $i = 1, 2$  are the integrated hazards of  $T_i$ . Here conditionally  $T_1$  and  $T_2$  are independent. We can derive unconditional bivariate survival function by integrating  $Z$  out

$$S(t_1, t_2) = E[\exp[-Z\{H_1(t_1) + H_2(t_2)\}]] = L[H_1(t_1) + H_2(t_2)]. \quad (4.6)$$

When  $T_1$  and  $T_2$  are dependent, conditionally on  $Z$ , the bivariate survival function is

$$S(t_1, t_2|Z) = \exp[-ZH(t_1, t_2)], \quad (4.7)$$

where  $H(t_1, t_2)$  is the bivariate integrated hazard of  $(T_1, T_2)$ . The corresponding unconditional bivariate survival function is derived by integrating  $Z$  out

$$S(t_1, t_2) = L[H(t_1, t_2)]. \quad (4.8)$$

For the multivariate set up, the unconditional multivariate survival function is

$$S(t_1, \dots, t_k) = L[H(t_1, \dots, t_k)]. \quad (4.9)$$

When  $T_1, \dots, T_k$  are independent the unconditional multivariate survival function is

$$S(t_1, \dots, t_k) = L\left[\sum_{i=1}^k H_i(t_i)\right]. \quad (4.10)$$

## 4.5 The Marginal Parametrization

From the relation  $S_i(t) = L(H_i(t))$ , it is found that  $H_i(t) = L^{-1}(S_i(t))$ . The bivariate survival function corresponding to the equation (4.6) is

$$S(t_1, t_2) = L(L^{-1}(S_1(t_1)) + L^{-1}(S_2(t_2))). \quad (4.11)$$

The multivariate version of this formula is

$$S(t_1, \dots, t_k) = L\left[\sum_{i=1}^k L^{-1}(S_i(t_i))\right]. \quad (4.12)$$

In terms of marginal integrated hazard functions, the multivariate version of this formula is

$$S(t_1, \dots, t_k) = L\left[\sum_{i=1}^k L^{-1}(\exp\{-H_i(t_i)\})\right]. \quad (4.13)$$

## 4.6 Frailty as a Model for Omitted Covariates

One explanation for including hidden heterogeneity effects on survival in the model is to account for the effects of omitted covariates (Hougaard et al., 1994, Keiding et al., 1997) in a Cox regression model (Cox, 1972). Let  $Y_0, \dots, Y_n$  be some covariates with associated regression coefficients  $\beta_0, \dots, \beta_n$ . In a Cox regression model the conditional survival function given the covariates is given by

$$S(x|Y_0, \dots, Y_n) = \exp(-e^{\beta_0 Y_0 + \dots + \beta_n Y_n} H(x)). \quad (4.14)$$

If the covariate  $Y_0$  is not observed and  $Y_0$  is independent of  $Y_1, \dots, Y_n$ , the respective survival function is given by:

$$S(x|Y_1, \dots, Y_n) = L(e^{\beta_1 Y_1 + \dots + \beta_n Y_n} H(x)) \quad (4.15)$$

where  $L(s)$  is the Laplace transform of “frailty”  $Z = \exp(\beta_0 Y_0)$ . In this case frailty is used to describe deviations from the proportional hazards assumption. This model is identifiable when the frailty  $Z$  has finite mean (Elbers and Ridder, 1982).

---

## 4.7 Frailty as a Model of Stochastic Hazard

Another way of introducing frailty is through the modeling of biological processes within the organism. The basic idea is that genes and environment influence the lifespan through individual age-specific hazard rates. The health history of the organism is described by a random process  $(Z_t)_{t \geq 0}$  which also incorporates the individual environment and genetic expression history. At each time  $t$  the survival chances are related to deviations of the health status from *homeostasis* (Yashin et al., 2000) which are described by the following conditional age-specific hazard rate:

$$\mu(t|\{Z_s, 0 \leq s \leq t\}) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} P(t \leq T < t + \Delta t | T \geq t, \{Z_s, 0 \leq s \leq t\}) \quad (4.16)$$

where we assume that the limit on the right-hand side exists.

In general, many components of  $Z_t$  are not observed, although recent developments in DNA microarray techniques may facilitate studies of individual gene expression histories. Therefore some simplification steps are necessary in order to make the problem of studying (4.16) analytically tractable.

**Simplification 1:** The hazard rate at age  $t$  depends only on  $Z_t$  and not on the whole past history  $\{Z_s | 0 \leq s \leq t\}$ , i.e.

$$h(t|\{Z_s, 0 \leq s \leq t\}) = h(t|Z_t).$$

**Simplification 2:** Genetic and environmental factors described by  $Z_t$  do not change with age, i.e.  $Z_t = Z, t \geq 0$  and consequently:

$$h(t|Z_t) = h(t|Z).$$

**Simplification 3:** First order Taylor approximation of  $h(t|Z)$  assuming that  $Z$  is one-dimensional and  $h(t|0) = 0$ , i.e. proportional hazards model holds:

$$h(t|Z) = Zh_0(t).$$

In the latter case all inter-individual differences in susceptibility to death at age  $t$  can be described by the non-negative random variable  $Z$  which is called *frailty* (Vaupel et al., 1979).

## 4.8 Identifiability of Frailty Models

Identifiability is an important property of a statistical model, determining whether the model parameters may be recovered from the observed data (McLachlan and Basford, 1988). A successful parameter estimation procedure or proof of consistency of the parameter estimates require that a model is identifiable to begin with. The study of identifiability property is especially important when dealing with semiparametric latent variable models (e.g. frailty models), since in these cases it is easy to specify a latent structure which is too complex to be identified from the data and also because in these models the relationship between the latent quantities and observed model characteristics is often not transparent (Iachine, 2006).

Frailty models (Vaupel et al., 1979) are used in survival analysis to account for unobserved heterogeneity in individual risks to disease and death. In a univariate frailty model the observed survival time  $T$  and unobserved frailty variable  $Z$  are related by the proportional hazards assumption:

$$h(t|Z) = Zh_0(t). \quad (4.17)$$

This model is determined by the distribution of the frailty variable  $Z$  and the underlying baseline hazard  $h_0(t)$  or equivalently the cumulative baseline hazard function  $H(t) = \int_0^t h_0(u)du$  (determined up to a multiplicative constant). This model is not identifiable from data on  $T$  alone unless additional parametric assumptions are made about the cumulative baseline hazard  $H(t)$  (Heckman and Singer, 1984).

In applications, data on observed covariates  $\mathbf{Y}$  are often available together with survival information. In this case the frailty model (4.17) may be extended to include the effects of the observed covariates:

$$h(t|Z, \mathbf{Y}) = Zr(\mathbf{Y})h_0(t), \quad (4.18)$$

where  $r(\mathbf{Y})$  is an unknown risk function. In fact, when  $r(\mathbf{Y}) = \exp(\beta^T \mathbf{Y})$  (i.e. the risk function in the regression model of Cox (1972)) the frailty variable in model (4.18) may be viewed as describing the effects of unobserved covariates  $\mathbf{Y}_u$ , i.e.  $Z = \exp(\beta^T \mathbf{Y}_u)$ .

The identifiability properties of the univariate frailty model (4.18) with unspecified functional forms of the frailty distribution, baseline hazard and risk function have been studied in detail. These properties are largely determined by the existence of a finite mean of the frailty distribution. Elbers and Ridder (1982) have shown the identifiability of model (4.18) using information on  $T$  and  $\mathbf{Y}$  when  $EZ < \infty$  (alternative conditions of identifiability were considered by Heckman and Singer (1984)). However, this model is not identifiable when frailty has an infinite mean (Hougaard, 1986).

Rider (1990) extended the result of Elbers and Ridder (1982) to generalized accelerated failure time (GAFT) models and showed that the frailty model (4.18) is a GAFT model. In essence, he proved the over identifiability property of the univariate frailty models with finite mean, a fact which was also noted by Melino and Sueyoshi (1990). In the GAFT framework the frailty model (4.18) corresponds to a particular error-term distribution structure of the GAFT regression model. Ridder (1990) shows that the GAFT model is identifiable without this additional structural assumption about the noise term.

To analyze bivariate data on pairs of related survival times  $(T_1, T_2)$  (e.g. matched pairs experiments, twin or family data), bivariate frailty models were suggested. Initially, such models exploited the data of shared frailty (Holt and Prentice, 1974, Clayton, 1978, Vaupel, 1988). The bivariate model was derived by extending (4.17) to the 2-dimensional case under the assumption of conditional independence of  $T_1, T_2$  given the shared frailty  $Z$ , resulting in the following bivariate survival function:

$$S(t_1, t_2|Z) = e^{-ZH(t_1)} e^{-ZH(t_2)}. \quad (14.19)$$

The identifiability property of this model using data on  $T_1, T_2$  was shown to hold by Honoré (1993) without the assumption of finite mean of frailty.

It turns out that univariate frailty models (4.17) without observed covariates and without any additional parametrical assumptions about  $h_0(x)$  are not identifiable from univariate survival data. Moreover, even in the presence of observed covariates there exist frailty models that cannot be identified (e.g. the positive stable distribution, Hougaard (1986b)) from univariate data alone. Bivariate survival data present a unique opportunity to identify the frailty distribution and the underlying hazards. For example, all *shared frailty* models (e.g. the positive stable) are identifiable from bivariate survival data without observed covariates (Iachine and Yashin, 1998).

# **Chapter 5**

---

## ***Various Frailty Models***

5.1	Introduction .....	81
5.2	Gamma Frailty .....	81
5.3	Positive Stable Frailty .....	83
5.4	Power Variance Function Frailty .....	85
5.5	Compound Poisson Frailty .....	86
5.6	Compound Poisson Distribution with Random Scale .....	89
5.7	Frailty Models in Hierarchical Likelihood .....	92
5.8	Frailty Models in Mixture Distributions .....	94
5.8.1	Gamma Frailty in Weibull Mixture .....	95
5.8.2	Positive Stable Frailty in Weibull Mixture .....	96
5.8.3	PVF Frailty in Weibull Mixture .....	96
5.9	Piecewise Gamma Frailty Model .....	97
5.9.1	Frailty Models and Dependence Function .....	99
5.9.2	Example: Epilepsy Data .....	102

---

### **5.1 Introduction**

In this chapter, we discuss some important frailty distributions which are used more often and they have a lot of applications. The main frailty distributions discussed in this chapter are gamma, positive stable, power variance function, lognormal, Weibull, and compound Poisson distributions. Frailty models can be expressed in terms of Laplace transform. Once the Laplace transform of frailty distribution is obtained, it is easy to obtain the estimates the parameters of frailty models. Now we will discuss various frailty models here one by one as follows.

---

### **5.2 Gamma Frailty**

Gamma distributions have been used for many years to generate mixtures in exponential and Poisson models. From a computational point of view, gamma models fit very well into survival models, because it is easy to derive the formulas for any number of events. This is due to simplicity of the derivatives of the Laplace transform. This is also the reason why this distribution has

been applied in most of the applications published until now. The probability density function (pdf) of gamma distribution as

$$f(z) = \theta^\alpha z^{\alpha-1} e^{-\theta z} / \Gamma(\alpha), \quad y > 0, \theta, \alpha > 0. \quad (5.1)$$

For many calculations and in order to solve non-identifiability problem, it makes sense to take  $E(Z) = 1$  which leads to restrict the scale parameter, and the standard restriction is  $\theta = \alpha$ , when  $\alpha \rightarrow \infty$ , the distribution becomes degenerate. The pdf and Laplace transform of gamma distribution are respectively as follows.

$$f(z) = \alpha^\alpha z^{\alpha-1} e^{-\alpha z} / \Gamma(\alpha), \quad y > 0, \theta, \alpha > 0 \quad (5.2)$$

$$L(s) = \left[1 + \frac{s}{\alpha}\right]^{-\alpha}, \quad \alpha > 0. \quad (5.3)$$

Let  $T$  be a survival times and  $Z$  be the frailty variable which is distributed as gamma in (5.2). The conditional survival function in this case is given by

$$S(t|z) = \exp(-zH(t))$$

and the unconditional survival function is given by integrating out  $Z$  from the above equation

$$S_\alpha(t) = E[S(t|z)] = L[H(t)] = \left[1 + \frac{H(t)}{\alpha}\right]^{-\alpha}, \quad \alpha > 0. \quad (5.4)$$

Let  $T_1$  and  $T_2$  be the life times of two individuals or twins or paired organs, then the unconditional bivariate survival function with gamma frailty is given by

$$\begin{aligned} S_\alpha(t_1, t_2) &= L(H(t_1, t_2)) \\ &= \frac{\alpha^\alpha}{[\alpha + H(t_1, t_2)]^\alpha} \\ &= \left\{1 + \frac{H(t_1, t_2)}{\alpha}\right\}^{-\alpha}. \end{aligned} \quad (5.5)$$

where  $H(t_1, t_2)$  is the bivariate integrated hazard function. When  $T_1$  and  $T_2$  are independent,  $H(t_1, t_2) = H_1(t_1) + H_2(t_2)$ , where  $H_1(t_1)$  and  $H_2(t_2)$  are the integrated hazard functions of the two lifetimes  $T_1$  and  $T_2$ . The various families of distributions have some theoretical advantages. If the frailty distribution is a natural exponential family, selection, that is truncation (updating when no events have happened), implies that the conditional distribution of the frailty is still within the same family. The hazard for death of individual 1 at time  $t_1$ , conditional on individual 2's being alive at time  $t_2$ , that is, conditional on the event  $(T_2 > t_2)$ , is  $h_1(t_1)\alpha[\alpha + H_1(t_1) + H_2(t_2)]^{-1}$ . The hazard for death of individual 1 at time  $t_1$ , conditional on death of individual 2 at time

$(T_2 = t_2)$ , the hazard is  $h_1(t_1)(\alpha + 1)[\alpha + H_1(t_1) + H_2(t_2)]^{-1}$ . This implies that the cross-ratio function is given by (See Clayton, 1978; Oakes, 1989)

$$\begin{aligned}\theta(t_1, t_2) &= \frac{\lambda(t_1 | T_2 = t_2)}{\lambda(t_1 | T_2 > t_2)} \\ &= \frac{\left(\frac{\partial^2 S_\alpha(t_1, t_2)}{\partial t_1 \partial t_2}\right)(S_\alpha(t_1, t_2))}{\left(\frac{\partial S_\alpha(t_1, t_2)}{\partial t_1}\right)\left(\frac{\partial S_\alpha(t_1, t_2)}{\partial t_2}\right)} \\ &= 1 + \alpha^{-1}.\end{aligned}\quad (5.6)$$

When  $T_1$  and  $T_2$  are independent,  $\theta(t_1, t_2)$  is independent of the lifetimes. The expression  $\theta(t_1, t_2)$  can be interpreted as the relative risk for an individual if the other one has experienced the event rather than being event free at a given time (See Liang, 1991). Therefore, it is an association function such that  $\theta(t_1, t_2) > 1$  represents positive association,  $\theta(t_1, t_2) < 1$  indicates negative association and  $\theta(t_1, t_2) = 1$  implies no association.

---

### 5.3 Positive Stable Frailty

In practice, the gamma frailty specification may not fit well (Shih, 1998; Glidden, 1999; Fan et al., 2000). The positive stable (PS) model (Hougaard, 2000) is a useful alternative, in part because it has the attractive feature that predictive hazard ratio decrease to 1 over time (Oakes, 1989). The property is observed in familial associations of the ages of onset of diseases with etiologic heterogeneity, where genetic cases occur early and long-term survivors are weakly correlated. The gamma model has predictive hazard ratios which are time invariant and may not be suitable for these patterns of failures (Fine et al., 2003). The probability density function (pdf) of positive stable distribution with two parameters  $\alpha$  and  $\delta$  is given by

$$f(z) = -\frac{1}{\pi z} \sum_{k=1}^{\infty} \frac{\Gamma(k\alpha + 1)}{k!} (-z^{-\alpha} \delta / \alpha)^k \sin(ak\pi), \quad z > 0, 0 < \alpha < 1, \delta > 0 \quad (5.7)$$

with Laplace transform

$$L(s) = E\{e^{-sZ}\} = e^{-\delta s^\alpha / \alpha}. \quad (5.8)$$

(See Hougaard, 2000, p. 503).

The unconditional survival function of the lifetime  $T$  with PS frailty is given by

$$S_\alpha(t) = E[S(t|z)] = L[H(t)] = \exp(\delta H(t)^\alpha / \alpha). \quad (5.9)$$

The unconditional bivariate survival function of the lifetimes  $T_1$  and  $T_2$  with PS frailty is given by

$$S_{\alpha,\delta}(t_1, t_2) = e^{-\delta[H(t_1, t_2)]^\alpha/\alpha}. \quad (5.10)$$

In order to solve the non-identifiability problem, we restrict the parameters ( $\alpha = \delta$ ) in the positive stable distribution. The pdf of positive stable distribution with ( $\alpha = \delta$ ) is given by

$$f(z) = \frac{1}{\pi} \sum_{k=1}^{\infty} \frac{\Gamma(k\alpha + 1)}{k!} \left(-\frac{1}{z}\right)^{\alpha k + 1} \sin(\alpha k \pi), \quad z > 0, 0 < \alpha < 1, \quad (5.11)$$

with Laplace transform

$$L(s) = E\{e^{-sZ}\} = e^{-s^\alpha} \quad (5.12)$$

and the unconditional survival function with PS frailty is given by

$$S_\alpha(t) = L[H(t)] = \exp(-H(t)^\alpha). \quad (5.13)$$

The unconditional bivariate survival function with positive stable frailty is given by

$$S_\alpha(t_1, t_2) = e^{-[H(t_1, t_2)]^\alpha}. \quad (5.14)$$

When  $\alpha = 1$ , the frailty distribution is degenerate at  $Z = 1$ .

### The Stable Weibull Model:

The Weibull distribution is particularly well suited to the positive stable frailty model. The bivariate Weibull model is obtained by assuming  $H_i(t) = \lambda_i t^{c_i}$ ,  $i = 1, 2$ . This means that conditionally on  $Z = z$ , the distribution of  $T_i$  is Weibull  $(\lambda_i, z, c_i)$ ,  $i = 1, 2$ . When  $T_1$  and  $T_2$  are independent, the bivariate survival function is

$$S_\alpha(t_1, t_2) = e^{-[\lambda_1 t_1^{c_1} + \lambda_2 t_2^{c_2}]^\alpha}. \quad (5.15)$$

The advantage of this model is that the marginal distributions are also of Weibull form. Also, the time to the first event,  $T_{(1)} = \min(T_1, T_2)$  is of Weibull form when  $c_1 = c_2$ .

The positive stable model has the advantage that it fits proportional hazards which means that if the conditional model has proportional hazards, so does the marginal distribution. This is an advantage, when considering the model as a random effects model.

## 5.4 Power Variance Function Frailty

Power variance function (PVF) distribution is a three-parameter family uniting gamma and positive stable distributions. The distribution is denoted  $\text{PVF}(\alpha, \delta, \theta)$ . For  $\alpha = 0$ , the gamma distributions are obtained, with same parametrization. Some formulas are valid, but many are others are different in this case. For  $\theta = 0$ , the positive stable distributions are obtained. For  $\alpha = 1/2$ , the inverse Gaussian distributions are obtained. For  $\alpha = -1$ , the non-central gamma distribution of shape parameter zero is obtained. For  $\alpha = 1$ , a degenerate distribution is obtained.

The parameter set is  $(\alpha \leq 1, \delta > 0)$ , with  $(\theta \geq 0 \text{ for } \alpha > 0)$ , and  $(\theta > 0 \text{ for } \alpha \leq 0)$ . The distribution is concentrated on the positive numbers for  $\alpha \geq 0$ , and is positive or zero for  $\alpha < 0$ . In the case  $\alpha > 0$ , the p.d.f. of PVF is given by (See Hougaard, 2000, p. 504)

$$f(z) = e^{-\theta z + \delta \theta^\alpha / \alpha} \frac{1}{\pi} \sum_{k=1}^{\infty} \frac{\Gamma(k\alpha + 1)}{k!} \left(-\frac{1}{z}\right)^{\alpha k + 1} \sin(\alpha k \pi), \quad z > 0.$$

In the case  $\alpha < 0$ , the  $\Gamma$ -term in the density is not necessarily defined, and therefore we can use the alternative expression for p.d.f. of PVF as (See Hougaard, 2000, p. 504)

$$f(z) = e^{-\theta z + \delta \theta^\alpha / \alpha} \frac{1}{z} \sum_{k=1}^{\infty} \frac{(-\delta z^\alpha / \alpha)^k}{k! \Gamma(-k\alpha)}, \quad z > 0. \quad (5.16)$$

This expression is valid for all  $\alpha$  values, except 0 and 1, with the convention that when the  $\Gamma$ -function in the denominator is undefined (which happens when  $k\alpha$  is a positive integer), the whole term in the sum is zero. For  $\alpha < 0$ , there is probability  $\exp(\delta \theta^\alpha / \alpha)$  of the random variable being zero. For  $\alpha \geq 0$ , the distribution is unimodal.

If  $Z_1$  and  $Z_2$  are independent, and  $Z_i$  follows  $\text{PVF}(\alpha, \delta_i, \theta)$ ,  $i = 1, 2$  the distribution of  $Z_1 + Z_2$  is  $\text{PVF}(\alpha, \delta_1 + \delta_2, \theta)$ . So, PVF distribution is infinitely divisible. When  $\theta > 0$ , all (positive) moments exist, and the mean is  $\delta \theta^{\alpha-1}$ . The variance is  $\delta(1-\alpha)\theta^{\alpha-2}$ .

In order to solve the non-identifiability problem, we assume  $E(Z) = 1$ . Thus  $E(Z) = 1$  is achieved by setting  $\delta = \theta^{1-\alpha}$ . The pdf of PVF when  $\delta = \theta^{1-\alpha}$  is given by

$$f(z) = e^{\theta z + \theta / \alpha} \frac{1}{z} \sum_{k=1}^{\infty} \frac{[z^\alpha / (\alpha \theta^{\alpha-1})]^k}{k! \Gamma(-k\alpha)}, \quad z > 0. \quad (5.17)$$

The Laplace transform of the above PVF distribution is

$$L(s) = e^{-\theta\{(1+s/\theta)^\alpha - 1\}/\alpha}. \quad (5.18)$$

The unconditional survival function of the lifetime  $T$  with PVF frailty is given by

$$S_{\alpha,\theta}(t) = L[H(t)] = \exp(-\theta\{[1 + H(t)/\theta]^\alpha - 1\}/\alpha). \quad (5.19)$$

The unconditional bivariate survival function of the lifetimes  $T_1$  and  $T_2$  with PVF frailty is given by

$$S_{\alpha,\theta}(t_1, t_2) = e^{-\theta\{[1 + H(t_1, t_2)/\theta]^\alpha - 1\}/\alpha}. \quad (5.20)$$

## 5.5 Compound Poisson Frailty

Aalen (1988, 1992) introduced a compound Poisson distribution as a mixing distribution in survival models which is an extension of one studied by Hougaard (1986b). The compound Poisson distribution plays a prominent role in this extension, being used here as a mixing distribution. Quite often hazard rates or intensities are raising at the start, reaching a maximum and then declining. Hence the intensity has a unimodal shape with finite mode. For example, 1) death rates for cancer patients, meaning that the longer the patient lives, beyond a certain time, the more improved are his or her chances, 2) divorce rates, the maximal rate of divorce which occurs after a few years which means most marriages are going through crisis and then improving (Aaberge et al., 1989). The population intensity starts to decline simply because the high-risk individuals have already died or been divorced, and so forth.

An additional feature which is often seen is that the total integral under the intensity (hazard rate) is to be finite; that is, the distribution is defective. In practical terms this means that some individuals have zero susceptibility; they will ‘survive forever’. For instance, some patients survive their cancer, some people never marry, some marriages are not prone to be dissolved, and so on. In medicine, there are several examples of diseases primarily attacking people with a particular susceptibility, for instance, a genetic kind, other people having virtually zero susceptibility of getting the disease. Another example is fertility. Some couples are unable to conceive children, so that the distribution of times to having first child births for a population of couples will be defective. In unemployment data, one is also faced with the fact that some people may be completely unable to get a job.

The use of the compound Poisson distribution for  $Z$  is not only mathematically convenient, but might also be seen as natural in a more substantial

sense. The distribution arises as a sum of a random number of independent gamma variables, where the number of terms in the sum is Poisson distributed. This might be viewed as a kind of shock model, where the vulnerability of the subject has been shaped by a random number of shocks, each of random size. The compound Poisson variable ( $Z$ ) can be defined as follows.

$$Z = \begin{cases} X_1 + X_2 + \dots + X_N, & N > 0 \\ 0, & N = 0 \end{cases} \quad (5.21)$$

where  $N$  is Poisson distributed with mean  $\rho$ , while  $X_1, X_2, \dots$  are independent and gamma distributed with scale parameter  $\nu$  and shape parameter  $\eta$ .

The distribution of  $Z$  consists of two parts; a discrete part which corresponds to the probability of zero susceptibility, and a continuous part on the positive real line. The discrete part is

$$P(Z = 0) = \exp(-\rho), \quad (5.22)$$

which decreases as  $\rho$  increases. The distribution of the continuous part can be found by conditioning  $N$  and using the fact that the  $X$ 's are gamma distributed. It can be written as

$$f(z; \eta, \nu, \rho) = \exp[-(\rho + \nu z)] \frac{1}{z} \sum_{n=1}^{\infty} \frac{\rho^n (\nu z)^{n\eta}}{\Gamma(n\eta) n!}. \quad (5.23)$$

The parameter set for the compound Poisson distribution is  $\eta, \nu, \rho > 0$ . The expectation and variance are given by

$$E(Z) = \rho\eta/\nu, \quad \text{Var}(Z) = \rho\eta(\eta + 1)/\nu^2. \quad (5.24)$$

The Laplace transforms of the gamma and Poisson distributions are given by  $L_X(s) = [\nu/(\nu + s)]^\eta$  and  $L_N(s) = \exp(-\rho + \rho e^{-s})$ , respectively. When  $N = 0$ ,  $L_N(s) = \exp(-\rho)$ . Now Laplace transform of compound Poisson distribution is

$$\begin{aligned} L(s) &= E(e^{-sZ}) \\ &= E(e^{-s(X_1+X_2+\dots+X_N)}) \\ &= E(L_X(s)^N) \\ &= L_N(-\ln(L_X(s))) \\ &= \exp \left\{ -\rho + \rho \left( \frac{\nu}{\nu + s} \right)^\eta \right\}. \end{aligned} \quad (5.25)$$

The above Laplace transform is same as the Laplace transform of Hougaard (1986) with different parametrization as follows.

$$L(s) = \exp \left\{ \frac{\alpha}{(1-\alpha)\delta} \left[ 1 - \left( 1 + \frac{\delta\gamma}{\alpha} s \right)^{1-\alpha} \right] \right\}, \quad \alpha, \delta \geq 0, \gamma > 0. \quad (5.26)$$

The unconditional survival function of the lifetime  $T$  is given by

$$\begin{aligned} S_{\rho,\nu,\eta}(t) &= L[H(t)] \\ &= \exp \left\{ \frac{\alpha}{(1-\alpha)\delta} \left[ 1 - \left( 1 + \frac{\delta\gamma}{\alpha} H(t) \right)^{1-\alpha} \right] \right\}, \quad \alpha, \delta \geq 0, \gamma > 0. \end{aligned} \tag{5.27}$$

If  $\alpha$  or  $\delta$  equals 0, then the distribution of  $Z$  is degenerate at  $\gamma$ . These cases correspond to no heterogeneity being present. When  $\alpha$  equals one, we get

$$L(s) = \left\{ \frac{1}{1+s\delta\gamma} \right\}^{1/\delta} \tag{5.28}$$

which is Laplace transform of a gamma distribution.

The connection between Laplace transforms in expressions (5.25) and (5.26) is derived as follows. By differentiation of (5.25) one finds the first and second moments. Let us define new parameters,  $\gamma$  is the mean,  $\delta$  is the squared coefficient of variation, and  $\alpha = \eta + 1$  and equating all these parameters we get  $\gamma = E(Z) = \rho\eta/\nu$ ,  $\delta = \text{Var}(Z)/(EZ)^2 = (\eta+1)/(\eta\rho)$ . The inverse of this parametrization is

$$\rho = \frac{\alpha}{\delta(\alpha-1)}, \quad \nu = \frac{\alpha}{\delta\gamma}, \quad \eta = \alpha - 1.$$

Inserting this into the Laplace transform (5.25) brings it into the form (5.26), which will be used in this section.

In order to solve the non-identifiability problem, we take  $E(Z) = 1$  which leads to  $\gamma = 1$  in the Laplace transform of  $Z$ . When individuals in a study belong to families or groups where there may be similarities in risk, then this association can be modeled within the present framework. Note that gene-specific quantity is shared by each of the twins. The joint survival function given the frailty  $Z = z$  is

$$S(t_1, t_2 | z) = e^{-z \cdot H(t_1, t_2)}. \tag{5.29}$$

The unconditional survival function of  $(T_1, T_2)$  is given by integrating out  $Z$  from the above equation

$$\begin{aligned} S(t_1, t_2) &= Ee^{-Z \cdot H(t_1, t_2)} \\ &= L(H(t_1, t_2)) \\ &= \exp \left\{ \frac{\alpha}{(1-\alpha)\delta} \left[ 1 - \left( 1 + \frac{\delta}{\alpha} H(t_1, t_2) \right)^{1-\alpha} \right] \right\}, \quad \alpha, \delta \geq 0, \gamma > 0 \end{aligned} \tag{5.30}$$

where  $L(.)$  is the Laplace transform of the distribution of  $Z$ . Thus, the bivariate survival function is easily expressed by means of the Laplace transform of the frailty distribution, evaluated at the total integrated conditional hazard.

## 5.6 Compound Poisson Distribution with Random Scale

An extension of the compound Poisson frailty model to family data, is to apply a probability distribution to the parameter  $\rho$  which was proposed by Moger and Aalen (2005). A probability density of  $\rho$  expresses the variation between families. The individuals of a given family are characterized by having a specific value of  $\rho$ , so they will have correlated frailties, while individuals from different families are independent. This yields a two level model, where the frailty has two components: A familial component, for instance relating to shared genes and environment, and an individual component, which could relate to exposure to individual environment. Thus, the model does not fit into the traditional dichotomy of shared frailty models. We would like to stress the importance of frailty models having clear biological content, corresponding to understand a problem from a substance point of view, as opposed to just making mathematical assumptions. Since compound Poisson distribution is included in the power variance function (PVF) distributions, this corresponds to randomizing a scale parameter in the PVF distributions.

This section will focus on densities for  $\rho$  which are included in the PVF distribution family. Specifically we consider the gamma, inverse Gaussian and positive stable distributions. As given in Hougaard (2000), the distributions can be united in a three-parameter family with parameter set  $\alpha \leq 1$ ,  $\epsilon > 0$ , with  $\theta \geq 0$  for  $\alpha > 0$ , and  $\theta > 0$  for  $\alpha \leq 0$ . For  $\alpha = 0$  the gamma distributions are obtained. The inverse Gaussian distributions are obtained for  $\alpha = 1/2$ , and for  $\theta = 0$  one gets the positive stable distributions. The positive stable distributions are absolutely continuous and nonnegative, with unimodal densities (Hougaard, 1986). For  $\alpha = 1$ , a degenerate distribution is obtained, at  $\epsilon$ , independent of  $\theta$ . This corresponds to independence within families. This is given by (5.23), with  $\eta = -\alpha$ ,  $\rho = -(\epsilon/\alpha)\theta^\alpha$  and  $\nu = \theta$ . Hence, values of  $\alpha < 0$  yield the compound Poisson distributions. The parameterization used in Section 2 is the most appropriate for  $\alpha < 0$ , while the parameterization by Hougaard is more easy to use for  $\alpha > 0$ . The expectation and variance of the distribution of  $\rho$  are

$$E(\rho) = \epsilon\theta^{\alpha-1}, \quad Var(\rho) = \epsilon(1-\alpha)\theta^{\alpha-2}. \quad (5.31)$$

Thus, the positive stable distribution has no finite expectation or variance. The Laplace transform of  $\rho$  is given by

$$L_\rho(s) = \exp\left\{-\frac{\epsilon}{\alpha}[(\theta + s)^\alpha - \theta^\alpha]\right\}. \quad (5.32)$$

In the case of the mixed compound Poisson distribution, the unconditional discrete part of  $Z$  is given by

$$P(Z=0) = E(\exp(-\rho)) = L_\rho(1). \quad (5.33)$$

The density of the unconditional continuous part of  $Z$  can be calculated in a similar manner by noting that

$$E[\rho^n \exp(-\rho)] = (-1)^n L_\rho^{(n)}(1), \quad (5.34)$$

where  $L^{(n)}(s)$  denotes the  $n$ -th derivative of the Laplace transform. By inserting the density of  $\rho$  into (5.23) and integrating out  $\rho$ , the density of  $Z$  may be put on the following form:

$$h(z; \eta, \nu, \alpha, \theta, \epsilon) = \exp(-\nu z) \frac{1}{z} \sum_{n=1}^{\infty} \frac{(\nu z)^{n\eta}}{\Gamma(n\eta)n!} (-1)^n L_\rho^{(n)}(1). \quad (5.35)$$

The derivatives of the Laplace transform for the power variance function distribution for  $\rho$  are of the form

$$L_\rho^{(n)}(s) = (-1)^n L_\rho(s) \sum_{j=1}^n c_{n,j}(\alpha) \epsilon^j (\theta + s)^{j\alpha - n}, \quad (5.36)$$

as shown in Hougaard (2000). The coefficients  $c_{n,j}(\alpha)$  are given by the recursive formula

$$\begin{aligned} c_{n,1}(\alpha) &= \Gamma(n-\alpha)/\Gamma(1-\alpha), & c_{n,n}(\alpha) &= 1, \\ C_{n,j}(\alpha) &= C_{n-1,j-1}(\alpha) + c_{n-1,j}(\alpha)[(n-1)-j\alpha]. \end{aligned}$$

The Laplace transform of  $\rho$ ,  $L_\rho(s)$ , combined with (5.25) yield the expression

$$L_Z(s) = L_\rho \left( 1 - \left( \frac{\nu}{\nu+s} \right)^\eta \right) \quad (5.37)$$

for the Laplace transform of  $Z$ . For the PVF distributed  $\rho$ , this equals

$$\begin{aligned} L_Z(s) &= \exp \left( -\frac{\epsilon}{\alpha} \left\{ \left[ \theta + 1 - \left( \frac{\nu}{\nu+s} \right)^\eta \right]^\alpha - \theta^\alpha \right\} \right) \\ &\quad \text{if } \alpha \leq 1, \alpha \neq 0, \end{aligned} \quad (5.38)$$

$$L_Z(s) = \left( \frac{\theta}{\theta + 1 - \left( \frac{\nu}{\nu+s} \right)^\eta} \right)^\epsilon \quad \text{if } \alpha = 0. \quad (5.39)$$

The Laplace transform of the gamma mixture distribution ( $\alpha = 0$ ) is obtained by taking the limit of the general Laplace transform. The positive stable mixture distribution ( $\theta = 0$ ) gives some nice properties when used as a frailty distribution. The Laplace transform of  $Z$  in this case is

$$L_Z(s) = \exp \left\{ -\frac{\epsilon}{\alpha} \left[ 1 - \left( \frac{\nu}{\nu+s} \right)^\eta \right]^\alpha \right\}. \quad (5.40)$$

Apart from the exponent  $\alpha$ , this is of the same form as the Laplace transform

of a compound Poisson distribution given in Equation (5.25), with  $\epsilon/\alpha$  playing the role of  $\rho$ .

Now the survival function of the lifetime  $T$  is given by

$$\begin{aligned} S(t) &= L_Z(H(t)) \\ &= \exp\left(-\frac{\epsilon}{\alpha}\left\{\left[\theta + 1 - \left(\frac{\nu}{\nu + H(t)}\right)^{\eta}\right]^{\alpha} - \theta^{\alpha}\right\}\right) \quad \text{if } \alpha \leq 1, \alpha \neq 0, \\ S(t) &= \left(\frac{\theta}{\theta + 1 - \left(\frac{\nu}{\nu + H(t)}\right)^{\eta}}\right)^{\epsilon} \quad \text{if } \alpha = 0. \end{aligned}$$

The mean and variance of  $Z$  can easily be found by means of (5.24), by noting that  $E(Z) = E[E(Z|\rho)]$  and that  $\text{Var}(Z) = \text{Var}[E(Z|\rho)] + E[\text{Var}(Z|\rho)]$ :

$$\beta = E(Z) = \frac{\epsilon\eta}{\nu\theta^{1-\alpha}}, \quad \text{Var}(Z) = \frac{\epsilon\eta[\theta + \eta(1 - \alpha + \theta)]}{\nu^2\theta^{2-\alpha}}. \quad (5.41)$$

Note that when a positive stable mixture distribution is used, the frailty distribution  $Z$  has no finite expectation or variance. When  $\rho$  is not stable distributed ( $\theta > 0$ ), the Laplace transform in (5.38) can be reparameterized by using the expectation  $\beta$  from (5.41) and the squared coefficient of variation

$$d = \frac{\text{Var}(Z)}{E(Z)^2} = \frac{[\theta + \eta(1 - \alpha + \theta)]}{\theta^{\alpha}\epsilon\eta}$$

as new parameters. The value  $d = 0$  corresponds to no heterogeneity.

When individuals in a study belong to families or groups where there may be similarities in risk, then this association can be modeled within the present framework. For simplicity, a paired twins or paired dental implants or paired organs are dependent with hazard function  $H(t_1, t_2)$ . Note that gene-specific quantity is shared by each of the twins. The joint survival function given the frailty  $Z = z$  is

$$S(t_1, t_2 | z) = e^{-z \cdot H(t_1, t_2)}. \quad (5.42)$$

When  $T_1$  and  $T_2$  are independent,  $H(t_1, t_2) = H_1(t_1) + H_2(t_2)$ , where  $H_i(t_i)$ ,  $i = 1, 2$  are the integrated hazards of  $T_1$  and  $T_2$  respectively. From this, we immediately derive the bivariate survival function by integrating  $Z$  out

$$\begin{aligned}
S(t_1, t_2) &= Ee^{-Z \cdot H(t_1, t_2)} \\
&= L(H(t_1, t_2)) \\
&= \exp \left( -\frac{\epsilon}{\alpha} \left\{ \left[ \theta + 1 - \left( \frac{\nu}{\nu + H(t_1, t_2)} \right)^{\eta} \right]^{\alpha} - \theta^{\alpha} \right\} \right) \\
&\quad \text{if } \alpha \leq 1, \alpha \neq 0, \\
&= \left( \frac{\theta}{\theta + 1 - \left( \frac{\nu}{\nu + H(t_1, t_2)} \right)^{\eta}} \right)^{\epsilon} \quad \text{if } \alpha = 0,
\end{aligned} \tag{5.43}$$

where  $L(\cdot)$  is the Laplace transform of the distribution of  $Z$ . Thus, the bivariate survival function is easily expressed by means of the Laplace transform of the frailty distribution, evaluated at the total integrated conditional hazard.

In order to solve the non-identifiability problem, we assume a mean of 1 for the frailty distributions. For the gamma distribution, this can be achieved by setting  $\theta = \epsilon$ . In the shared PVF model,  $E(Z) = 1$  is achieved by setting  $\epsilon = \theta^{1-\alpha}$ . The shared frailty models are compared to a compound Poisson model where  $\rho$  is gamma distributed, yielding a compound Poisson-gamma model. To secure a unit mean for the frailty, we get  $\epsilon = \nu\theta/\eta$ . In this section, we assume the distribution of frailty as compound Poisson-gamma distribution for the bivariate survival data. The bivariate survival function based on this frailty is given by

$$S(t_1, t_2) = \left( \frac{\theta}{\theta + 1 - \left( \frac{\nu}{\nu + H(t_1, t_2)} \right)^{\eta}} \right)^{\nu\theta/\eta}. \tag{5.44}$$

## 5.7 Frailty Models in Hierarchical Likelihood

The generalized linear model (GLIM) can be extended by incorporating frailty in the linear predictor. This generalized linear mixed model (GLMM) is very useful for accommodating the over-dispersion occurring in the data. Such research has been carried out by Schall (1991), who emphasized the linear logistic regression model with frailty. Schall (1991) proposed an algorithm to estimate the fixed effects, frailty and dispersion components in a GLMM, where this algorithm yields approximate ML and restricted maximum likelihood (REML) estimators for the variance of the frailty terms. Another method adjusting for over-dispersion was developed by Breslow and Clayton (1993), using Laplace's method to obtain the marginal quasi-likelihood (MQL). The

integrated quasi-likelihood proposed by Breslow and Clayton (1993), however, cannot be evaluated in closed form. Therefore, Laplace's method was used to approximate the integral. This approach eventually led to estimating equations based on penalized quasi-likelihood (PQL) for the mean parameter and pseudo-likelihood for the variance. The pseudo ML method is used to obtain an estimator by maximizing a likelihood function associated with family of distributions which does not necessarily contain the true distribution. In most cases, the frailty term is assumed to have a Gaussian distribution with mean of zero and a variance of  $\sigma^2$ . Penalized likelihood was specifically exploited by Green and Silverman (1994) to be used for semiparametric regression analysis, whereas Hastie and Tibshirani (1986) applied the penalized likelihood procedure to generalized additive models (GAM).

McGilchrist and Aisbett (1991) and McGilchrist (1993) considered the proportional hazard model

$$h(t; y_i) = h_0(t) \exp(y_i' \beta + u_i), \quad i = 1, \dots, n \quad (5.45)$$

where  $y_i$  is the p-component covariate vector associated with the  $i$ -th individual or system,  $\beta$  is the p-component regression coefficient vector, and  $u_i$  the corresponding individual or system effect in  $n$  individuals or systems. Since the frailties  $z_i = \exp(u_i)$  are assumed to have a lognormal distribution, then  $u_i$  has a normal distribution; relocating  $u_i$  (by absorption into the hazard or intercept) leads to a zero mean. McGilchrist (1994) adapted Henderson's (1975) and Harville's (1977) best linear unbiased predictor (BLUP) procedures for mixed normal linear models for use in GLMM.

If the density function of a failure time  $T_i$ , conditional on  $u_i$  is  $g(t_i; \beta | u_i)$ , then  $l_1 = \sum \ln g(t_i; \beta | u_i)$  will be the corresponding (conditional) log likelihood function. Then  $l(\beta, \mathbf{u}) = l_1 + l_2$  becomes the joint log likelihood of  $\beta$  and  $\mathbf{u} = (u_1, u_2, \dots, u_n)'$ , where  $l_2$  is the log likelihood function of the frailty effects. The BLUP procedure avoids having to integrate out the frailty distribution; an important advantage since in most cases the required numerical integration is not practical (Schall, 1991). A similar approach to Henderson's method has been developed by Lee and Nelder (1996), who introduced a class of hierarchical generalized linear models (HGLM) which incorporated random components in the model. In their approach the distribution of frailty could come from any distribution and they paid special attention to the conjugate distribution, where the distribution of frailty is conjugate to that of the response variable. For example, if frailty is assumed to have beta distribution then the conditional distribution of the response given frailty is the binomial distribution or if frailty is assumed to have a inverse gamma distribution then the conditional distribution of the response given frailty is the gamma distribution.

The likelihood function  $L(t_1, t_2, \mathbf{u}; \beta, \sigma)$ , where  $\sigma$  is the standard deviation of  $u$ , is constructed from the product of  $L_1(t_1, t_2, \beta | \mathbf{u})$ , the partial likelihood of paired failure times ( $T_1, T_2$ ) conditional on frailty, and  $L_2(\mathbf{u}, \sigma)$ , the likelihood of the frailty. However, the frailty are not directly observable, and the

joint likelihood is not a standard likelihood in the conventional sense because it is based on the non-observable random variables  $\mathbf{u}$ . Lee and Nelder (1996) argued that by using hierarchical likelihood (h-likelihood) one avoids integrating out the frailty and also the h-likelihood retains the properties of the likelihood. Also for more complicated problems, such as when the number of parameters goes to infinity, the marginal likelihood cannot be reduced beyond a high dimensional integral, and numerical integration is no longer feasible or reliable because it is difficult to integrate out the frailty.

Henderson's mixed model equations provide efficient ways of computing the BLUP estimates of  $\beta$  and  $\sigma$ . The ML estimates of  $\beta$  is in fact the same as that arising from the BLUP procedure. Maximizing the density of the residuals with respect to  $\sigma$  produces restricted maximum likelihood (REML) estimates. This estimation procedure using the BLUP technique has been reviewed extensively by Robinson (1991) for normal linear models. His paper mentions that BLUP is an effective method to estimate the frailty, since the random variables  $\mathbf{u}$  are predictable.

Schall (1991) adopted Henderson's procedure to GLMM's. Then McGilchrist (1994) relaxed the exponential family assumption for the conditional distribution of  $T$  given  $u_i$ , but required the log frailty of  $u_i$  to be normally distributed. Lee and Nelder's (1996) HGLM allows for any positive distribution of frailty.

## 5.8 Frailty Models in Mixture Distributions

Hanagal (2007a, 2007b, 2008a) developed the frailty models in mixture distribution in general. He also obtained the estimation procedures in gamma and positive stable frailty models. The mixture distribution in terms of survival function is

$$\begin{aligned} S(t) &= pS_1(t) + (1-p)S_2(t) \\ &= pe^{-H_1(t)} + (1-p)e^{-H_2(t)}. \end{aligned} \quad (5.46)$$

There are two ways of obtaining frailty models. The first one is

$$\begin{aligned} S_M(t|z) &= pS_1(t|z) + (1-p)S_2(t|z) \\ &= pe^{-zH_1(t)} + (1-p)e^{-zH_2(t)}. \end{aligned} \quad (5.47)$$

This is called mixture of frailty models or mixture frailty. The second one is

$$S_F(t|z) = e^{z \ln[pS_1(t) + (1-p)S_2(t)]}. \quad (5.48)$$

This is called frailty of the mixture distributions in order to make the distinction between the two types. The same technique can be generalized to mixture

of more than two distributions. When we integrate with respect to frailty ( $Z$ ), we get survival function of mixture distribution in terms of frailty parameter.

### Example: Weibull Mixtures:

The survival function of two Weibull distributions is

$$\begin{aligned} S(t) &= pS_1(t) + (1-p)S_2(t) \\ &= pe^{-\lambda_1 t^{c_1}} + (1-p)e^{-\lambda_2 t^{c_2}}. \end{aligned} \quad (5.49)$$

The mixture frailty is given by

$$S_M(t|z) = pe^{-z\lambda_1 t^{c_1}} + (1-p)e^{-z\lambda_2 t^{c_2}}. \quad (5.50)$$

The frailty of the mixture is given by

$$S_F(t|z) = e^{z \ln[pe^{-\lambda_1 t^{c_1}} + (1-p)e^{-\lambda_2 t^{c_2}}]}. \quad (5.51)$$

The scale parameters  $\lambda_1$  and  $\lambda_2$  in the Weibull mixtures can be expressed in terms of regression parameters in the following way

$$\lambda_1 = \lambda_2 = \exp(\beta' z)$$

where  $\beta' = (\beta_0, \beta_1, \dots, \beta_p)$  and  $z = (1, z_1, \dots, z_p)$ .

If we want to make a distinction between the two scale parameters, one can express  $\lambda_1$  and  $\lambda_2$  as

$$\begin{aligned} \lambda_1 &= \theta_1 \exp(\beta' z) \\ \lambda_2 &= \theta_2 \exp(\beta' z). \end{aligned} \quad (5.52)$$

After substituting  $\lambda_1$  and  $\lambda_2$  in (5.50) and (5.51), we get frailty regression models in Weibull mixtures.

#### 5.8.1 Gamma Frailty in Weibull Mixture

Assuming the distribution of  $Z$  as gamma distribution and integrating over  $Z$ , in Eq. (5.50), we get Weibull mixture model with gamma mixture frailty, given by

$$S_M(t) = p \left[ 1 + \frac{\lambda_1 t^{c_1}}{\alpha} \right]^{-\alpha} + (1-p) \left[ 1 + \frac{\lambda_2 t^{c_2}}{\alpha} \right]^{-\alpha}. \quad (5.53)$$

The pdf corresponding to the above survival function is

$$f_M(t) = S_{M1}^{\frac{\alpha+1}{\alpha}} p \lambda_1 c_1 t^{c_1-1} e^{-\lambda_1 c_1 t^{c_1-1}} + S_{M2}^{\frac{\alpha+1}{\alpha}} (1-p) \lambda_2 c_2 t^{c_2-1} e^{-\lambda_2 c_2 t^{c_2-1}}, \quad (5.54)$$

where  $S_{M1} = \left[1 + \frac{e^{-\lambda_1 t^{c_1}}}{\alpha}\right]^{-\alpha}$  and  $S_{M2} = \left[1 + \frac{e^{-\lambda_2 t^{c_2}}}{\alpha}\right]^{-\alpha}$ .

Assuming the distribution of  $Z$  as gamma distribution and integrating over  $Z$ , in Eq. (5.51), we get Weibull mixture model with gamma frailty, given by

$$S_F(t) = \left[1 + \frac{\ln[pe^{-\lambda_1 t^{c_1}} + (1-p)e^{-\lambda_2 t^{c_2}}]}{\alpha}\right]^{-\alpha}. \quad (5.55)$$

The pdf corresponding to the above survival function is

$$f_M(t) = S_F^{\frac{\alpha+1}{\alpha}} \frac{p\lambda_1 c_1 t^{c_1-1} e^{-\lambda_1 c_1 t^{c_1-1}} + (1-p)\lambda_2 c_2 t^{c_2-1} e^{-\lambda_2 c_2 t^{c_2-1}}}{[pe^{-\lambda_1 t^{c_1}} + (1-p)e^{-\lambda_2 t^{c_2}}]}. \quad (5.56)$$

### 5.8.2 Positive Stable Frailty in Weibull Mixture

Assuming the distribution of  $Z$  as positive stable distribution and integrating over  $Z$ , in Eq. (5.50), we get Weibull mixture model with positive stable mixture frailty, given by

$$S_M(t) = pe^{-(\lambda_1 t^{c_1})^\alpha} + (1-p)e^{-(\lambda_2 t^{c_2})^\alpha}. \quad (5.57)$$

The pdf corresponding to the above survival function is

$$f_M(t) = pe^{-(\lambda_1 t^{c_1})^\alpha} c_1 \alpha \lambda_1^\alpha t^{c_1 \alpha - 1} + (1-p)e^{-(\lambda_2 t^{c_2})^\alpha} c_2 \alpha \lambda_2^\alpha t^{c_2 \alpha - 1} \quad (5.58)$$

which is again a mixture of two Weibull distributions. In this case, positive stable frailty model is stable with Weibull mixtures.

Assuming the distribution of  $Z$  as positive stable distribution and integrating over  $Z$ , in Eq. (5.51), we get Weibull mixture model with positive stable frailty, given by

$$S_F(t) = e^{(\ln[pe^{-\lambda_1 t^{c_1}} + (1-p)e^{-\lambda_2 t^{c_2}}])^\alpha}. \quad (5.59)$$

The pdf corresponding to the above survival function is

$$\begin{aligned} f_F(t) &= S_F(t) \alpha (\ln[pe^{-\lambda_1 t^{c_1}} + (1-p)e^{-\lambda_2 t^{c_2}}])^{\alpha-1} \\ &\times \left( \frac{p\lambda_1 c_1 t^{c_1-1} e^{-\lambda_1 t^{c_1}} + (1-p)\lambda_2 c_2 t^{c_2-1} e^{-\lambda_2 t^{c_2}}}{[pe^{-\lambda_1 t^{c_1}} + (1-p)e^{-\lambda_2 t^{c_2}}]} \right). \end{aligned} \quad (5.60)$$

### 5.8.3 PVF Frailty in Weibull Mixture

Assuming the distribution of  $Z$  as PVF distribution and integrating over  $Z$ , in Eq. (5.50), we get Weibull mixture model with PVF mixture frailty, given by

$$S_M(t) = \mu \left[ pe^{-\delta(\theta+\lambda_1 t^{c_1})^\alpha/\alpha} + (1-p)e^{-\delta(\theta+\lambda_2 t^{c_2})^\alpha/\alpha} \right] \quad (5.61)$$

where  $\mu = e^{\delta\theta^\alpha}/\alpha$ .

The pdf corresponding to the above survival function is

$$\begin{aligned} f_M(t) &= \mu p e^{-\delta(\theta + \lambda_1 t^{c_1})^\alpha/\alpha} \delta(\theta + \lambda_1 t^{c_1})^{\alpha-1} \lambda_1 c_1 t^{c_1-1} \\ &\quad + \mu(1-p) e^{-\delta(\theta + \lambda_2 t^{c_2})^\alpha/\alpha} \delta(\theta + \lambda_2 t^{c_2})^{\alpha-1} \lambda_2 c_2 t^{c_2-1}. \end{aligned} \quad (5.62)$$

Assuming the distribution of  $Z$  as PVF distribution and integrating over  $Z$ , in Eq. (5.51), we get Weibull mixture model with PVF frailty, given by

$$S_F(t) = \mu e^{\delta(\theta + \ln[pe^{-\lambda_1 t^{c_1}} + (1-p)e^{-\lambda_2 t^{c_2}}])^\alpha/\alpha}. \quad (5.63)$$

The pdf corresponding to the above survival function is

$$\begin{aligned} f_F(t) &= S_F(t) \delta(\theta + \ln[pe^{-\lambda_1 t^{c_1}} + (1-p)e^{-\lambda_2 t^{c_2}}])^{\alpha-1} \\ &\quad \times \left( \frac{p\lambda_1 c_1 t^{c_1-1} e^{-\lambda_1 t^{c_1}} + (1-p)\lambda_2 c_2 t^{c_2-1} e^{-\lambda_2 t^{c_2}}}{[pe^{-\lambda_1 t^{c_1}} + (1-p)e^{-\lambda_2 t^{c_2}}]} \right). \end{aligned} \quad (5.64)$$

## 5.9 Piecewise Gamma Frailty Model

Paik et al. (1994) worked on a generalization of a multivariate frailty model by introducing additional frailty terms for different time-intervals, Wintrebert et al. (2004) proposed a multivariate frailty model with a power parameter which allows a center-specific frailty to vary among individuals and extended this model to allow the center-specific frailty to change with time, Yau et al. (1998) allowed the frailties to be time varying according to an AR(1) process and proposed ML and REML methods for estimation, while Manda and Meyer (2005) presented a similar model, within a Bayesian framework. In genetic studies where the outcome is the time to the event of interest, failure times among family members may not be independent. In this case, conventional survival analysis may yield consistent estimates of the marginal hazard if the marginal hazard is incorrectly modeled (Huster et al., 1989). However, variance estimates overestimate the true variance when the independent variables vary within a unit, and underestimate when the independent variables are constant within a unit, leading to incorrect inferences.

Let  $S_{ij}^*(t_{ij})$  be the baseline survivor function of the  $j$ -th member in unit  $i$  at time  $T_{ij} = t_{ij}$ . In frailty models, conditionally on a random effect of a unit, say  $Z_i$ , the survivor function takes a form of the Lehmann family of alternatives, and the failure times for subunits in unit  $i$ ,  $T_{i1}, T_{i2}, \dots, T_{in}$ , are assumed to be independent. Then a joint survivor function of  $(t_{i1}, t_{i2}, \dots, t_{in})$  conditioning on  $Z_i$ , say  $S_i(t_{i1}, t_{i2}, \dots, t_{in}|Z_i)$ , is

$$S_i(t_{i1}, t_{i2}, \dots, t_{in}|Z_i) = \prod_{j=1}^n S_{ij}(t_{ij}|Z_i) = \prod_{j=1}^n S_{ij}^*(t_{ij})^{Z_i}.$$

Additionally,  $Z_i$  is assumed to be distributed as  $g(Z_i)$ . With these two assumptions, the joint survivor function is

$$\begin{aligned} S_i(t_{i1}, t_{i2}, \dots, t_{in}) &= Pr(T_{i1} > t_{i1}, T_{i2} > t_{i2}, \dots, T_{in} > t_{in}) \\ &= \int \prod_{j=1}^n S_{ij}^*(t_{ij})^{Z_i} g(Z_i) dZ_i. \end{aligned} \quad (5.65)$$

The marginal survivor function is

$$S_j(t_{ij}) = \int S_{ij}^*(t_{ij})^{Z_i} g(Z_i) dZ_i.$$

$S_i(t_{ij})$  can be interpreted as the average survivor function for a population. Frailty models are attractive especially in applications to genetic studies for two reasons. First,  $Z_i$  can be interpreted as a genetic factor as well as environmental components, especially all environmental factors that cause the correlation among failure times in a unit or family. Second, in most genetic studies, the number of subjects per family varies. In frailty models, the varying number of subjects can be handled easily due to the conditional independence structure. See Aalen (1988) and Clayton (1988) for more discussion.

Frailty models assume the proportionality of the random effects conditionally and the form of the density of random effects  $g(Z_i)$ . However, random effects are unobservable; therefore assumptions on random effects cannot be verified directly from data. In previous studies, a gamma distribution (Clayton, 1978; Oakes, 1982) or a positive stable distribution (Hougaard, 1986b) was chosen for  $g(Z_i)$  because of computational convenience. See Oakes (1989) for other choices of distributions. The assumption on  $g(Z_i)$  is equivalent to the assumption on the dependence structure among failure times (Oakes, 1989). To see what was assumed about the dependence structure by choosing  $g(Z_i)$ , define a measure of dependence between the bivariate failure times:

$$\begin{aligned} \theta^*(t_1, t_2) &= \frac{S(t_1, t_2) D_1 D_2 S(t_1, t_2)}{D_2 S(t_1, t_2) D_1 S(t_1, t_2)} \\ &= h(t_1 | T_2 = t_2) / h(t_1 | T_2 > t_2), \end{aligned} \quad (5.66)$$

where  $D_j$  denotes  $\partial / \partial t_j$ ,

$$\begin{aligned} h(t_1 | T_2 = t_2) &= \lim_{\Delta \rightarrow 0} \Delta^{-1} Pr(t_1 \leq T_1 < t_1 + \Delta | T_1 > t_1, T_2 = t_2) \\ &= \frac{D_1 D_2 S(t_1, t_2)}{D_2 S(t_1, t_2)}, \end{aligned}$$

$$h(t_1 | T_2 > t_2) = \lim_{\Delta \rightarrow 0} \Delta^{-1} Pr(t_1 \leq T_1 < t_1 + \Delta | T_2 > t_2) = \frac{D_1 S(t_1, t_2)}{S(t_1, t_2)},$$

$$S(t_1, t_2) = Pr(T_1 > t_1, T_2 > t_2).$$

This is a ratio of two hazards: the hazard of one member given the other

member has failed vs the other member has survived.  $\theta^*(t_1, t_2)$  is determined by an assumption about the distribution of  $Z_i$ . For instance, suppose that  $S_i$  arises from a gamma distribution with coefficient of variation  $1/c$ . Then straightforward calculation following (5.65) and (5.66) shows that

$$\theta^*(t_1, t_2) = (1 + c)/c. \quad (5.67)$$

Therefore, by choosing a gamma frailty, one assumes that  $\theta^*(t_1, t_2)$  is constant over time. A positive stable frailty yields  $\theta^*(t_1, t_2)$  as a decreasing function over  $t_1$  and  $t_2$ . So far no frailty models are known to yield increasing or bathtub-shaped dependence function. It would be undesirable to assume the shape of  $\theta^*(t_1, t_2)$  via  $g(Z_i)$  being constant or decreasing in advance, because researchers do not have previous information about  $\theta^*(t_1, t_2)$  or the dependence structure among failure times can be of interest to the study.

In this chapter we present a frailty model similar to (5.65). However, to accommodate the biological nature of our research, we modify the assumption on frailty made by (5.65), that is, frailty or liability specific to family and uniform throughout all time as follows. Suppose that in genetic studies the genetic effect persists only in early age, or only in late age, or early and late age (not in between), corresponding to decreasing, increasing, or bathtub-shaped dependence functions, respectively. Then, it is more appropriate for a frailty to vary across time intervals by having a nested structure  $Z_{ijk} = \alpha_i + \epsilon_{ijk}$  for the  $k$ -th interval and the  $j$ -th member in the  $i$ -th family. This was first proposed by Paik et al. (1994). With this structure of the frailties,  $\theta^*(t_1, t_2)$  is defined in each time interval and could vary over time in any fashion. Using the resulting model, we can avoid prespecifying the structure of dependence among failure times via  $g(Z_i)$ , which is the main disadvantage of frailty models. Therefore, we can estimate the regression parameters for the marginal hazard as well as the dependence function without assuming the functional form of  $\theta^*(t_1, t_2)$ . Also this model includes existing models such as the independent piecewise exponential (Breslow, 1974) or the gamma frailty model by Clayton (1978) and Oakes (1982) as special cases.

### 5.9.1 Frailty Models and Dependence Function

In a multivariate setting, define a measure of dependence among failure times as

$$\theta^*(t) = \frac{S(t)D_k D_j S(t)}{D_k S(t)D_j S(t)},$$

where  $t = (t_1, t_2, \dots, t_n)$  and  $D_j$  denotes  $\partial/\partial t_j$ . This is a ratio of two hazards:  $-D_k D_j S(t)/D_k S(t)$  denotes the hazard of member  $j$  given that member  $k$  failed at  $T_k = t_k$  while other members survived as of  $t_i$ , i.e.,  $T_i > t_i$  for  $i = 1, 2, \dots, j-1, j+1, \dots, k-1, k+1, \dots, n$ ; and  $-D_j S(t)/S(t)$  denotes the hazard of member  $j$  given  $T_i > t_i$  for  $i = 1, 2, \dots, j-1, j+1, \dots, n$ . In the bivariate case, Oakes (1989) showed that  $\theta^*(t)$  is a function of  $t$  only

through  $S(t)$  and  $g(Z)$  can uniquely determine  $\theta^*(t)$  as a function of  $S(t)$ . This result extends straightforwardly to the case of the multivariate distribution. However, when more than two failure times are considered, there are many other dependence functions to describe the dependence structure. To aid further discussion, we can rewrite  $\theta^*(t)$  as a function of  $S(t)$ ,  $\theta^*(t) = \theta^\#(S(t))$ . Similarly

$$\theta^\#(D_m S(t)) = \frac{D_m S(t) D_k D_j D_m S(t)}{D_j D_m S(t) D_k D_m S(t)}$$

is a ratio of two hazards: the hazard of member  $j$  given that member  $m$  and member  $k$  have failed while the rest of the members have not, and the hazard of member  $j$  given that member  $m$  has failed and the rest of members including  $k$  have not.

We define piecewise gamma frailty using the following three assumptions:

- (i) The conditional hazard function on a random effect follows the independent piecewise exponential model. Consider a set of discrete time-points  $a_0 < a_1 < \dots < a_K$ . The time axis can be divided into  $K$  intervals  $I_i = [a_{i-1}, a_i], i = 1, \dots, K$ , with  $a_0 = 0$  and  $a_K < \infty$ . Let  $Z_{ijk}$  be the unobservable random effect of the  $j$ -th member of the  $i$ -th family for the  $k$ -th time interval where  $k = 1, \dots, K$ . Also let  $h_{ijk}$  be the hazard function for the  $j$ -th individual in the  $i$ -th unit in  $I_k$ . Conditioning on  $Z_{ijk}$ , we assume that

$$h_{ijk} = Z_{ijk} h_{ijk}^*, \quad (5.68)$$

where  $h_{ijk}^* = \exp(y_{ij}\beta + \phi_k)$  and  $S^*(e_{ijk}) = \exp(-e_{ijk}h_{ijk}^*)$  are the baseline hazard and survivor functions, respectively,  $y_{ij}$  is a  $1 \times p$  vector of independent variables,  $\beta$  is a  $p \times 1$  vector of unknown parameters of interest, and  $\exp(\phi_k)$  is the underlying hazard for the  $k$ -th time interval. Expression (5.68) implies that the baseline hazard,  $h_{ijk}^*$ , is modified by a varying amount,  $Z_{ijk}$ , over each time interval.

- (ii) Frailty has a nested structure.  $Z_{ijk}$  can be further decomposed into a unit-specific factor and pure random fluctuation:

$$Z_{ijk} = \alpha_i + \epsilon_{ijk}.$$

- (iii)  $\alpha_i$  and  $\epsilon_{ijk}$  are independently distributed as gamma  $g(\alpha_i; \mu_1, \nu)$  and  $g(\epsilon_{ijk}; \mu_2, \gamma_k)$ , where

$$\begin{aligned} g(\alpha_i; \mu_1, \nu) &= \Gamma(\mu_1/\nu)^{-1} \{ \exp(-\alpha_i/\nu) \} (\alpha_i/\nu)^{\mu_1/\nu} / \alpha_i, \\ g(\epsilon_{ijk}; \mu_2, \gamma_k) &= \Gamma(\mu_2/\gamma_k)^{-1} \{ \exp(-\epsilon_{ijk}/\gamma_k) \} (\epsilon_{ijk}/\gamma_k)^{\mu_2/\gamma_k} / \epsilon_{ijk}, \\ \mu_1 + \mu_2 &= 1. \end{aligned}$$

The nested structure of  $Z$  is assumed for the following reasons. The

dependence among failure times within a unit is due to the correlation between  $Z_{ijk}$  and  $Z_{ims}$ . By sharing  $\alpha_i$  within a unit,  $t_{ij}$  and  $t_{im}$  are marginally correlated for  $j \neq m$ . In addition, since the variance of  $\epsilon_{ijk}$  changes over the time interval, the correlation between  $t_{ij}$  and  $t_{im}$  is distinct in each joint interval of  $t_{ij}$  and  $t_{im}$ . The restriction  $\mu_1 + \mu_2 = 1$  is imposed because the random effects play a role of inducing the dependence among failure times but not a role of reflecting changes in the conditional hazard.

Assumption (i) gives

$$\Pr(T_{ij} > t_{ij} | Z_{ij1}, Z_{ij2}, \dots, Z_{ijk}) = \prod_{k=1}^K \exp\{-Z_{ijk} e_{ijk} \exp(y_{ij}\beta + \phi_k)\},$$

where

$$e_{ijk} = \begin{cases} a_{k+1} - a_k & \text{if } t_{ij} \geq a_{k+1} \\ t_{ij} - a_k & \text{if } a_k \leq t_{ij} < a_{k+1} \\ 0 & \text{if } t_{ij} < a_k \end{cases}.$$

Based on (i), (ii), and (iii), the resulting joint survivor function of the piecewise gamma frailty model is

$$\begin{aligned} S(t_{i1}, t_{i2}, \dots, t_{in}) &= \int \cdots \int \prod_{j,k} \Pr(T_{ij} > t_{ij} | Z_{ij1}, Z_{ij2}, \dots, Z_{ijk}) g(\alpha_i; \mu_1, \nu) d\alpha_i \\ &\quad \times g(\epsilon_{ijk}; \mu_2, \gamma_k) \prod_{j,k} d\epsilon_{ijk} \\ &= \left( \frac{1}{1 + \nu A_i} \right)^{\mu_1/\nu} \prod_{j,k} \left( \frac{1}{1 + \gamma_k H_{ijk}} \right)^{\mu_2/\gamma_k}, \end{aligned}$$

where

$$H_{ijk} = -\ln S^*(e_{ijk}) \quad \text{and} \quad A_i = \sum_{j=1}^n \sum_{k=1}^K H_{ijk}. \quad (5.69)$$

The marginal hazard when  $t_{ij} \in I_k$  is

$$h(0, \dots, t_{ij}, \dots, 0) = h_{ijk}^* \left\{ \mu_1 \left( 1 + \nu \sum_{k=1}^K H_{ijk} \right)^{-1} + \mu_2 (1 + \gamma_k H_{ijk})^{-1} \right\}. \quad (5.70)$$

The bivariate marginal dependence function at  $t_{ij} \in I_k$  and  $t_{im} \in I_s$  is

$$\begin{aligned}\theta^*(0, , 0, t_{ij}, 0, \dots, t_{im}, \dots, 0) \\ = 1 + \left\{ \mu_1 \nu \left\{ 1 + \nu \sum_{k=1}^K (H_{ijk} + H_{imk}) \right\}^{-2} \right\} / \\ \left\{ \left\{ \mu_1 \left( 1 + \nu \sum_{k=1}^K (H_{ijk} + H_{imk}) \right)^{-1} + \mu_2 (1 + \gamma_k H_{ijk})^{-1} \right\} \right. \\ \left. \left\{ \mu_1 \left( 1 + \nu \sum_{k=1}^K (H_{ijk} + H_{imk}) \right)^{-1} + \mu_2 (1 + \gamma_s H_{ims})^{-1} \right\} \right\}.\end{aligned}$$

The piecewise gamma frailty model includes other models as special cases. First notice that when  $\mu_1 = 1$ , parameters in the unknown vector  $\gamma$  are not defined. When  $\nu = 0$  and  $\mu_1 = 1$ , representing a degenerate distribution of  $\alpha_i$  and  $Z_{ijk} = 1$ , the model reduces to the independent piecewise exponential model by Breslow (1974). When  $\nu = 0$  and  $0 < \mu_1 < 1$ , the model is not piecewise exponential but still assumes independence. In these cases  $\theta^*(0, 0, t_{ij}, 0, \dots, t_{im}, \dots, 0) = 1$  for  $j \neq m$ , implying independence among failure times. When  $\mu_1 = 1$  and  $\nu > 0$ , representing a degenerate distribution of  $\epsilon_{ijk}$ , the model reduces to the gamma frailty model and

$$\theta^*(0, 0, \dots, t_{ij}, 0, \dots, t_{im}, \dots, 0) = 1 + \nu.$$

This is identical with (5.68), a dependence function of Clayton (1978), Oakes (1982) and their multivariate versions.

Note that the frailty model discussed by Oakes (1989) is a subclass of the Archimedean distribution studied by Genest and MacKay (1986), but the piecewise gamma frailty model is not.

### 5.9.2 Example: Epilepsy Data

The Epilepsy Family Study of Columbia University (EFSCU; Ottman, Hauser, and Susser, 1985) was designed to test various genetic and nongenetic hypotheses relating to the familial distribution of epilepsy. In this study, 1957 adults ( $\geq 18$  years) with epilepsy (probands) were ascertained through a telephone survey of 10 voluntary organizations for epilepsy, conducted between 1985 and 1988. Each proband was interviewed about the occurrence of seizures and related disorders in close relatives.

Previous studies have demonstrated higher risks of epilepsy in relatives of affected persons than in the general population, but the role of genetic factors in causing this familial aggregation is unclear. An important aspect of the familial aggregation is the consistent observation of higher risks of epilepsy in offspring of affected women than in offspring of affected men (Ottman et

al., 1988). This finding is inconsistent with any conventional genetic model, including X-linkage, and is a central focus of the study. Also of interest is the comparison of risks in offspring of parents with early vs late age at onset, and examination of the persistence of genetic effect with increasing age of the offspring.

In this example, EFSCU data on all 695 offspring from 302 probands were used to answer the aforementioned research interests. Since samples were obtained through ascertainment of cases, failure times of the probands are left truncated; that is, one can sample probands only if the current age is greater than the age of onset of epilepsy. The main interest is the association between failure rate of offspring and characteristics of offspring who were followed from birth. Unless the failure time of the offspring depends on the current age of the proband, left truncation does not affect the results. Paik et al. (1994) obtained the estimation procedures and fitted piecewise gamma model for this data.



# Chapter 6

---

## *Estimation Methods for Shared Frailty Models*

6.1	Introduction .....	105
6.2	Inference for the Shared Frailty Model .....	106
6.3	The EM Algorithm .....	108
6.4	The Gamma Frailty Model .....	110
6.5	The Positive Stable Frailty Model .....	111
6.6	The Lognormal Frailty Model .....	113
6.6.1	Application to Seizure Data .....	113
6.7	Modified EM (MEM) Algorithm for Gamma Frailty Models .....	114
6.8	Application .....	116
6.9	Discussion .....	117

---

### 6.1 Introduction

In this chapter, we discuss different method of estimation for shared frailty models which are used more often and they have a lot of applications. Frailties are usually viewed as unobserved covariates. This led to the use of the EM algorithm as an tool. However, the algorithm is slow, variance estimates require further computation. Penalized partial likelihood (PPL) approach provides an alternate approach. The frailty terms are treated as additional regression coefficients which are constrained by a penalty function added to the log-likelihood. They are computationally similar to other shrinkage methods for penalized regression such as ridge regression, the lasso and smoothing splines. Standard algorithms for fitting Cox semiparametric and parametric models can be simply extended to include penalty functions. These methods usually converge quickly and produce both point and variance estimates for model parameters. The EM algorithm and PPL approach give the same results in the case of the gamma frailty model. It is proved theoretically in Duchateau and Janssen (2008). That is why PPL approach can be safely used for the gamma frailty density.

## 6.2 Inference for the Shared Frailty Model

Ebrahimi et al. (2004) presented a review of both a semiparametric and a parametric approach to estimating the risk coefficients and the frailty parameters for the gamma frailty, the positive stable frailty and the frailty parameters for the gamma frailty, the positive stable frailty and the lognormal frailty models. They also presented an application of frailty models to seizure data. We present briefly some of their work on this topic. To derive the general form of the likelihood function, it is assumed that the common factor causes dependence between individuals in a given group, and conditional on that, all individuals within the group are independent. Thus, for one group of  $n$  individuals, the conditional joint survival distribution of failure times  $T_1, T_2, \dots, T_n$  is given by

$$\begin{aligned} P(T_1 > t_1, \dots, T_n > t_n | z) &= P(T_1 > t_1 | z)p(T_2 > t_2 | z) \cdots P(T_n > t_n | z) \\ &= \exp \left\{ -z \sum_{j=1}^n H_0(t_j) \exp(\beta Y_j) \right\}. \end{aligned} \quad (6.1)$$

The above joint conditional survival distribution holds for any group. Integrating the frailty out, we get the joint survival function for this group as

$$\begin{aligned} S(t_1, \dots, t_n) &= P(T_1 > t_1, \dots, T_n > t_n) \\ &= \int_0^\infty P(T_1 > t_1, \dots, T_n > t_n | z) f(z) dz \\ &= \int_0^\infty \exp \left\{ -z \sum_{j=1}^n H_0(t_j) \exp(\beta Y_j) \right\} f(z) dz \\ &= L \left[ \sum_{j=1}^n H_0(t_j) \exp(\beta Y_j) \right], \end{aligned} \quad (6.2)$$

where  $L$  is the Laplace transform of the density function  $f$  and  $H_0(t) = \int_0^t h_0(z) dz$ .

From (6.2) it is clear that the joint survival function for one group is the Laplace transform of the frailty density function  $f$  with parameter  $\sum_{j=1}^n H_0(t_j) \exp(\beta Y_j)$ . In principle, any distribution on the positive numbers can be applied as a frailty distribution. In this chapter, we concentrate on the gamma, the lognormal, and the positive stable distributions. For other distributions see Hougaard (2000) and Ohman and Eberly (2001).

From (6.2) one can derive the likelihood function for one group as follows: If the failure time is observed for the  $j$ -th individual at time  $t_j$ , its probability is given by

$$\begin{aligned}
P(T_j = t_j, T_1 > t_1, T_2 > t_2, \dots) \\
&= -\frac{\partial S(t_1, \dots, t_n)}{\partial t_j} \\
&= -h_0(t_j) \exp(\beta Y_j) L^{(1)} \left( \sum_{j=1}^n H_0(t_j) \exp(\beta Y_j) \right), \tag{6.3}
\end{aligned}$$

where  $L^{(1)}(s)$  denotes the first derivative of  $L(s)$  with respect to  $s$ . Let  $D. = \sum \delta_j$ , the total number of failures in the group, and  $\theta$  be the parameter of the frailty distribution. Then, using Eq. (6.3), the likelihood for one group is given by

$$(-1)^{D.} \left\{ \prod_{j=1}^n h_0(t_j)^{\delta_j} \exp(\delta_j \beta Y_j) \right\} L^{(D.)} \left( \sum_{j=1}^n H_0(t_j) \exp(\beta Y_j) \right). \tag{6.4}$$

The likelihood function for all individuals is constructed by multiplying the group likelihoods together. Specifically, if  $D_i$  denotes the number of failures in the  $i$ -th group, and  $D = \sum_{i=1}^G D_i$ , then the likelihood function is given by

$$(-1)^D \prod_{i=1}^G \left\{ \prod_{j=1}^{n_i} h_0(t_{ij})^{\delta_{ij}} \exp(\delta_{ij} \beta Y_{ij}) \right\} L^{(D_i)} \left( \sum_{j=1}^{n_i} H_0(t_{ij}) \exp(\beta Y_{ij}) \right). \tag{6.5}$$

If we assume a parametric form for  $h_0$ , we can handle the estimation in the usual way by differentiating the log likelihood function. If a parametric form is not assumed for  $h_0$ , there are several estimation methods available to handle this semiparametric model. These methods are described below.

The full conditional approach is to use the likelihood function (6.5) and insert non-parametric expressions for  $H_0(t)$ ,  $H_0(t) = \sum_{t_k \leq t} h_{0k}$ , assuming a discrete contribution  $h_{0k}$  at each time of failure. Here  $t_k$  is the  $k$ -th smallest failure time, regardless of the subgroup,  $M$  is the number of distinct failures, and  $d_k$  is the number of failures at  $t_k$ ,  $k = 1, \dots, M$ . Thus we get the function with parameters  $\beta, \theta, h_{01}, \dots, h_{0M}$ . Now, we follow the traditional method for estimation: differentiate the log likelihood with respect to all parameters and use the Newton-Raphson method to obtain estimates. This approach takes much more iterations for convergence than other methods, so it is not preferable when we deal with a typical frailty distribution. However, for the data set with complicated dependence structures, we may end up with the likelihood function which cannot be handled by simpler methods. In this case, the full conditional approach may be the only resolution.

The EM algorithm, which is a simpler method, can also be used for estimation since we can treat the frailty as covariates. Kelin (1992) developed

the estimation based on the EM algorithm for the gamma frailty. A similar method was developed by Wang et al. (1995) for the positive stable frailty. Shu and Kelin (1999) also offer SAS macros for the gamma and the positive stable frailty models on the internet. We discuss this algorithm in more detail in the next section.

S-plus (version 6) offers the frailty function based on the penalized likelihood approach (Good and Gaskin, 1971). An excellent example of the penalized likelihood estimates is given by Tapia and Thompson (1978). The penalized approach has some similarities to the EM algorithm. The penalized likelihood is constructed by the product of the partial likelihood, including the frailty terms as parameters, and a penalty function which describes the roughness of the curve under consideration. An iterative estimation procedure is used by maximizing the partial likelihood first, then modifying the frailties by the penalized likelihood. In S-plus, there are two penalty functions available within the frailty function to simulated gamma frailty and the lognormal frailty. For more details about this approach see Hougaard (2000). Most of the analysis in this chapter was done by using corresponding functions in S-plus.

### 6.3 The EM Algorithm

The most commonly applied estimation method for parallel data with covariates is the EM algorithm. The EM algorithm is a combination of an expectation step (E-step) and a maximization step (M-step). This method does not use the likelihood function (6.5), rather the full likelihood function. In the E-step of the algorithm the expected value of  $L_{full}$  is computed given the current estimates of the parameters and the observable data. In the M-step, plug estimates of frailties into the modified partial likelihood, update the estimates of  $\beta$  and  $h_0$ ; plug into  $L_{full}$ , update the estimate  $\theta$ . In the M-step estimates of parameters which maximize the expected value of  $L_{full}$  from the E-step are obtained. We define the full likelihood as the product of the conditional and the density of frailties.

$$L_{full} = \prod_{i=1}^G \prod_{j=1}^{n_i} (z_i)^{\delta_{ij}} h_0(t_{ij})^{\delta_{ij}} \exp(\delta_{ij}\beta Y_{ij}) \exp\{-z_i H_0(t_{ij}) \exp(\beta Y_{ij})\} f(z_i). \quad (6.6)$$

In Eq. (6.6), the first term is a standard survival likelihood given the frailties. The most basic difference between the likelihood (6.5) and the full likelihood is in the way frailty is treated. For the likelihood, the frailty is unobservable and thus inestimable; the only information about the frailty is the assumed distribution  $f(z)$  with the unknown parameter  $\theta$ . So keep only  $\theta$  in the likelihood function. However, for the full likelihood, pretend that the frailties are estimable. In this approach, first estimate the frailty for each

group based on the observable data and the initial values of  $h_0, \beta, \theta$ . Then plug in the estimates of  $z_1, \dots, z_G$  into the modified partial likelihood to derive the estimates of  $\beta$  and  $h_0$  and then into the likelihood of  $\theta$  to get  $\hat{\theta}$ . An advantage of estimating  $z_1, \dots, z_G$  is that one can use the hazard functions,  $z_i h_0(t_{ij}) \exp(\beta Y_{ij})$ ,  $i = 1, \dots, G$ , to construct the corresponding partial likelihood, where one can obtain the estimate of  $\beta$  and  $h_0$ . Furthermore, estimate of  $\theta$  is easy to derive by inserting  $\hat{z}_i$ ,  $i = 1, \dots, G$ , into the corresponding likelihood. Compared with the full conditional approach, where the likelihood is maximized directly, the EM algorithm is much simpler.

More specifically, the steps of the EM algorithm are as follows:

- (1) Provide initial values of  $\beta, h_0$  and  $\theta$ .
- (2) In the E-step, plug values of  $\beta, h_0$  and  $\theta$  into the full likelihood (6.6) and calculate the conditional expectation of  $z_i$  and  $\ln(z_i)$  given the observable data. Parner (1997) suggested a general formula for the expectation of frailties,

$$E(z_i) = -\frac{L^{(D_i+1)}[\sum_j \hat{H}_0(t_{ij}) \exp(\hat{\beta} Y_{ij})]}{L^{(D_i)}[\sum_j \hat{H}_0(t_{ij}) \exp(\hat{\beta} Y_{ij})]}, \quad i = 1, \dots, G. \quad (6.7)$$

In this step, the unobserved terms in the log-likelihood are removed by substitution with the mean value given the observations.

- (3) In the M-step, plug the expectation of frailties into the modified partial likelihood, update the estimates of  $\beta$  and  $h_0$ ; plug into Eq. (6.6), update the estimate of  $\theta$ . In this case, the partial likelihood turns out to be

$$L(\beta) = \prod_{k=1}^M \frac{\exp(\hat{z}_k(\beta s_k))}{\sum_{l \in R(t_k)} \hat{z}_l \exp(\beta Y_l)] d_k}, \quad (6.8)$$

where  $t_k$  is the  $k$ -th smallest failure time, regardless of subgroup,  $d_k$  is the number of failures at  $t_k$ ,  $D_k$  is the set of all individuals who fail at time  $t_k$ , and  $s_k = \sum_{j \in D_k} Y_j$ . Furthermore, the maximum likelihood estimate of  $h_{0_k}$  is

$$\hat{h}_{0_k} = \frac{d_k}{\sum_{l \in R(t_k)} \hat{z}_l \exp(\hat{\beta} Y_l)}, \quad k = 1, \dots, M. \quad (6.9)$$

Here, the frailty values are considered fixed and known.

- (4) Repeat the E-step and M-step until the estimates converge.

The standard errors of the estimates of  $h_0, \beta$ , and  $\theta$  can be obtained from the inverse of the observed information matrix.

## 6.4 The Gamma Frailty Model

Clayton (1978) proposed the gamma distribution for the frailty. Since then, the gamma frailty model has been used extensively because the derivatives of its Laplace transformation are quite simple. The density function of the frailty is

$$f(z) = \frac{z^{1/\theta-1} \exp(-z/\theta)}{\Gamma(1/\theta)\theta^{1/\theta}} \quad (6.10)$$

with the Laplace transform  $L(s) = (1 + \theta s)^{-1/\theta}$ . Usually, we use the one parameter gamma distribution denoted by  $\text{Gamma}(\theta)$ . Thus the mean of the frailty is 1, which is the desired property of the frailty distribution; the variance is  $\theta$ , which reflects the degree of dependence in the data. Large  $\theta$  indicates strong dependence.

From Eq. (6.2), it is easily seen that the marginal hazard is

$$\frac{h_0(t_j) \exp(\beta Y)}{1 + \theta H_0(t_j) \exp(\beta Y)}. \quad (6.11)$$

Consequently, for two randomly selected individuals with covariate values  $Y$  and  $Y^*$ , the marginal hazards are not proportional over time and the relative risk is given by

$$\exp(\beta(Y - Y^*)) \frac{1 + \theta H_0(t) \exp(\beta Y^*)}{1 + \theta H_0(t) \exp(\beta Y)}, \quad (6.12)$$

which depends on time through  $H_0(t)$ . From Eq. (6.12), it is clear that the relative risk starts at  $\exp(\beta(Y - Y^*))$  because no failure occurs, and converges to 1 as  $t \rightarrow \infty$  since  $H_0(t) \rightarrow \infty$  in this case. If we are interested in the relative risk at a given time  $t$ , we can calculate it by replacing  $\theta, \beta$  and  $H_0(t)$  with their estimates in Eq. (6.12).

One can derive the likelihood function as follows. The  $p$ -th derivative of the Laplace transform is

$$L^{(p)}(s) = (-1)^p \theta^p (1 + \theta s)^{-1/\theta-p} \Gamma\left(\frac{1}{\theta} + p\right) / \Gamma\left(\frac{1}{\theta}\right). \quad (6.13)$$

Following Eq. (6.5), the likelihood for all individuals is given by

$$\begin{aligned} & (-1)^D \prod_{i=1}^G \frac{\theta^{D_i} \Gamma(1/\theta + D_i)}{\Gamma(1/\theta)} \left\{ \prod_{j=1}^{n_i} h_0(t_{ij})^{\delta_{ij}} \exp(\delta_{ij}\beta Y_{ij}) \right\} \\ & \times \left\{ 1 + \theta \sum_{j=1}^{n_i} H_0(t_{ij}) \exp(\beta Y_{ij}) \right\}^{-1/\theta-D_i}. \end{aligned} \quad (6.14)$$

If we decide to use the EM algorithm, the full likelihood is

$$\begin{aligned} & \prod_{i=1}^G \prod_{j=1}^{n_i} h_o(t_{ij})^{\delta_{ij}} \exp(\delta_{ij}\beta Y_{ij}) \exp\{-z_i H_0(t_{ij}) \exp(\beta Y_{ij})\} \\ & \times \prod_{i=1}^G \frac{z_i^{1/\theta+D_i-1} \exp(-z_i/\theta)}{\Gamma(1/\theta)\theta^{1/\theta}}. \end{aligned} \quad (6.15)$$

Now the following procedure must be followed: Find estimates of  $\beta$  and  $H_0(t)$  in the model without frailty. This corresponds to setting  $\theta = 0$ . Use them as initial estimates of  $\beta$  and  $H_0(t)$  with  $\theta = 0$  as an initial estimate for  $\theta$ . For the E-step, obtain the expectation of frailty  $z_i$ . We can get it using Eq. (6.7) directly. However, for the gamma frailty, a short-cut is available. Based on  $L_{\text{full}}$ , the distribution of  $z_i$  given the observable data is still a gamma with shape parameter  $\tilde{\alpha}_i = 1/\theta + D_i$  and scale parameter  $\tilde{\theta}_i = 1/\theta + \sum_j H_0(t_{ij}) \exp(\beta Y_{ij})$ .

In general, it can be proved that for the gamma  $(\alpha, \theta)$ ,

$$E(z) = \alpha/\theta, \quad E(\ln(z)) = \psi(\alpha) - \ln \theta,$$

where  $\psi(\alpha)$  is the digamma function  $\Gamma'(\alpha)/\Gamma(\alpha)$ . Therefore, for the E-step, the expectation of  $z_i$  and the expectation of  $\ln(z_i)$  are  $\tilde{\alpha}_i/\tilde{\theta}_i$  and  $\psi(\tilde{\alpha}_i) - \ln(\tilde{\alpha}_i)$ , respectively. The general formula in Eq. (6.7) gives us exactly the same result for  $E(z_i)$ . For the M-step, obtain the estimate of  $\beta$  based on Eq. (6.8). The estimate of  $h_{0k}$  is given by Eq. (6.9) for  $k = 1, \dots, M$ , and the estimate of  $\theta$  is derived by maximizing the likelihood of  $\theta$ .

$$L(\theta) = \Gamma(1/\theta)^{-G} \theta^{-G/\theta} \prod_{i=1}^G \hat{z}_i^{1/\theta+D_i-1} \exp(-\hat{z}_i/\theta). \quad (6.16)$$

To get the standard errors of the estimates, we have to derive the information matrix and then plug in estimates of  $\beta, h_{01}, \dots, h_{0M}, \theta$ . The covariance matrix is the inverse of the observed information matrix.

## 6.5 The Positive Stable Frailty Model

In this section we assume the positive stable distribution for the frailty, see Hougaard (2000). For most frailty models, the marginal hazard functions are not proportional. The positive stable frailty model is an exception. This is an advantage of the positive stable frailty model.

Suppose the frailty has the positive stable distribution with parameter  $\theta$ .

We restrict  $0 < \theta \leq 1$  to get a distribution with positive numbers. The Laplace transform is  $L(s) = \exp(-s^\theta)$ . The marginal distribution of  $T_j$  is given by

$$P(T_j > t) = \exp\{-H_0(t_j)^\theta \exp(\theta\beta Y)\}. \quad (6.17)$$

Thus, the integrated hazard and the hazard function are  $H_0(t_j)^\theta \exp(\theta\beta Y)$ , and  $\theta h_0(t_j) H_0(t_j)^{\theta-1} \exp(\theta\beta Y)$ , respectively. For the marginal distributions of two individuals with covariate values  $Y$  and  $Y^*$ , the relative risk is the constant  $\exp(\theta\beta(Y - Y^*))$ .

The  $p$ -th derivative of Laplace transform is

$$L^{(p)}(s) = (-1)^p \exp(-s^\theta) \sum_{m=1}^p c_{p,m} \theta^m s^{m\theta-p}, \quad (6.18)$$

where  $c_{p,m}$  is a polynomial in  $\theta$  of degree  $m$  and is defined recursively by

$$c_{p,p} = 1, \quad c_{p,1} = \Gamma(p-\theta)/\Gamma(1-\theta),$$

and

$$c_{p,m} = c_{p-1,m-1} + c_{p-1,m} \{(p-1) - m\theta\}.$$

Following Eq. (6.5), the likelihood function for all individuals is given by

$$\begin{aligned} & \prod_{i=1}^G \left\{ \prod_{j=1}^{n_i} h_0(t_{ij})^{\delta_{ij}} \exp(\delta_{ij}\beta Y_{ij}) \right\} \exp \left\{ - \left[ \sum_{j=1}^{n_i} H_0(t_{ij}) \exp(\beta Y_{ij}) \right]^\theta \right\} \\ & \times \sum_{m=1}^{D_i} c_{D_i,m} \theta^m \left[ \sum_{j=1}^{n_i} H_0(t_{ij}) \exp(\beta Y_{ij}) \right]^{m\theta-D_i}. \end{aligned} \quad (6.19)$$

One may use the EM algorithm for estimation. For the E-step, using the general formula Eq. (6.7), the expectation of frailty  $z_i$  is given by

$$E(z_i) = \frac{\sum_{m=1}^{D_i+1} c_{D_i+1,m} \theta^m [\sum_j H_0(t_{ij}) \exp(\beta Y_{ij})]^{m\theta-D_i-1}}{\sum_{m=1}^{D_i} c_{D_i,m} \theta^m [\sum_j H_0(t_{ij}) \exp(\beta Y_{ij})]^{m\theta-D_i}}.$$

For the M-step, one has a trouble in estimating  $\theta$  because of its likelihood function, involving the positive stable density, is really complicated. Wang et al. (1995) suggested a practical algorithm for providing values for  $\theta$  and searching for the value which maximizes Eq. (6.19).

---

## 6.6 The Lognormal Frailty Model

McGilchrist and Aisbett (1991) proposed the lognormal distribution with mean  $\xi$  and variance  $\sigma^2$  for the frailty. Therefore, the density of frailty is given by

$$f(z) = \frac{1}{z(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{(\ln(z))^2}{2\sigma^2}\right\}. \quad (6.20)$$

Unfortunately, the Laplace transform of this distribution is theoretically intractable. Hence, the explicit form of marginal hazard is not available to us. They offered the penalized partial likelihood (PPL) approach for estimation. This approach is implemented in the frailty function in S-plus (version 6). This approach is explained in detail by Duchateau and Janssen (2008).

### 6.6.1 Application to Seizure Data

In this section Ebrahimi et al. (2004) used the frailty model described in the previous section to asses the rate of seizure recurrence after a first unprovoked seizure in childhood. The seizure data came from a cohort of 407 children (one month through 19 years of age) prospectively identified for their first unprovoked seizure in Bronx, New York between 1983 and 1992 and followed through September 1998. (See Shinnar et al., 2000). The dates of occurrence of the first ten seizure recurrences were recorded, along with several covariates. The first covariate is the etiology of the seizure which is classified as remote symptomatic (RS) (having neurological abnormality associated with an increased risk of seizures and which is taken to be the presumed cause of the seizure) versus all other causes. The second covariate is the asleep-aware state, which is coded as 1 if the child was asleep when the initial seizure occurred, and given 0 if the child was awake. The third covariate is electroencephalogram (EEG). There are three categories based on children's EEG: normal, abnormal, and not performed. Ebrahimi et al. (2004) used two categories to denote these categories. EEG is coded as 1 for abnormal EEG, and 0 for others. A second variable (MissEEG) is coded as 1 if the child did not have an EEG, and 0 for others. Ebrahimi et al. (2004) analyzed seizure data based on the gamma frailty model. They showed that based on the AIC, the gamma distribution is a better fit.

## 6.7 Modified EM (MEM) Algorithm for Gamma Frailty Models

Klein and Moeschberger (2003) provide an EM algorithm (KM-EM) for semiparametric PH frailty model and use the information matrix to calculate the variances of the MLEs. Therneau et al. (2000) and Ripatti and Palmgren (2000) propose a penalized partial likelihood (PPL) method. However, both methods involve the calculation of matrix inverses, which depends on the group number  $G$ , the number of  $\beta$  parameters  $p$ , and the number of distinct even times  $r$ . The information matrix is a  $(p+r+1) \times (p+r+1)$  non-diagonal matrix in the KM-EM algorithm. The PPL method involves creating and calculating the inverse of  $(G+p) \times (G+p)$  non-diagonal matrices. As  $G$  and  $r$  get large, especially when greater than 10,000, it is impossible to find the inverse of high dimension matrices.

The KM-EM algorithm is modified in that M-step is accomplished by a Newton-Raphson step for frailty parameter  $\theta$ , and it is standard statistical procedure for the parameters  $(\beta, \lambda_0)$ . This method can handle the data involving a large number of clusters (groups) and distinct event times.

The KM-EM algorithm (Klein and Moeschberger, 2003) has been used to estimate Gamma frailty models. If the frailties  $\mathbf{Z} = \{z_i, i = 1, \dots, G\}$  were observed, then the survival times  $\mathbf{T} = (t_{i1}, \dots, t_{in_i})$  are independently conditioned on  $z_i$ . The loglikelihood function for the complete data  $\mathbf{Y} = (\mathbf{T}, \delta, \mathbf{Y}, \mathbf{Z})$  is

$$L_F(\theta, \beta, \lambda_0) = \sum_{i=1}^G \left[ \ln g(z_i) + \sum_{j=1}^{n_i} \{ \delta_{ij} \ln(\lambda(t_{ij}|Y_{ij})z_i) - z_i \Lambda(t_{ij}|Y_{ij}) \} \right]$$

which can be decomposed into two parts  $L_F(\theta, \beta, \lambda_0) = L_1(\theta) + L_2(\beta, \lambda_0)$ , where

$$L_1(\theta) = -G \left\{ \frac{\ln \theta}{\theta} + \ln \Gamma \left( \frac{1}{\theta} \right) \right\} + \sum_{i=1}^G \left\{ \left( \frac{1}{\theta} + D_i - 1 \right) \ln z_i - \frac{z_i}{\theta} \right\},$$

$$L_2(\beta, \lambda_0) = \sum_{i=1}^G \sum_{j=1}^{n_i} \{ \delta_{ij} \ln \lambda(t_{ij}|Y_{ij}) - z_i \Lambda(t_{ij}|Y_{ij}) \}.$$

In the E-step, the expected value of complete data log-likelihood  $L_F(\theta, \beta, \lambda_0)$  is computed, given the current parameter estimates and the observed data. In the M-step, the maximum likelihood estimates are obtained by maximizing the expected value of  $L_F(\theta, \beta, \lambda_0)$ .

To apply the E-step, the expectation of  $z_i$  and  $\log z_i$  are calculated given the data and current estimates  $(\beta, \theta)$ . The frailties  $z_i$  are i.i.d. gamma random

variables with shape parameter  $A_i = 1/\theta + D_i$  and scale parameter  $C_i = 1/\theta + \sum_{j=1}^{n_i} \Lambda_0(t_{ij}) \exp(\beta Y_{ij})$ . Thus,

$$E[z_i] = \frac{A_i}{C_i} \quad \text{and} \quad E[\ln z_i] = \psi(A_i) - \ln(C_i),$$

where

$$\psi(\alpha) = \frac{d \ln \Gamma(\alpha)}{d\alpha}$$

is the digamma function. The expectation of  $L_2(\beta, \lambda_0)$  is

$$E(L_2(\beta, \lambda_0) | \delta_{ij}, t_{ij}, Y_{ij}; z_i) = \sum_{i=1}^G \sum_{j=1}^{n_i} \left\{ \delta_{ij} [\ln \lambda(t_{ij} | Y_{ij})] - \frac{A_i}{C_i} \Lambda(t_{ij} | Y_{ij}) \right\}. \quad (6.21)$$

Let  $G_i = \ln(A_i/C_i)$ . The expected log-likelihood (6.21) can be considered as the partial log-likelihood of the Cox model with a group specific covariate  $G_i$  with a known coefficient 1.

The M-step by Klein and Moeschberger (2003) uses the SAS interactive matrix language. Here, the modified M-step uses a standard SAS procedure to find the MLE of regression parameters  $(\beta, \lambda_0)$  and uses Newton-Raphson step to find the MLE of frailty parameter  $\theta$ . In the modified M-step, the MLEs of log-likelihood (6.21) can be obtained by the SAS procedure PROC PHREG using option OFFSET for additional covariate  $G_i$ . Given the additional covariate  $G_i$ , the cumulative hazard is

$$\Lambda_{ij}(t_{ij}) = \Lambda(t_{ij} | Y_{ij}) z_i = \Lambda_0(t_{ij}) \exp(\beta^t Y_{ij} + G_i).$$

The estimates of  $\Lambda_{ij}(t_{ij})$  can be output directly from PROC PHREG. Hence,

$$\sum_{j=1}^{n_i} \Lambda_0(t_{ij}) \exp(\beta^t Y_{ij}) = \frac{\sum_{j=1}^{n_i} \Lambda_{ij}(t_{ij})}{\exp(G_i)} = C_i - \frac{1}{\theta}$$

can be used to calculate  $C_i$  in the next E-step.

Let  $\alpha = 1/\theta$ . The expected log-likelihood

$$\begin{aligned} & E(L_1(\alpha) | \delta_{ij}, t_{ij}, Y_{ij}; z_i) \\ &= G \times [\alpha \ln \alpha - \ln \Gamma(\alpha)] + \alpha \sum_{i=1}^G \left[ \psi(A_i) - \ln C_i - \frac{A_i}{C_i} \right]. \end{aligned}$$

Let  $Y = [\psi(A_i) - \ln C_i - A_i/C_i]/G$ . Then equivalently, we can just maximize

$$m(\alpha) = \alpha \ln \alpha - \ln \Gamma(\alpha) + \alpha Y.$$

Variable	EM algorithm		PPL	
	Estimate	S.E.	Estimate	S.E.
Age	0.0035	0.0111	0.0034	0.0111
Sex	-1.4746	0.3578	-1.4715	0.3579
GN	0.0908	0.4066	0.0894	0.4068
AN	0.3520	0.4001	0.3518	0.4002
PKD	-1.4237	0.6307	-1.4277	0.6309
$\theta$	0.0043	0.6607	$5 \times 10^{-7}$	N/A

TABLE 6.1: Parameter estimates of  $(\beta, \theta)$  with standard errors

The derivatives are  $m'(\alpha) = \ln \alpha + 1 + Y - \psi(\alpha)$  and  $m''(\alpha) = 1/\alpha - \psi(1, \alpha)$ , where

$$\psi(\alpha) = \frac{d \ln \Gamma(\alpha)}{d \alpha} \quad \text{and} \quad \psi(1, \alpha) = \frac{d^2 \ln \Gamma(\alpha)}{d \alpha^2}$$

can be obtained by the SAS functions DIGAMMA( $\alpha$ ) and TRIGAMMA( $\alpha$ ). The Newton-Raphson step to obtain the MLE of  $\alpha$  can be achieved by

$$\alpha_{k+1} = \alpha_k - \frac{m'(\alpha_k)}{m''(\alpha_k)}.$$

The MEM algorithm avoids complicated matrix computation and can handle large data set with thousands of groups and distinct event times. A profile likelihood EM algorithm suggested by Klein and Moeschberger (2003) and Nielsen et al. (1992) can be also naturally implemented. The profile likelihood of parameter  $\theta$  is defined as

$$PL(\theta) = \sup_{\beta_\theta} L(\theta, \beta_\theta). \quad (6.22)$$

For each fixed value of  $\theta$  in a specific range, the EM algorithm described above is used to obtain an estimate of  $\beta_\theta$ . The value of the profile likelihood is then given by  $L(\theta, \beta_\theta)$ . The value of  $\theta$  which maximizes this quantity is, then, the MLE.

## 6.8 Application

McGilchrist and Aisbett (1991) analyze the data on the recurrence times from the point of insertion of the catheter to infection for kidney patients using

portable dialysis equipment. Catheters may be removed for reasons other than infection, in which case the observation is censored. There are 38 patients and each has exactly two observations. The risk variables are age, sex (0 = male, 1 = female), and disease type coded as 0 = Glomerulo Nephritis (GN), 1 = Acute Nephritis (AN), 2 = Polycystic Kidney Disease (PKD), 3 = Other. The five regression variables fitted are age, sex, and presence/absence of disease types GN, AN, PKD.

For comparison, gamma frailty model is fitted by both the MEM algorithm and the PPL method (See Yu, 2006). Because lognormal distribution was used for the frailty by McGilchrist and Aisbett (1991), the parameter estimates are different. From [Table 6.1](#), we see that the parameter estimates from the MEM algorithm, especially the standard errors, are very close to the values from the PPL method.

---

## 6.9 Discussion

The modified EM algorithm utilizes the standard statistical procedures to find out the MLE, then multiple imputation is used to find the standard errors. This approach avoids the calculation of inverse of matrix. Based on the simulations, the three estimation methods produced similar parameter estimates when the number of groups is small, say less than 100. The EM algorithms give the standard error for the frailty parameter estimate  $\hat{\theta}$ , whereas the PPL method does not. For a medium number of groups, from one thousand and ten thousand, the KM-EM algorithm cannot run because the calculation of inverse of the matrix exhausts computer resources. Yu (2006) pointed out that when the number of groups gets even larger (greater than 100,000), the MEM algorithm is the only method that can handle such data set. In summary, when the number of groups is small or medium, the three methods produces similar results and are competitive. For a huge data set, the MEM algorithm is preferred. The EM algorithm and PPL approach give the same results in the case of the gamma frailty model, it is proved theoretically by Duchateau and Janssen (2008).



# Chapter 7

---

## *Analysis of Survival Data in Shared Frailty Models*

7.1	Introduction .....	119
7.2	Analysis for Bone Marrow Transplantation (BMT) Data .....	119
7.3	Analysis for Acute Leukemia Data .....	122
7.4	Analysis for HLA Data .....	125
7.5	Analysis for Kidney Infection Data .....	129
7.6	Analysis of Litters of Rats .....	131
7.7	Analysis for Diabetic Retinopathy Data .....	133

---

### 7.1 Introduction

Therneau and Grambsch (2000) have given some data analysis on frailty models using SPLUS and SAS software packages. In this chapter, we analyze some more data which was discussed in Chapter 1 using R-package which has free open access to everyone.

---

### 7.2 Analysis for Bone Marrow Transplantation (BMT) Data

Consider the data set on bone marrow transplantation (BMT) given in Section 1.2. We analyze the data using gamma frailty and Gaussian frailty models for the parametric baseline models using R statistical package. Let us consider lognormal distribution as the baseline model which has the highest loglikelihood among parametric family. The following is the R program.

```
bmt=read.table(file="C:/David/Book2/Data/BMT2.txt",header=T)
library(survival)
fit1=survreg(Surv(T,I)^~PAge+PSex+hospital+g,data=bmt,
dist='lognormal')
summary(fit1)
```

Here the variable of interest (response variable,  $T$ ) is time to death or relapse,  $I$  is an indicator (1-dead or relapsed, 0-alive or disease free). PAge, PSex, hospital, g are the covariates indicating patient age, patient sex (1-male, 0-female), disease group (1-ALL, 2-AML-low risk, 3-AML-high risk). The output of the R program is given below.

```
Call:
survreg(formula = Surv(T, I) ~ PAge + PSex + hospital + g,
        data = bmt, dist = "lognormal")

      Value Std. Error     z      p
(Intercept) 7.6086    0.6347 11.99 4.11e-33
PAge       -0.0190    0.0162 -1.17 2.42e-01
PSex        0.4147    0.3034  1.37 1.72e-01
hospital     0.3008    0.1361  2.21 2.71e-02
g           -0.3512    0.1936 -1.81 6.96e-02
Log(scale)   0.7553    0.0703 10.75 5.89e-27

Scale= 2.13

Log Normal distribution
Loglik(model)= -1008.1 Loglik(intercept only)= -1014
    Chisq= 11.84 on 4 degrees of freedom, p= 0.019
Number of Newton-Raphson Iterations: 3
n= 274
```

From the output of R program it is observed that the test for the regression parameters equal to zero is rejected with chi-square value 11.84 for 4 df and p-value is 0.019. Hospital is the most effective variable which is related to the survival of patients with p-value 0.027. The loglikelihood (intercept only) is -1014 for the lognormal distribution as a baseline. For other baseline distributions, the loglikelihood (intercept only) for Weibull baseline is -1024.1 and for loglogistic baseline is -1017.1. Among these three distribution, lognormal has the highest loglikelihood with value -1014. R package do not have gamma distribution as a baseline model for the survival function. Now we introduce frailty component in the above regression model. First we introduce gamma frailty and then Gaussian frailty models. R package do not have a positive stable, power variance function, Weibull as a frailty distributions. Following is the R program for gamma frailty model.

```
fit2=survreg(Surv(T,I)~PAge+PSex+hospital+g+frailty(pat),
data=bmt,dist='lognormal')
summary(fit2)
```

In the above R command patient is the frailty variable which has gamma distribution by default when not specified. The output of the R program is given below.

Call:

```
survreg(formula = Surv(T, I) ~ PAge + PSex + hospital + g
+ frailty(pat), data = bmt, dist = "lognormal")
```

	Value	Std. Error	z	p
(Intercept)	7.1304	0.6964	10.240	1.32e-24
PAge	-0.0116	0.0200	-0.579	5.63e-01
PSex	0.3341	0.3301	1.012	3.12e-01
hospital	0.1870	0.1463	1.278	2.01e-01
g	-0.0864	0.2360	-0.366	7.14e-01
Log(scale)	-1.1182	0.0606	-18.453	4.96e-76

Scale= 0.327

Log Normal distribution

```
Loglik(model)= -719.6 Loglik(intercept only)= -1014
Chisq= 588.81 on 115.5 degrees of freedom, p= 0
Number of Newton-Raphson Iterations: 8 38
n= 274
```

From the output of R program it is observed that the test for the regression parameters equal to zero is rejected with chi-square value 588.81 for 115.5 df and p-value is 0. The effect of frailty component is significant as compared to without the frailty term. Now the loglikelihood has been increased to -719.6 from -1008.1. The chi-square value is 577 for 1 df and p-value is 0. Following is the R program for Gaussian frailty model.

```
fit3=survreg(Surv(T,I)~PAge+PSex+hospital+g+frailty(pat,
dist='gauss'),data=bmt,dist='lognormal')
summary(fit3)
```

In the above R command patient is the frailty variable which has Gaussian distribution. The output of the R program is given below.

Call:

```
survreg(formula = Surv(T, I) ~ PAge + PSex + hospital + g
+ frailty(pat, dist = "gauss"), data = bmt,
dist = "lognormal")
```

	Value	Std. Error	z	p
(Intercept)	6.7119	0.6054	11.09	1.46e-28
PAge	-0.0169	0.0156	-1.09	2.77e-01
PSex	0.2363	0.2916	0.81	4.18e-01
hospital	0.1976	0.1279	1.55	1.22e-01
g	-0.2008	0.1880	-1.07	2.85e-01
Log(scale)	-1.1215	0.0606	-18.52	1.50e-76

Scale= 0.326

Log Normal distribution

Loglik(model)= -721.3 Loglik(intercept only)= -1014  
 Chisq= 585.28 on 122.4 degrees of freedom, p= 0  
 Number of Newton-Raphson Iterations: 5 27  
 n= 274

The frailty model for the nonparametric Cox's proportional hazards model, there is a problem of convergence while estimating the parameters for this data. The following are the R commands without output for this data.

```
fit4=coxph(Surv(T,I)~PAge+PSex+hospital+g+frailty(pat),data=bmt)
fit5=coxph(Surv(T,I)~PAge+PSex+hospital+g+frailty(pat,
dist='gauss'),data=bmt)
```

---

### 7.3 Analysis for Acute Leukemia Data

Consider the data set on remission duration from a clinical trial for acute leukemia given in Section 1.3. Here the variable of interest is time to relapse for placebo and 6-MP patients. The cens is an relapse indicator (0-censored, 1-relapse), status is the covariate indicating the remission status at randomization (1-partial, 2-complete). We analyze the data using ordinary Cox's proportional hazard (PH) model, gamma frailty, and Gaussian frailty models for the nonparametric Cox's PH model using R statistical package. The following are the R commands with output.

```
> fit1=coxph(Surv(time,cens)~status,data=remission)
> summary(fit1)
Call:
coxph(formula = Surv(time, cens) ~ status, data = remission)

n= 42

      coef  exp(coef)  se(coef)     z Pr(>|z|)
status -0.2070    0.8130    0.4143 -0.5    0.617

      exp(coef)  exp(-coef) lower .95 upper .95
status    0.813        1.23    0.361     1.831

Rsquare= 0.006  (max possible= 0.988 )
Likelihood ratio test= 0.24  on 1 df,   p=0.6231
Wald test            = 0.25  on 1 df,   p=0.6174
Score (logrank) test = 0.25  on 1 df,   p=0.6168
```

```

> fit1$loglik
[1] -93.18427 -93.06351

> fit2=coxph(Surv(time,cens)~status+frailty(pat),
  data=remission)
> summary(fit2)

Call:
coxph(formula = Surv(time, cens) ~ status + frailty(pat),
      data = remission)

n= 42
            coef    se(coef)   se2   Chisq DF p
status     -0.207 0.414    0.414 0.25   1 0.62
frailty(pat)                      0.00   0 0.95

            exp(coef) exp(-coef) lower .95 upper .95
status      0.813      1.23     0.361      1.83

Iterations: 6 outer, 23 Newton-Raphson
Variance of random effect= 5e-07   I-likelihood = -93.1
Degrees of freedom for terms= 1 0
Rsquare= 0.006 (max possible= 0.988 )
Likelihood ratio test= 0.24 on 1 df,   p=0.623
Wald test           = 0.25 on 1 df,   p=0.617
> fit2$loglik
[1] -93.18427 -93.06350

> fit3=coxph(Surv(time,cens)~status+frailty(pat,dist='gauss'),
  data=remission)
> summary(fit3)

Call:
coxph(formula = Surv(time, cens) ~ status + frailty(pat,
  dist = "gauss"), data = remission)

n= 42
            coef    se(coef)   se2   Chisq DF p
status     -0.208 0.415    0.414 0.25  1.00 0.62
frailty(pat, dist = "gaus" )                  0.02  0.03 0.64

            exp(coef) exp(-coef) lower .95 upper .95
status      0.812      1.23     0.36      1.83

Iterations: 10 outer, 34 Newton-Raphson
Variance of random effect= 0.00130
Degrees of freedom for terms= 1 0
Rsquare= 0.007 (max possible= 0.988 )
Likelihood ratio test= 0.29 on 1.03 df,   p=0.603
Wald test           = 0.25 on 1.03 df,   p=0.628

> fit3$loglik
[1] -93.18427 -93.03973

```

The log partial-likelihood when the vector  $\beta = 0$  is -93.18427. In the ordinary Cox's PH model, the final fit is -93.06351 and in the gamma frailty model, the variance of the random effect is estimated to be almost close to zero, corresponding 0 df. A likelihood ratio test for the frailty is twice the difference between the log partial-likelihood with the frailty term integrated out, shown as 'I-likelihood' in the printout, and the loglikelihood of a no-frailty model, or  $2(93.1 - 93.063) = 0.074$ . It has one degree of freedom with p-value = 0.785. The third fit is Gaussian frailty fit which chooses the estimate of the random effect using a REML criterion. REML method has slightly larger estimated frailty variance 0.0013 as compared to gamma frailty variance which is close to zero. The integrated likelihood is not printed for Gaussian frailty. Now we analyze the data with parametric baseline models. The following are the R commands with output.

```
> fit4=survreg(Surv(time,cens)~status,data=remission,
  dist='lognormal')
> fit4
Call:
survreg(formula = Surv(time, cens) ~ status, data = remission,
  dist = "lognormal")

Coefficients:
(Intercept)      status
      2.045857     0.238918

Scale= 1.109509

Loglik(model)= -115.2   Loglik(intercept only)= -115.4
    Chisq= 0.33 on 1 degrees of freedom, p= 0.56
n= 42

> fit5=survreg(Surv(time,cens)~status+frailty(pat),
  data=remission, dist='lognormal')
> fit5
Call:
survreg(formula = Surv(time, cens) ~ status + frailty(pat),
  data = remission, dist = "lognormal")

            coef  se(coef)  se2   Chisq DF   p
(Intercept) 2.046  0.747    0.745 7.51  1.00 0.0061
status       0.239  0.415    0.414 0.33  1.00 0.5700
frailty(pat)                         0.05  0.07 0.5600

Scale= 1.11

Iterations: 10 outer, 31 Newton-Raphson
Variance of random effect= 0.00263   I-likelihood = -55.7
Degrees of freedom for terms= 1.0 1.0 0.1 1.0
Likelihood ratio test=0.43 on 1.1 df, p=0.535 n= 42
```

```

> fit5$loglik
[1] -115.3930 -115.1776

> fit6=survreg(Surv(time,cens)~status+frailty(pat,
  dist='gauss'),data=remission,dist='lognormal')
> fit6
Call:
survreg(formula = Surv(time, cens) ~ status + frailty(pat,
  dist = "gauss"), data = remission, dist = "lognormal")

            coef  se(coef)   se2    Chisq DF   p
(Intercept) 2.046  0.747   0.746 7.51  1.00 0.0061
status        0.239  0.414   0.414  0.33  1.00 0.5600
frailty(pat, dist = "gaus")                      0.02  0.03 0.6300

Scale= 1.11

Iterations: 10 outer, 28 Newton-Raphson
Variance of random effect= 0.00130
Degrees of freedom for terms= 1 1 0 1
Likelihood ratio test=0.38 on 1 df, p=0.549 n= 42
> fit6$loglik
[1] -115.3930 -115.2024

```

It is observed from the above R output that the loglikelihood (intercept only) for the lognormal distribution is -115.4. For the other parametric distribution, the loglikelihood for Weibull is -116.4 and for loglogistic it is -115.4 and we take here lognormal distribution as the baseline model to analyze frailty models. The estimate of the variance for gamma frailty is 0.00263 and for Gaussian frailty is 0.0013. The loglikelihood of a no-frailty model (in gamma), or  $2(115.4 - 115.17) = 0.46$ . It has one degree of freedom with p-value = 0.497. Similarly, the loglikelihood of a no-frailty (in Gaussian), or  $2(115.4 - 115.20) = 0.4$  which is chi-square for one df with p-value = 0.527. So far there is clear cut way to say which frailty distribution is best.

## 7.4 Analysis for HLA Data

Consider the data set on kidney dialysis (HAL) given in Section 1.7. Here the variable of interest is time to graft rejection. The cens is an relapse indicator (1-censored, 0-graft rejected). HAL is the covariate indicating the HAL matching (1-good, 0-poor). We analyze the data using ordinary Cox's proportional hazard (PH) model, gamma frailty, and Gaussian frailty models for the nonparametric Cox's PH model using R statistical package. The following are the R commands with output.

```

> fit1=coxph(Surv(time,cens==1)~HAL,data=hal)
> summary(fit1)
Call:
coxph(formula = Surv(time, cens == 1) ~ HAL,
      data = hal)

n= 34

      coef  exp(coef)  se(coef)      z Pr(>|z|)
HAL -1.0812    0.3392   0.4257 -2.54   0.0111 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

      exp(coef)  exp(-coef) lower .95 upper .95
HAL     0.3392       2.948   0.1473   0.7812

Rsquare= 0.183  (max possible= 0.99 )
Likelihood ratio test= 6.89  on 1 df,  p=0.008688
Wald test            = 6.45  on 1 df,  p=0.01108
Score (logrank) test = 7  on 1 df,  p=0.008167

> fit1$loglik
[1] -78.58933 -75.14636

> fit2=coxph(Surv(time,cens==1)~HAL+frailty(pat),data=hal)
> summary(fit2)
Call:
coxph(formula = Surv(time, cens == 1) ~ HAL + frailty(pat),
      data = hal)

n= 34
      coef  se(coef)  se2   Chisq DF   p
HAL      -1.36  0.49    0.463  7.68 1.00 0.0056
frailty(pat)                      27.54  9.45 0.0015

      exp(coef)  exp(-coef) lower .95 upper .95
HAL      0.257       3.89    0.0984   0.672

Iterations: 6 outer, 55 Newton-Raphson
Variance of random effect= 0.925  I-likelihood = -73.4
Degrees of freedom for terms= 0.9 9.4
Rsquare= 0.671  (max possible= 0.99 )
Likelihood ratio test= 37.8  on 10.3 df,  p=5.34e-05
Wald test            = 7.68  on 10.3 df,  p=0.69

> fit3=coxph(Surv(time,cens==1)~HAL+frailty(pat,dist='gauss'),
  data=hal)
> summary(fit3)
Call:
coxph(formula = Surv(time, cens == 1) ~ HAL + frailty(pat,
      dist = "gauss"), data = hal)

n= 34
      coef  se(coef)  se2   Chisq DF   p

```

```

HAL          -1.52 0.515      0.475  8.76  1 0.00310
frailty(pat, dist = "gaus"                  34.17 11 0.00034

      exp(coef) exp(-coef) lower .95 upper .95
HAL      0.218       4.59     0.0795    0.598

Iterations: 8 outer, 47 Newton-Raphson
Variance of random effect= 1.79
Degrees of freedom for terms= 0.9 11.0
Rsquare= 0.723 (max possible= 0.99 )
Likelihood ratio test= 43.6 on 11.8 df,   p=1.57e-05
Wald test           = 8.76 on 11.8 df,   p=0.712

> fit4=coxph(Surv(time,cens==1)~HAL+frailty(pat,dist='gauss',
  method='aic'),data=hal)
> summary(fit4)
Call:
coxph(formula = Surv(time, cens == 1) ~ HAL + frailty(pat,
  dist = "gauss", method = "aic"), data = hal)

n= 34
              coef  se(coef)  se2   Chisq DF  p
HAL          -2.37 0.727   0.699 10.6  1.0 1.e-03
frailty(pat, dist = "gaus"                   75.1 15.1 5.9e-10

      exp(coef) exp(-coef) lower .95 upper .95
HAL      0.0936      10.7     0.0225    0.39

Iterations: 10 outer, 118 Newton-Raphson
Variance of random effect= 19.3
Degrees of freedom for terms= 0.9 15.1
Rsquare= 0.803 (max possible= 0.99 )
Likelihood ratio test= 55.3 on 16.1 df,   p=3.34e-06
Wald test           = 10.6 on 16.1 df,   p=0.836

```

The log partial-likelihood when the vector  $\beta = 0$  is -78.59. In the ordinary Cox's PH model, the final fit is -75.14 and in the gamma frailty model, the variance of the random effect is estimated to be 0.925, corresponding 9.45 df. HAL matching is significant covariate in all the models. A likelihood ratio test for the frailty is twice the difference between the log partial-likelihood with the frailty term integrated out, shown as 'L-likelihood' in the printout, and the loglikelihood of a no-frailty model, or  $2(75.14 - 73.4) = 3.48$ . It has one degree of freedom with p-value = 0.062. The frailty variable (patient) is not significantly related to time to graft rejection. The third fit is Gaussian frailty fit which chooses the estimate of the random effect using a REML criterion. The fourth fit is Gaussian frailty fit which uses Akaike's information criterion (AIC) to choose the variance of random effect, that is, it maximizes the value of (LR test - df), two quantities that are printed on the last line. The estimate of the variance of random effect is 19.3 as compared to REML method which is 1.79. The AIC method has a larger estimated frailty variance.

REML method has slightly larger estimated frailty variance 1.79 as compared to gamma frailty variance which is 0.925. The integrated likelihood is not printed for Gaussian frailty. Now we analyze the data with parametric baseline models. The following are the R commands with output.

```
> fit5=survreg(Surv(time,cens==1)~HAL,data=hal,dist='lognormal')
> summary(fit5)

Call:
survreg(formula = Surv(time, cens == 1) ~ HAL, data = hal,
       dist = "lognormal")
      Value Std. Error      z      p
(Intercept) 3.060     0.121 25.23 2.04e-140
HAL          0.556     0.183  3.03 2.43e-03
Log(scale)   -0.653     0.135 -4.83 1.36e-06

Scale= 0.521

Log Normal distribution
Loglik(model)= -119.1  Loglik(intercept only)= -123.1
                 Chisq= 8.07 on 1 degrees of freedom, p= 0.0045
Number of Newton-Raphson Iterations: 4
n= 34

> fit6=survreg(Surv(time,cens==1)~HAL+frailty(pat),data=hal,
+ dist='lognormal')
> summary(fit6)

Call:
survreg(formula = Surv(time, cens == 1) ~ HAL + frailty(pat),
       data = hal, dist = "lognormal")
      Value Std. Error      z      p
(Intercept) 3.120     0.176 17.76 1.35e-70
HAL          0.492     0.105  4.67 2.98e-06
Log(scale)   -1.351     0.133 -10.16 2.98e-24

Scale= 0.259

Log Normal distribution
Loglik(model)= -97  Loglik(intercept only)= -123.1
                 Chisq= 52.2 on 13.8 degrees of freedom, p= 2.2e-06
Number of Newton-Raphson Iterations: 10 39
n= 34

> fit7=survreg(Surv(time,cens==1)~HAL+frailty(pat,dist='gauss'),
+ data=hal,dist='lognormal')
> summary(fit7)

Call:
survreg(formula = Surv(time, cens == 1) ~ HAL + frailty(pat,
       dist = "gauss"), data = hal, dist = "lognormal")
      Value Std. Error      z      p
(Intercept) 3.039     0.121 25.02 4.02e-138
HAL          0.489     0.106  4.62 3.78e-06
```

```

Log(scale) -1.327      0.139 -9.52  1.70e-21
Scale= 0.265

Log Normal distribution
Loglik(model)= -98.2   Loglik(intercept only)= -123.1
                  Chisq= 49.81 on 12.3 degrees of freedom, p= 1.9e-06
Number of Newton-Raphson Iterations: 6 23
n= 34

```

It is observed from the above R output that the loglikelihood (intercept only) for the lognormal distribution is -123.1. For the other parametric distribution, the loglikelihood for Weibull is -128.1 and for loglogistic it is -123.5 and we take here lognormal distribution (which has the highest loglikelihood) as the baseline model to analyze frailty models. The covariate HAL matching is significant in all the parametric regression models. The estimates of the variance for gamma frailty and for Gaussian frailty are not printed in the R output. The loglikelihood of a no-frailty model (in gamma), or  $2(119.1 - 97.0) = 44.2$ . It has one degree of freedom with p-value = 0.0. Similarly, the loglikelihood of a no-frailty (in Gaussian), or  $2(119.1 - 98.2) = 41.8$  which is chi-square for one df with p-value = 0.0. From both frailty models, it is observed from the R output that frailty variable (patient) is very highly significantly related to the time to graft rejection.

---

## 7.5 Analysis for Kidney Infection Data

Consider the data set on kidney infection given in Section 1.5. Here the variable of interest is first or second recurrence time. The cens is an recurrence indicator (0-censored, 1-recurrence occurred). The covariates are age, sex (0-male, 1-female) and three disease indicator variables GN, AN, and PKD. Therneau and Grambsch (2000) analyzed this data using gamma and Gaussian frailty models. When all the three covariates are included in the gamma frailty, the variance of the random effect is essentially zero. When the disease variable is dropped out of the random effects model, the estimate of the variance of the random effect is 0.408. The LR test for no-frailty gives chi-square value 5.4 for one df with p-value of 0.02. Now we analyze the data with parametric baseline models and compare the results with nonparametric models. The following are the R commands with output.

```

> fit1=survreg(Surv(time,cens)^~age+gender+disease,data=kidney1,
  dist='lognormal')
> summary(fit1)

```

Call:

```
survreg(formula = Surv(time, cens) ~ age + gender + disease,
        data = kidney1, dist = "lognormal")
      Value Std. Error      z      p
(Intercept) 3.346170    0.4728  7.0771 1.47e-12
age         -0.000918    0.0114 -0.0805 9.36e-01
gender       1.476469    0.3172  4.6543 3.25e-06
diseaseAN   -0.620735    0.4097 -1.5153 1.30e-01
diseaseGN   -0.322634    0.4124 -0.7824 4.34e-01
diseasePKD   0.695716    0.5735  1.2132 2.25e-01
Log(scale)   0.104581    0.0914  1.1446 2.52e-01
```

Scale= 1.11

Log Normal distribution

```
Loglik(model)= -326  Loglik(intercept only)= -337.1
Chisq= 22.24 on 5 degrees of freedom, p= 0.00047
Number of Newton-Raphson Iterations: 4
n= 76
```

```
> fit2=survreg(Surv(time,cens)~age+gender+disease+frailty(id),
  data=kidney1,dist='lognormal')
> summary(fit2)
```

Call:

```
survreg(formula = Surv(time, cens) ~ age + gender + disease +
        frailty(id), data = kidney1, dist = "lognormal")
      Value Std. Error      z      p
(Intercept) 3.55693    0.5700  6.240 4.38e-10
age         -0.00331    0.0129 -0.257 7.97e-01
gender       1.48908    0.3763  3.957 7.59e-05
diseaseAN   -0.63131    0.4624 -1.365 1.72e-01
diseaseGN   -0.31810    0.4615 -0.689 4.91e-01
diseasePKD   0.80826    0.6549  1.234 2.17e-01
Log(scale)   -0.26487   0.1049 -2.526 1.15e-02
```

Scale= 0.767

Log Normal distribution

```
Loglik(model)= -302.5  Loglik(intercept only)= -337.1
Chisq= 69.2 on 20.2 degrees of freedom, p= 2.8e-07
Number of Newton-Raphson Iterations: 10 33
n= 76
```

It is observed from the above R output that the loglikelihood (intercept only) for the lognormal distribution is -337.1. For the other parametric distribution, the loglikelihood for Weibull is -338.6 and for loglogistic it is -339.1 and we take here lognormal distribution (which has the highest loglikelihood) as the baseline model to analyze frailty models. The covariate gender and disease AN are significant factors in all the lognormal regression models. We analyze the data using gamma frailty model only. The loglikelihood of a no-frailty model, or  $2(326.0 - 302.5) = 47$ . It has one degree of freedom with p-value

= 0.0. From the gamma frailty model, it is observed from the R output that frailty variable (patient) is very highly significantly related to the recurrence time.

---

## 7.6 Analysis of Litters of Rats

Consider the data set on litters of rats given in Section 1.6. Here the variable of interest is follow up time. The status is an indicator (0-tumor, 1-censored). The covariate is treatment indicator. Therneau and Grambsch (2000) analyzed this data using gamma and Gaussian frailty models. They observed the estimate of the variance of the random effect is highest when Gaussian frailty with AIC method is used. Now we analyze the data with gamma frailty and test for no-frailty in gamma frailty model under the Cox's PH model.

```
> fit1=coxph(Surv(time,status==1)~ind,data=rats)
> summary(fit1)
Call:
coxph(formula = Surv(time, status == 1) ~ ind, data = rats)

n= 150

      coef  exp(coef)  se(coef)      z Pr(>|z|)
ind  0.9047    2.4713   0.3175  2.849  0.00438 ** 
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

      exp(coef)  exp(-coef) lower .95 upper .95
ind     2.471      0.4046    1.326     4.605

Rsquare= 0.052  (max possible= 0.916 )
Likelihood ratio test= 7.98 on 1 df,  p=0.004741
Wald test            = 8.12 on 1 df,  p=0.004379
Score (logrank) test = 8.68 on 1 df,  p=0.003217

> fit1$loglik
[1] -185.6556 -181.6677

> fit2=coxph(Surv(time,status==1)~ind+frailty(litter),data=rats)
> summary(fit2)
Call:
coxph(formula = Surv(time, status == 1) ~ ind + frailty(litter),
      data = rats)

n= 150
      coef  se(coef)  se2   Chisq DF   p
ind     0.914  0.323   0.319  8.01  1.0 0.0046
frailty(litter)                      17.69 14.4 0.2400
```

```

exp(coef) exp(-coef) lower .95 upper .95
ind      2.50      0.401     1.32      4.7

Iterations: 6 outer, 24 Newton-Raphson
Variance of random effect= 0.499   I-likelihood = -180.8
Degrees of freedom for terms= 1.0 14.4
Rsquare= 0.222   (max possible= 0.916 )
Likelihood ratio test= 37.6 on 15.4 df,   p=0.00124
Wald test           = 8.01 on 15.4 df,   p=0.934

```

The log partial-likelihood without frailty in the first fit with covariate is -181.6 and log partial-likelihood with gamma frailty in the second fit with covariate is -180.8. The LR test for no-frailty with chi-square value 1.6 for one 1 df and p-value of 0.2. Next we analyze the data using parametric models.

```
> fit3=survreg(Surv(time,status==1)~ind,data=rats,dist='weibull')
> summary(fit3)
```

```
Call:
survreg(formula = Surv(time, status == 1) ~ ind, data = rats,
        dist = "weibull")
            Value Std. Error      z      p
(Intercept) 4.983     0.0833 59.81 0.00e+00
ind         -0.239     0.0891 -2.68 7.42e-03
Log(scale)  -1.333     0.1439 -9.26 2.01e-20
```

Scale= 0.264

Weibull distribution  
 Loglik(model)= -242.3 Loglik(intercept only)= -246.3  
 Chisq= 8 on 1 degrees of freedom, p= 0.0047  
 Number of Newton-Raphson Iterations: 7  
 n= 150

```
> fit6=survreg(Surv(time,status==1)~ind+frailty(litter),
  data=rats,dist='weibull')
> summary(fit6)
```

```
Call:
survreg(formula = Surv(time, status == 1) ~ ind
        + frailty(litter), data = rats, dist = "weibull")
            Value Std. Error      z      p
(Intercept) 5.021     0.1339 37.49 1.53e-307
ind         -0.200     0.0679 -2.95 3.16e-03
Log(scale)  -1.712     0.1262 -13.57 6.20e-42
```

Scale= 0.181

Weibull distribution  
 Loglik(model)= -204.3 Loglik(intercept only)= -246.3  
 Chisq= 83.88 on 37.9 degrees of freedom, p= 2.5e-05  
 Number of Newton-Raphson Iterations: 10 30

```

n= 150

> fit7=survreg(Surv(time,status==1)~ind+frailty(litter,
  dist='gauss'),data=rats,dist='weibull')
> summary(fit7)

Call:
survreg(formula = Surv(time, status == 1) ~ ind +
  frailty(litter, dist = "gauss"), data = rats,
  dist = "weibull")
      Value Std. Error      z      p
(Intercept) 4.871     0.0636  76.54 0.00e+00
ind        -0.182     0.0665 -2.74 6.13e-03
Log(scale) -1.649     0.1459 -11.30 1.25e-29

Scale= 0.192

Weibull distribution
Loglik(model)= -225.7 Loglik(intercept only)= -246.3
    Chisq= 41.06 on 16.4 degrees of freedom, p= 0.00068
Number of Newton-Raphson Iterations:  8 37
n= 150

```

It is observed from the above R output that the loglikelihood (intercept only) for the Weibull distribution is -246.3. For the other parametric distribution, the loglikelihood for lognormal is -247.1 and for loglogistic it is -246.5 and we take here Weibull distribution (which has the highest loglikelihood) as the baseline model to analyze frailty models. The covariate treatment indicator is significant factor in all the Weibull regression models. We analyze the data using gamma frailty and Gaussian frailty models. The LR of a no-frailty model (in gamma) is  $2(242.3 - 204.3) = 78$ . It has chi-square one degree of freedom with p-value = 0.0. The LR of a no-frailty (in Gaussian) is  $2(242.3 - 225.7) = 33.2$ . It has chi-square one degree of freedom with p-value = 0.0. From the Gaussian and gamma frailty models, it is observed from the R output that frailty variable (litter) is very highly significantly related to the recurrence time.

## 7.7 Analysis for Diabetic Retinopathy Data

Consider the data set on diabetic retinopathy given in Section 1.8. Here the variable of interest is follow up time. The status is an indicator (0-tumor, 1-censored). We consider two covariate in the analysis, one is laser type (treatment) and the second is type of diabetes (1-juvenile, 2-adult). Therneau and Grambsch (2000) analyzed this data using gamma and Gaussian frailty models. They observed the estimate of the variance of the random effect is highest when Gaussian frailty with AIC method is used. Now we analyze the data

with Gaussian and gamma frailties and test for no-frailty in the parametric regression model.

```
> fit1=survreg(Surv(time,status)~treat+adult, data=diabetics,
   dist='lognormal')
> summary(fit1)

Call:
survreg(formula = Surv(time, status) ~ treat + adult,
   data = diabetics, dist = "lognormal")
      Value Std. Error      z      p
(Intercept) 4.74357    0.5122  9.2608 2.03e-20
treat       -0.28471   0.2291 -1.2425 2.14e-01
adult        0.00771   0.2314  0.0333 9.73e-01
Log(scale)   0.66247   0.0637 10.4034 2.39e-25

Scale= 1.94

Log Normal distribution
Loglik(model)= -842.1  Loglik(intercept only)= -842.9
   Chisq= 1.55 on 2 degrees of freedom, p= 0.46
Number of Newton-Raphson Iterations: 3
n= 394

> fit2=survreg(Surv(time,status)~treat+adult+frailty(id),
   data=diabetics,dist='lognormal')
> summary(fit2)

Call:
survreg(formula = Surv(time, status) ~ treat + adult +
   frailty(id), data = diabetics, dist = "lognormal")
      Value Std. Error      z      p
(Intercept) 4.8461    0.5601  8.652 5.08e-18
treat       -0.2352   0.2511 -0.936 3.49e-01
adult        -0.1146   0.2524 -0.454 6.50e-01
Log(scale)  -0.0183   0.0578 -0.317 7.51e-01

Scale= 0.982

Log Normal distribution
Loglik(model)= -701.3  Loglik(intercept only)= -842.9
   Chisq= 283.18 on 126.1 degrees of freedom, p= 4.3e-14
Number of Newton-Raphson Iterations: 8 31
n= 394

> fit3=survreg(Surv(time,status)~treat+adult+frailty(id,
   dist='gauss'), data=diabetics,dist='lognormal')
> summary(fit3)

Call:
survreg(formula = Surv(time, status) ~ treat + adult +
   frailty(id, dist = "gauss"), data = diabetics,
   dist = "lognormal")
```

	Value	Std. Error	z	p
(Intercept)	4.3241	0.4454	9.709	2.76e-22
treat	-0.2531	0.2015	-1.256	2.09e-01
adult	-0.0102	0.2032	-0.050	9.60e-01
Log(scale)	0.0585	0.0663	0.882	3.78e-01

Scale= 1.06

Log Normal distribution

Loglik(model)= -726.9 Loglik(intercept only)= -842.9

Chisq= 231.98 on 106.1 degrees of freedom, p= 2.2e-11

Number of Newton-Raphson Iterations: 6 23

n= 394

It is observed from the above R output that the loglikelihood (intercept only) for the lognormal distribution is -842.9. For the other parametric distribution, the loglikelihood for weibull is -848 and for loglogistic it is -845.6 and we take here lognormal distribution (which has the highest loglikelihood) as the baseline model to analyze frailty models. No covariates are significant factors in all the lognormal regression models. We analyze the data using gamma frailty and Gaussian frailty models. The LR of a no-frailty model (in gamma) is  $2(842.1 - 701.3) = 281.6$ . It has chi-square one degree of freedom with p-value = 0.0. The LR of a no-frailty (in Gaussian) is  $2(842.1 - 726.9) = 230.4$ . It has chi-square one degree of freedom with p-value = 0.0. From the Gaussian and gamma frailty models, it is observed from the R output that frailty variable (patient) is very highly significantly related to the follow up time.



# Chapter 8

---

## *Tests of Hypotheses in Frailty Models*

8.1	Introduction .....	137
8.2	Tests for Gamma Frailty Based on Likelihood Ratio and Score Tests .....	138
8.2.1	The Model and the Main Results .....	139
8.2.2	Analysis of Diabetic Retinopathy .....	142
8.3	Logrank Tests for Testing $\beta = 0$ .....	143
8.3.1	Notations and Review .....	144
8.3.2	Parametric Tests for Uncensored Samples .....	146
8.3.3	Nonparametric Tests for Uncensored Samples .....	148
8.3.4	Effect of Censoring .....	151
8.3.5	Some Numerical Examples .....	154
8.4	Test for Heterogeneity in Kidney Infection Data .....	155
8.4.1	Models and Methods .....	157

---

### 8.1 Introduction

In this chapter, we discuss some well-known tests for frailty which are used more often and they have a lot of applications. In frailty models there is a clear need for inference on the heterogeneity parameter which measures the association between the survival outcomes in a specific cluster. The model specification of frailty models typically requires the heterogeneity parameter to be positive or, in case of homogeneity, to be zero. Therefore hypothesis testing problems for homogeneity against heterogeneity is described by a one-sided alternative hypothesis and, under the null hypothesis, the parameter is at the boundary of the parameter space which is  $(0, \infty)$ . In this chapter, three different test procedures are discussed. We first discuss tests for gamma frailty based on likelihood ratio and score tests and analyze diabetic retinopathy data. In Section 8.3, we discuss the logrank test for testing  $\beta = 0$  in parametric and nonparametric setup for uncensored and censored data and we give some numerical examples. In the last section, we discuss test for homogeneity, i.e., all frailties have common distribution and we analyze kidney infection data.

## 8.2 Tests for Gamma Frailty Based on Likelihood Ratio and Score Tests

Tests for the presence of heterogeneity in frailty models use an alternative hypothesis in which the heterogeneity parameter is subject to an inequality constraint. As a result the classical likelihood ratio asymptotic chi-square distribution theory is no longer valid.

Even though one-sided testing has a long history, going back to Chernoff (1954), the asymptotic distributional behavior of the test statistic has been studied few years ago. An important reference on this problem is Self and Liang (1987), who consider likelihood ratio tests for independent and identically distributed observations. Silvapulle and Silvapulle (1995) study one-sided score tests, see also Verbeke and Molenberghs (2003) in the context of mixed linear models. Vu and Zhou (1997) derive general theoretical results. More information is contained in the overview paper by Sen and Silvapulle (2002).

One of the important methodological questions is to provide information on the asymptotic distributional behavior of the likelihood ratio test for heterogeneity. To test for heterogeneity (to test within cluster correlation) we consider the following hypotheses testing problem. Assume that the random effect, present in the mixed proportional hazards model, has variance  $\theta$ . The relevant hypotheses testing problem is:

$$H_0 : \theta = 0 \quad \text{versus} \quad H_a : \theta > 0$$

From the theory of mixed effects models we know that the asymptotic distribution theory for the likelihood ratio statistic for such hypotheses testing problems does not follow the classical chi-square limit theory. The reason is that, under the null hypothesis, the parameter of interest is at the boundary of the parameter space (in the alternative hypothesis the heterogeneity parameter is subject to an inequality constraint). A consequence of this is that the classical conditions needed for the likelihood ratio theory are not satisfied. We therefore need to develop likelihood ratio theory under non-standard conditions. This phenomenon has been recognized in the literature on frailty models. Vaida and Xu (2000) wrote that for the likelihood ratio test a correction for the null distribution, which is no longer a chi-square distribution, is needed as discussed in similar set-ups by Stram and Lee (1994) and Self and Liang (1987) in the context of mixed effects models. Duchateau et al. (2002) simulate the limit distribution of the likelihood ratio test and conjecture that the simulated distribution is a 50:50 mixture of a  $\chi^2_0$  and a  $\chi^2_1$  distribution.

In this section we give the shared gamma frailty with a Weibull baseline hazard (which includes the exponential baseline hazard as a special case). Nguti et al. (2004) assume that the survival data are complete (censoring makes the formulas much more complicated) and that there are no covariates. A final simplification is that we assume each cluster contains two observations.

Bjarnason and Hougaard (2000) use this model to study the Fisher information. The idea behind the simplification is to fully understand a statistical property for a simple, but relevant, model. In this case Nguti et al. (2004) derived the asymptotic null distribution for the likelihood ratio test and the score test for heterogeneity.

### 8.2.1 The Model and the Main Results

We observe a set of  $n$  independent and identically distributed random vectors  $T_i = (T_{i1}, T_{i2})$ ,  $i = 1, 2, \dots, n$ . Each vector is considered as a cluster of size two. We assume that, conditional on the frailty variables  $Z_i$ , the lifetimes  $T_{i1}$  and  $T_{i2}$  are independent with (for  $Z_i = z$ ) a Weibull( $z\lambda, \gamma$ ) distribution, i.e., the conditional hazard is

$$h(t|z) = z\lambda\gamma t^{\gamma-1}$$

with  $\lambda > 0$  and  $\gamma > 0$  and where  $Z_i$  has the gamma density

$$f(z) = z^{\frac{1}{\theta}-1} \exp(-\frac{z}{\theta}) / [\Gamma(1/\theta)\theta^{\frac{1}{\theta}}].$$

The key idea is that within cluster dependence is caused by the frailty variables  $Z_1, \dots, Z_n$  representing unobserved common risk factors. The frailty variables are assumed to be independent. Also note that  $\text{Var}(Z_i) = \mu$ . Given  $Z_i = z$ , the conditional survival function of  $(T_{i1}, T_{i2})$  is

$$\begin{aligned} S(t_1, t_2|z) &= P(T_{i1} > t_1, T_{i2} > t_2 | Z = z) \\ &= \exp\{-z\lambda(t_1^\gamma + t_2^\gamma)\}. \end{aligned}$$

The unconditional survival function is

$$S(t_1, t_2) = E[\exp\{-z\lambda(t_1^\gamma + t_2^\gamma)\}].$$

The corresponding joint density is

$$f(t_1, t_2) = \frac{(1+\theta)\lambda^2\gamma^2t_1^{\gamma-1}t_2^{\gamma-1}}{[1+\theta\lambda(t_1^\gamma + t_2^\gamma)]^{\frac{1}{\theta}+2}}.$$

For  $\theta > 0$  (heterogeneity between clusters) the components of the vector  $(T_{i1}, T_{i2})$  are correlated (within cluster correlation). To quantify the within cluster dependence one can use Kendalls coefficient of concordance which, in terms of the joint density and survival function, is given by

$$4 \int_0^\infty \int_0^\infty f(t_1, t_2)S(t_1, t_2)dt_1 dt_2 - 1,$$

see Hougaard (2000, p. 132) and Bjarnason and Hougaard (2000). For this model Kendall's coefficient of concordance is  $\theta/(2+\theta)$  which is zero for  $\theta = 0$  (homogeneity between clusters). Moreover one can easily obtain that

$$\lim_{\theta \rightarrow 0} f(t_1, t_2) = (\lambda\gamma t_1^{\gamma-1} e^{-\lambda t_1^\gamma})(\lambda\gamma t_2^{\gamma-1} e^{-\lambda t_2^\gamma}),$$

i.e.,  $T_{i1}$  and  $T_{i2}$  are independent Weibull distributed random variables. The likelihood for the data is given by

$$\prod_{i=1}^n \frac{(1+\theta)\lambda^2\gamma^2 T_{i1}^{\gamma-1} T_{i2}^{\gamma-1}}{[1+\theta\lambda(T_{i1}^\gamma + T_{i2}^\gamma)]^{\frac{1}{\theta}+2}}$$

with corresponding loglikelihood

$$\begin{aligned} L = & \sum_{i=1}^n [2 \ln \lambda + 2 \ln \gamma + \ln(1+\theta) + (\gamma-1)(\ln T_{i1} + \ln T_{i2}) \\ & - (\frac{1}{\theta} + 2) \ln(1 + \theta\lambda(T_{i1}^\gamma + T_{i2}^\gamma))]. \end{aligned}$$

To test the within cluster dependence we consider the testing problem  $H_0 : \theta = 0$  against  $H_a : \theta > 0$ . For the further discussion it is convenient to work with the following transformed Weibull parameters:  $\eta = -\ln \lambda$  and  $\alpha = -\ln \gamma$ ; or  $\lambda = \exp(-\eta)$  and  $\gamma = \exp(-\alpha)$ . Further we use  $\tau$  as shorthand notation for the set of model parameters  $(\theta, \eta, \alpha)$  and  $\nu = (\eta, \alpha)$  for the set of nuisance parameters.

The corresponding likelihood ratio statistic is

$$\Lambda = 2(\max L(\tau) - \max L_0(\tau))$$

where  $\max L(\tau)$  and  $\max L_0(\tau)$  are maximum loglikelihoods under  $H_a$  and  $H_0$ , respectively. Under the null hypothesis the parameter vector of interest is at the boundary of the parameter space. Therefore the standard asymptotic distribution theory for likelihood ratio tests does not work. Instead we have the following theorem.

**Theorem 8.1** *The likelihood ratio statistic  $\Lambda$  for testing the heterogeneity has an asymptotic null distribution which is an equal mixture of a point mass at zero and a chi-square distribution with one degree of freedom, abbreviated as  $(\chi_0^2 + \chi_1^2)/2$ .*

**Remark:** This result is important since it has an immediate impact on how to determine (asymptotic) critical values and P-values for likelihood ratio test for heterogeneity. Erroneously relying on classical chi-square distribution theory will lead to P-values that are too big (or critical values that are too large), which means that not using the appropriate statistical inference leads to a conservative strategy in rejecting the null hypothesis of independence.

Under standard conditions, that is, when parameters constrained under the null hypothesis belong to the interior of the parameter space, it is well known that likelihood ratio, Wald, and score statistics have asymptotically the same distribution under the null hypothesis. Under inequality constraints in the alternative hypothesis, a score statistic is no longer uniquely defined,

see Silvapulle and Silvapulle (1995). Robertson, Wright, and Dykstra (1988, p. 320) propose a Wald and score statistic the latter of which has the disadvantage of requiring estimation of model parameters both under the null and alternative hypothesis. Silvapulle and Silvapulle (1995) propose a different score-type statistic which only requires estimation under the null hypothesis. Under mild regularity conditions, they obtain that under the null hypothesis asymptotically the score statistic follows the same mixture distribution as the likelihood ratio statistic.

The explicit expressions for the components of the score vector  $S_n(\tau) = (S_{n,\theta}(\tau), S_{n,\nu}(\tau))$  with  $S_{n,\theta}(\tau) = \partial L(\tau)/\partial\theta$  and  $S_{n,\nu}(\tau) = (\partial L(\tau)/\partial\eta, \partial L(\tau)/\partial\alpha)'$  are given in Nguti et al. (2004). Using a Taylor series expansion, Silvapulle and Silvapulle (1995) rewrite the likelihood ratio statistic as the difference of the minimum of two quadratic forms, of which the minimization of the first one under the null hypothesis can be performed exactly. We state the resulting score statistic in the following theorem given by Nguti et al. (2004). Let  $\hat{\nu}$  be the maximum likelihood estimator of the nuisance parameters under the null hypothesis and let  $S_{n,\theta}(0, \hat{\nu})$  denote the score vector evaluated at  $(0, \hat{\nu})$ .

**Theorem 8.2** (i) For a shared gamma frailty model with exponential baseline hazard a score statistic for testing the heterogeneity is given by

$$S_n = \frac{1}{3n^2} [S_{n,\theta}(0, \hat{\eta})]^2 - 3n \inf_{\geq 0} \left\{ \left( \frac{1}{3n^{3/2}} S_{n,\theta}(0, \hat{\eta}) - b \right)^2 \right\}.$$

(ii) For a Weibull distribution as the baseline hazard function a score statistic for testing the heterogeneity is given by

$$S_n = \frac{1}{3n^2} \frac{\pi^2}{\pi^2 - 4} [S_{n,\theta}(0, \hat{\nu})]^2 - 3n \left( 1 - \frac{4}{\pi^2} \right) \inf_{\geq 0} \left\{ \left( \frac{1}{3n^{3/2}} \frac{\pi^2}{\pi^2 - 4} S_{n,\theta}(0, \hat{\nu}) - b \right)^2 \right\}.$$

For both models the corresponding score statistic under the null hypothesis has asymptotic distribution  $(\chi_0^2 + \chi_1^2)/2$ .

Note that in the exponential baseline hazard model when  $S_{n,\theta}(0, \hat{\eta}) \geq 0$ , the expression of the score statistic simplifies to

$$S_n = \frac{1}{3n^2} [S_{n,\theta}(0, \hat{\eta})]^2.$$

A similar simplification holds for the Weibull baseline hazard model.

Wald-type test statistics for testing heterogeneity may be employed as well. Robertson, Wright, and Dykstra (1988) construct a Wald statistic for the situation where the alternative hypothesis is described by inequalities. Their test statistic requires estimation of model parameters under both the

null and alternative hypothesis. Sen and Silvapulle (2002) state a Wald test statistic as a difference of the minimum of two quadratic forms which has under the null hypothesis the same asymptotic distribution as the score and likelihood ratio statistic. For more details, see the review paper by Sen and Silvapulle (2002).

### 8.2.2 Analysis of Diabetic Retinopathy

Claeskens et al. (2005) rigorously establish the limiting distribution of the likelihood ratio statistic when covariate information is present. They also discussed the asymptotic distribution of the related score test. They use the diabetic retinopathy data (Huster et al., 1989) to test the heterogeneity by considering time to blindness in each eye of 197 patients with diabetic retinopathy. One eye of each patient is randomly selected for treatment and the other eye is observed without treatment. The data are bivariate right censored data with a treatment indicator as covariate. They assumed Weibull distribution as a baseline for this data with scale parameter  $\lambda$  and shape parameter  $\gamma$  and the positive stable frailty distribution with parameter  $\theta$ . They obtained the parameter estimates and their standard errors which are presented in Table 8.1.

Model	$\theta$	$\lambda$	$\gamma$	$\beta$
Full	0.712(0.145)	0.011(0.190)	0.888(0.006)	0.382(0.046)
Null	—	0.015(0.126)	0.799(0.005)	0.280(0.027)

TABLE 8.1: Diabetic retinopathy study: Parameter estimates (standard errors)

The log-likelihood values for null and full model are -846.499 and -841.272, respectively. This gives the observed value 10.454 of the likelihood ratio statistic. The corresponding p-value for the one-sided likelihood ratio test equals 0.0006. The presence of heterogeneity is confirmed by the construction of a profile likelihood based confidence interval for  $\theta$ . For a given value of  $\theta$ , we maximize  $L(\theta, \lambda, \gamma, \beta)$  with respect to  $\lambda$ ,  $\gamma$ , and  $\beta$ ; for a given  $\theta$  Claeskens et al. (2006) denote these maximizers as  $\lambda(\theta)$ ,  $\gamma(\theta)$ , and  $\beta(\theta)$ . They obtained the profile likelihood  $L(\theta, \lambda(\theta), \gamma(\theta), \beta(\theta))$  and follow the method similar as explained in Morgan (1992) to obtain the profile likelihood based confidence interval. They obtained 95% confidence limits for  $\theta$  are (0.32,1.20) which does not include zero, hence heterogeneity is present in the data.

### 8.3 Logrank Tests for Testing $\beta = 0$

The logrank test proposed from different viewpoints by Savage (1956), Mantel (1966), Peto and Peto (1972) among others is the optimal nonparametric test for testing the null hypothesis  $\beta = 0$  in the proportional hazards model

$$h_i(t) = \lim_{\Delta \rightarrow 0} \frac{1}{\Delta} Pr(T \leq t + \Delta \mid T \geq t) = \exp(\beta y_i) h_0(t) \quad (8.1)$$

for the influence of the explanatory variable  $y_i$  on the survival time  $T_i$  of individual  $i$ , ( $i = 1, 2, \dots, n$ ), when the baseline hazard function  $h_0(t)$  is totally unknown. The logrank test is easily derived as the score (Rao) test from Cox's proportional hazards model. It is well known also, see e.g., Cox (1959), Kalbfleisch (1974), Efron (1977), or Oakes (1977), that this test has full asymptotic efficiency locally at  $\beta = 0$  relative to the best parametric test under the model with  $h_0(t)$  known up to a scale factor,  $h_0(t) = \eta \tilde{h}_0(t)$ , say, with  $\tilde{h}_0(t)$  known. If the covariate is centered, so that  $\sum y_i = 0$ , the full asymptotic efficiency of the logrank test at  $\beta = 0$  extends to comparisons of the parametric test when the baseline hazard  $h_0(t)$  is fully known. The unit Fisher information about  $\beta$  at  $\beta = 0$  is  $I(0) = n^{-1} \sum y_i^2$ . The logrank test remains a valid test of the hypothesis  $\beta = 0$  under arbitrary patterns of independent right censorship of the  $T_i$ , but full asymptotic efficiency relative to the best parametric test will hold only if the explanatory variable  $y_i$  does not influence the potential censoring times, a condition which can be termed as "equal censoring".

Suppose however that there is a further covariate  $w$ , asymptotically orthogonal to  $y$  in the sense that as  $n \rightarrow \infty$ ,  $\lim n^{-1} \sum y_i g(w_i) \rightarrow 0$ , for any smooth function  $g(.)$  and which also influences  $T$ , so that the correct model is

$$h_i(t) = \exp(\beta y_i + \gamma w_i) b_0(t) \quad (8.2)$$

for a (different) unknown baseline hazard function  $b_0(t)$ . For example  $y$  could be the treatment assigned by randomization in a clinical trial and  $w$  could be a baseline covariate. If the values of  $w_i$  are observed, then the score test from Cox's model with  $\gamma$  estimated as a nuisance parameter is optimal for testing  $\beta = 0$  with  $\gamma$  unspecified, and is again locally fully efficient asymptotically in the absence of censoring, with unit Fisher information  $I(0) = n^{-1} \sum y_i^2$ , (Schoenfeld, 1981). We will call this test the adjusted logrank test. The primary focus is the testing of the hypothesis  $\beta = 0$  in (8.2) when the values of  $w$  are not observed, that is, of the performance of the logrank test and its parametric analog when there are omitted covariates. This problem has been considered explicitly by numerous authors, for example Lagakos and Schoenfeld (1984), Morgan (1986), and Struthers and Kalbfleisch (1986). Oakes and Jeong (2003) used the notation and some results from the theory of frailty

models (see e.g. Vaupel et al., 1979; Hougaard, 1984; Oakes, 1989) to connect this work with that of Gill (1980), Harrington and Fleming (1982) and others on weighted logrank tests. The key to this connection is that when  $w_i$  is unobserved, the model (8.2) induces a model for the conditional distribution of  $T_i$  given  $y_i$  that is usually not of proportional hazards form. The simple unweighted logrank test is then no longer optimal for this induced model.

Oakes and Jeong (2003) used the concept of the efficacy  $e(U)$  (cf Pitman, 1979, Chapter 5) of a test statistic  $U$  of the null hypothesis  $\beta = 0$ . Often  $U = U(0)$ , where  $U(\beta)$  is the derivative of a correctly or incorrectly specified log-likelihood function in  $\beta$ . Let  $\mu(\beta) = E_\beta U$  and assume that  $\mu(0) = 0$ . Note that expectations and variances are always taken under the null hypothesis  $\beta = 0$  unless indicated otherwise by subscript. Then

$$e(U) = \lim_{n \rightarrow \infty} \frac{1}{n} \frac{\{\mu'(0)\}^2}{\text{var}\{U(0)\}}.$$

When  $U$  is an asymptotically efficient test, for example when it is based on the score statistic  $U(\beta)$  from the true log-likelihood, the efficacy equals the unit Fisher information. The asymptotic relative efficiency (ARE) of a test statistic  $\bar{U}$  to another test  $U$  of the same hypothesis against the same class of alternatives is the ratio of their efficacies. It is well known that when  $U$  is an efficient test, this ratio is just  $\rho^2(\bar{U}, U)$ , the squared correlation between the two test statistics under the null hypothesis.

### 8.3.1 Notations and Review

Following Vaupel et al. (1979) we define a frailty to be a random unobserved multiplicative factor in the hazard function. We write  $z_i = \exp(\gamma w_i)$  for the unobserved factors and assume that the  $z_i$  are realizations of independent and identically distributed positive random variables  $Z_i$  with common Laplace transform  $p(s) = E \exp(-sZ)$ . This assumption implies independence of  $Z$  and  $y$ . Let  $\theta_i = \exp(\beta y_i)$  denote the hazard ratios of the observed explanatory variable, so that the null hypothesis  $\beta = 0$  gives  $\theta_i = 1$ . The model (8.2) becomes  $h_i(t | y_i, w_i) = z_i \theta_i b_0(t)$ , or in terms of the corresponding survivor functions  $S_i(t) = \exp(- \int h_i(\tilde{t}) d\tilde{t})$  and  $A_0(t) = \exp(-b_0(\tilde{t}) d\tilde{t})$ , with the integrals running from  $\tilde{t} = 0$  to  $\tilde{t} = t$ ,  $S_i(t | y_i, z_i) = \{A_0(t)\}^{\theta_i z_i}$ . Removing the conditioning on the value of  $Z_i$  gives

$$S_i(t) = E \exp[-Z_i \theta_i \{-\ln A_0(t)\}] = p(\theta_i B), \quad (8.3)$$

where

$$B = B(t) = \int_0^t b_0(\tilde{t}) d\tilde{t} = -\ln A_0(t), \quad (8.4)$$

is the integrated baseline hazard. In the sequel the argument  $t$  or  $t_i$  will generally be omitted for notational simplicity: functions with arguments  $B$  or  $B_i$

are understood to depend on  $t$  or  $t_i$  respectively via (8.4). Equation (8.3) is fundamental to the theory. It shows that the  $\theta_i$ , which enter (8.2) as hazard ratios, become scale change parameters on the integrated hazard scale  $B_i$ . Note that a scale factor common to all subjects may be absorbed into  $b_0(t)$ , so that frailty distributions are effectively defined only up to an overall scale factor. The density and hazard functions of  $T_i$  become respectively  $-\theta_i b_0(t)p'(\theta_i B)$  and

$$= -\theta_i b_0(t) \frac{p'(\theta_i B)}{p(\theta_i B)}. \quad (8.5)$$

The hazard function may also be written  $\theta_i b_0(t)E(Z_i | T_i \geq t)$ . The conditional expectation typically depends on  $\theta_i$  as well as  $t$ , so that for different  $\theta_i$  the resulting hazard functions are generally no longer proportional as functions of  $t$ . The gamma frailties of Vaupel et al. (1979) were used by Greenwood and Yule (1920) and are still the most popular. See e.g. Clayton (1978), Clayton and Cuzick (1985), Klein (1992), Nielsen et al. (1992), Andersen et al. (1993, Chapter 10), and the asymptotic theory of Murphy (1994, 1995). The gamma distribution with unit mean and variance  $k^{-1}$  has Laplace transform

$$p(s) = \left( \frac{1}{1 + s/k} \right)^k.$$

Note that the limit as  $k \rightarrow \infty$  of  $p(s)$  is,  $\exp(-s)$ , the Laplace transform of the degenerate distribution which places constant mass at unity. The induced model (8.3) does not satisfy the proportional hazards property. The ratio of the hazard functions at  $t$  for an individual with  $\theta_i = \theta > 1$  to an individual with  $\theta_i = 1$  decreases monotonically from  $\theta$  to unity as  $t \rightarrow \infty$ . Hougaard (1985, 1986) noted that the proportional hazards property does hold for the induced model if  $p(s) = \exp(-s^\alpha)$ . For  $0 < \alpha < 1$  this is the Laplace transform of the positive stable distribution. The proportionality factor in (8.5) is attenuated from  $\theta$  to  $\theta^\alpha$ , which is always closer to unity. Another possible frailty distribution is the inverse Gaussian. The Laplace transform of the inverse Gaussian frailty distribution with unit mean and variance  $(2\psi)^{-1}$  is

$$p(s) = \exp \left[ -s \{\psi(\psi + s)\}^{1/2} + 2\psi \right]. \quad (8.6)$$

The hazard ratio as a function of  $t$  decreases monotonically from  $\theta$  to  $\sqrt{\theta}$  as  $t \rightarrow \infty$ . For the two-point distribution, with  $p(s) = \alpha \exp(-s) + (1 - \alpha) \exp(-as)$ , in which  $Pr(Z = 1) = \alpha$ , and  $Pr(Z = a) = 1 - \alpha$  the hazard ratio is nonmonotone, approaching  $\theta$  for large  $t$  as well as at  $t = 0$ . Finally, the displaced Poisson distribution has  $p(s) = \exp\{-as + k(e^{-s} - 1)\}$ , so that  $Z = \alpha + J$ , where  $J$  follows a Poisson distribution with mean  $k$ . The hazard ratio from the induced model need not exceed unity for all  $t$ . Given any frailty distribution with Laplace transform  $p_0(s)$ , Oakes and Jeong (1998) defined a family of distributions using the tilting operation, defined for all  $\tau > 0$ ,  $p_\tau(s) = p_0(\tau + s)/p_0(\tau)$  (Hougaard, 1984). For example the inverse Gaussian

distributions are obtained by tilting the positive stable distribution with index  $\alpha = 1/2$ .

### 8.3.2 Parametric Tests for Uncensored Samples

Oakes and Jeong (1998) proposed parametric test in uncensored samples for testing  $\beta = 0$ . Suppose that the baseline hazard  $b_0(t)$  in (8.2) is known and that the value of the covariate  $w_i$  is observed for every individual as well as that of  $y_i$ . The log-likelihood in  $(\beta, \gamma)$  is

$$l(\beta) = \beta \sum y_i - \gamma \sum w_i + n \ln b_0(t_i) - \sum \exp(\beta y_i + \gamma w_i) B_i.$$

The scores in  $(\beta, \gamma)$  are

$$\frac{\partial l}{\partial \beta} = \sum y_i - \sum y_i \theta_i z_i B_i = U_1^{(p)}(\beta),$$

say, the superscript  $p$  denoting ‘parametric’,

$$\frac{\partial l}{\partial \gamma} = \sum w_i - \sum w_i \theta_i z_i B_i,$$

and the observed information matrix has elements

$$\begin{aligned} -\frac{\partial^2 l}{\partial \beta^2} &= \sum y_i^2 \theta_i z_i B_i, \\ -\frac{\partial^2 l}{\partial \beta \partial \gamma} &= \sum y_i w_i \theta_i z_i B_i, \\ -\frac{\partial^2 l}{\partial \gamma^2} &= \sum w_i^2 \theta_i z_i B_i. \end{aligned}$$

Viewed as random variables, each term  $\theta_i z_i B_i$  has a unit exponential distribution, so the elements of the unit Fisher information matrix are simply  $\mathcal{I}_{\beta\beta}^{(p)} = n^{-1} \sum y_i^2$ ,  $\mathcal{I}_{\beta\gamma}^{(p)} = n^{-1} \sum y_i w_i \rightarrow 0$ ,  $\mathcal{I}_{\gamma\gamma}^{(p)} = n^{-1} \sum w_i^2$ . Hence  $\beta$  and  $\gamma$  are asymptotically orthogonal, and the effective limiting unit Fisher information for  $\beta$  is  $\lim n^{-1} \text{var}[U_1^{(p)}(0)] = \lim \mathcal{I}_1^p(\beta) = \lim n^{-1} \sum y_i^2$  whether or not  $\gamma$  is known. Note that, provided the expectations are taken conditionally on the  $z_i$  as they should be, these results hold even if the moments of  $Z$  do not exist, as is the case for the positive stable frailty distribution.

Suppose now that  $B(t) = B(t, \alpha)$  is a regular parametrization of the cumulative baseline hazard. Then

$$n^{-1} \frac{\partial^2 l}{\partial \alpha \partial \beta} = -n^{-1} \sum \theta_i y_i z_i \frac{\partial B(t_i, \alpha)}{\partial \alpha}.$$

The expectation of this quantity when  $\beta = 0$  may be written

$$-n^{-1} \sum y_i E \left\{ \exp(\gamma w_i) \frac{\partial B(t_i, \alpha)}{\partial \alpha} \right\} \rightarrow 0,$$

since the terms in braces are free of  $y_i$  and  $\sum n^{-1} y_i g(w_i) \rightarrow 0$  for any function  $g(\cdot)$  by assumption. Hence the condition for orthogonality holds with respect to  $\alpha$  as well as  $\gamma$ : the limiting effective unit Fisher information for  $\beta$  is still  $n^{-1} \sum y_i^2$ . This is the condition for *adaptation* with respect to  $b_0(t, \alpha)$ .

Suppose now that the  $w_i$  are not observed, so that the  $Z_i$  are independent and identically distributed random variables with a common distribution that has Laplace transform  $p(s) = E \exp(-sZ)$ . The survivor function and density for subject  $i$  are respectively  $S_i(t, \beta) = p(\theta_i B)$  and  $f_i(t, \beta) = -\theta_i p'(\theta_i B) b_0(t)$ . The total log-likelihood is (recalling that  $\sum y_i = 0$ )

$$l(\beta) = - \sum \ln\{-p'(\theta_i \beta_i)\} + \sum \ln b_0(t_i).$$

The score in  $\beta$  is

$$U_2^{(p)}(\beta) = \frac{\partial l}{\partial \beta} = \sum \theta_i B_i y_i \frac{p''(\theta_i B_i)}{p'(\theta_i B_i)}.$$

Substitution of  $\beta = 0$  gives the test statistic

$$U_2^{(p)} = \sum B_i y_i \frac{p''(B_i)}{p'(B_i)}.$$

The unit Fisher information at  $\beta = 0$  is

$$\begin{aligned} \mathcal{I}_2^{(p)}(0) &= n^{-1} \operatorname{var}(U_2^{(p)}) = n^{-1} y_i^2 \operatorname{var} \left\{ \frac{B p''(B)}{p'(B)} \right\} \\ &= n^{-1} \sum y_i^2 \left[ \int \frac{\{B p''(B)\}^2}{-p'(B)} dB - 1 \right]. \end{aligned}$$

Since both  $U_2^{(p)}$  and  $U_1^{(p)}$  are score statistics from correctly specified log-likelihoods, the former based on less data than the latter, their relative efficacy  $e(2 : 1)$  is the ratio of the two information, namely

$$e(2 : 1) = \int \frac{\{B p''(B)\}^2}{-p'(B)} dB - 1.$$

Oakes and Jeong (1998) considered the score test obtained under the incorrect assumption that the  $w_i = 1$ , the proportional hazards model. They choose the parametrization of the baseline distribution so that it agrees with the true density when  $\beta = 0$ . The density of  $T_i$  under this model would be  $f_i(t) = -\theta p(B_i)^{\theta_i-1} p'(B_i) b_0(t)$ . The score statistic evaluated at  $\beta = 0$  is

$$U_3^{(p)} = \sum y_i \ln p(B_i).$$

Knowledge of the functional form of  $p(\cdot)$  is not needed to write down  $U_3^{(p)}$ , since it depends on the data only through the observable marginal survivor

function  $S(t) = p\{B(t)\}$ . It is easily shown that  $\text{var}(U_3^{(p)}) = \sum y_i^2$ . The ARE of this test relative to the test based on  $U_2^{(p)}$ , which is best for this model, is

$$e(3 : 2) = e(U_3^{(p)})/e(U_2^{(p)}) = \rho^2(U_2^{(p)}, U_3^{(p)}).$$

Integration by parts gives

$$\begin{aligned} \text{cov}(U_2^{(p)}, U_3^{(p)}) &= \sum y_i^2 \text{cov} \left\{ \ln p(B_i), \frac{B_i p''(B_i)}{p'(B_i)} \right\} \\ &= \sum y_i^2 \int \frac{B \{p'(B)\}^2}{p(B)} dB. \end{aligned}$$

Hence the ARE of the unadjusted test  $U_3^{(p)}$  relative to the test  $U_2^{(p)}$  that is the absence of knowledge of the covariate is

$$e(3 : 2) = \frac{[\int B \{p'(B)\}^2 / p(B) dB]^2}{\int \{B p''(B)\}^2 / \{-p'(B)\} dB - 1}.$$

Relative to the adjusted test  $U_1^{(p)}$ , the ARE of  $U_3^{(p)}$  is just

$$e(3 : 1) = e(3 : 2)e(2 : 1) = \left[ \int \frac{B \{p'(B)\}^2}{p(B)} dB \right]^2.$$

### 8.3.3 Nonparametric Tests for Uncensored Samples

Oakes and Jeong (1998) considered nonparametric tests of the hypothesis  $\beta = 0$ . As it happens, these have received far more attention in the literature than the corresponding parametric tests described in Section (8.5). It is convenient to begin with the simple unweighted logrank test statistic, denoted here by  $U_3^{(np)}$ . This is calculated by setting  $\beta = 0$  in the score statistic from Cox's (1972, 1975) partial likelihood,

$$U_3^{(np)}(\beta) = \sum \int \left\{ y_i - \frac{\sum y_j \theta_j V_j(t)}{\sum \theta_j V_j(t)} \right\} dN_i(t).$$

Here, in the usual counting process notation,  $N_i(t) = 1\{T_i \leq t\}$  and  $V_i(t) = 1\{T_i \geq t\}$  are the indicator functions for the events that subject  $i$  is observed to fail at or before time  $t$  and that subject  $i$  is still at risk of failure at time  $t$  respectively. (Under centering condition  $\sum y_i = 0$  the first term may be omitted but one can prefer to keep it for further work below). Setting  $\beta = 0$  gives  $\theta_i = 1$ , so that the logrank test essentially compares the value of the covariate  $y$  for each individual who fails with the average value among individuals who survive beyond that time. It is well known that this statistic has full asymptotic efficiency relative to  $U_3^{(p)}$  when the model is correctly specified, i.e., when there is no omitted covariate in Eq. (8.2). If the values

of the omitted covariates are known, then the same formula can be used, but with  $z_i$  replaced by  $\exp(\gamma w_i)$ , where the  $\gamma$  is estimated from a similar score statistic. Results of Efron (1977) and Oakes (1977) can be used to show that the resulting statistic  $U_1^{(np)}$  is fully efficient with respect to the corresponding parametric statistic  $U_1^{(p)}$  when model (8.2) is correct - this result does not require that  $\gamma = 0$  so long as  $\lim n^{-1} \sum y_i w_i = 0$ .

As noted above, when the covariates  $w_i$  are not observed, the induced model for  $T$  given  $y$  is no longer of the proportional hazards form. A wider class of test statistics can be defined by introducing a weight function  $\hat{l}(t)$  into the logrank statistic, giving

$$U_{\hat{l}}^{(np)}(0) = \sum \int \hat{l}(t) \left\{ y_i - \frac{\sum y_j V_j(t)}{\sum V_j(t)} \right\} dN_i(t). \quad (8.7)$$

As the notation implies,  $\hat{l}(t)$  may be data-dependent, but it must be predictable, that is, for each  $t$ , the value of  $\hat{l}(t)$  must be determined by the history of failures up to, but not including, time  $t$ . In practice  $\hat{l}(t)$  is usually taken to be some function of the overall estimated survival probability at time  $t$ , the reason for this will become clear below.

Gill (1980) considered the two-sample problem of testing equality of two survival distributions when hazard functions are respectively  $h_\theta(t)$  and  $h_1(t)$ , so that the null hypothesis corresponds to  $\theta = 1$ . He showed that a weighted logrank test of the form (8.7) is asymptotically fully efficient relative to the best nonparametric test statistic provided that the weight function  $\hat{l}(t)$  converges to a deterministic limit  $l(t)$  proportional to

$$\lambda(t) = \left. \frac{\partial \ln h_\theta(t)}{\partial \theta} \right|_{\theta=1}.$$

Thus the optimal limiting weight functions is proportional, as a function of  $t$ , to the sensitivity of the log-hazard function to changes in the value of the parameter  $\theta$ . His results extend easily to the case of a general variable  $y$ . The efficacy of the best test, which we naturally term  $U_2^{(np)}$ , is

$$\lim \frac{1}{n} \sum y_i^2 \int \lambda(t)^2 \{-p'(B)\} b_0(t) dt.$$

Under the model, with  $h_\theta(t) = -\partial \ln p(\theta B) / \partial t$ , this optimal weight function becomes

$$\lambda(t) = 1 + \frac{B p''(B)}{p'(B)} - \frac{B p'(B)}{p(B)}, \quad (8.8)$$

a form given, in a somewhat less transparent notation, by Morgan (1986), correcting work of Lagakos and Schoenfeld (1984). The same formula appears in Struthers and Kalbfleisch (1986). Since  $p(\beta)$  is a Laplace transform,  $\lambda(t) < 1$

for all  $t > 0$ . Also,  $\lambda(0) = 1$  provided  $\text{var}(W) < \infty$ . Interestingly,  $\lambda(t)$  may be negative for some values of  $t$  as is shown by the displaced Poisson distribution, for which (8.8) gives  $\lambda(t) = 1 - \lambda B(\lambda + \alpha e^B)^{-1}$ , which is negative over an interval if  $\alpha$  is not too large.

Morgan (1986) also gave a formula equivalent in our notation to

$$e(3 : 1) = \left[ \int \lambda(t) \{-p'(B)\} dB \right]^2$$

for the ARE of the simple logrank test to the adjusted logrank test.

In fact, integration by parts shows that

$$\{e(3 : 1)\}^{1/2} = \int \lambda(t) \{-p'(B)\} b_0(t) dt = \int B \frac{\{p'(B)\}^2}{p(B)} dB,$$

$$e(2 : 1) = \int \lambda(t)^2 \{-p'(B)\} b_0(t) dt = \int B \frac{\{B p''(B)\}^2}{-p'(B)} dB - 1,$$

showing that the efficacies of the three nonparametric test statistics are the same as those of the three parametric tests. This is not surprising since under the assumption of equal censoring each nonparametric test is asymptotically fully efficient relative to its parametric counterpart.

Expression of the weight function  $\lambda(t)$  in terms of the observable survivor function  $v = p(B)$ , yields a new function

$$\tilde{\lambda}(v) = 1 - \frac{q(v)q''(v)}{\{q'(v)\}^2} - \frac{q(v)}{vq'(v)}.$$

Here  $q(v)$  is the inverse function of  $p(B)$  which is always well-defined on  $0 < v \leq 1$  (Oakes, 1989).

The ARE formulas become

$$\begin{aligned} \{e(3 : 1)\}^{1/2} &= - \int_0^1 \frac{q(v)}{vq'(v)} dv = \int_0^1 \tilde{\lambda}(v) dv, \\ e(2 : 1) &= \int_0^1 \left\{ \frac{q(v)q''(v)}{q'(v)^2} \right\}^2 dv = \int_0^1 \tilde{\lambda}(v)^2 dv. \end{aligned}$$

There is a relation with the “Kendall tau” process which arises in bivariate frailty models (Genest et al., 1993). They introduce a distribution function  $K(v) = v - q(v)/q'(v)$ . We see that

$$\tilde{\lambda}(v) = \frac{1}{v} K(v) - K'(v).$$

For the tilted family defined in [Section 8.4](#), it is easily seen that

$$\lambda_\tau(t) = \frac{\tau}{\tau + B} + \frac{B\lambda_0(B + \tau)}{B + \tau}.$$

The inverse function is given by  $q_\tau(v) = q_0(\phi v) - \tau$ , where  $\phi = p_0(\tau)$ , so that

$$1 - \tilde{\lambda}_\tau = \frac{q_0(\phi v) - \tau}{q_0(\phi v)} \{1 - \tilde{\lambda}_0(v)\}.$$

Oakes (1994) considered interior and exterior power families of Laplace transforms based on a specified transform  $p_1(s)$ . The interior power family is defined for  $0 < \alpha < 1$  as  $p_\alpha(s) = p_1(s^\alpha)$ . If  $Z$  is a frailty variable with Laplace transform  $p_1(s)$  then  $p_\alpha(s)$  is the Laplace transform of the variable  $Z^{1/\alpha}Q$ , where  $Q$  is a positive stable variable with index  $\alpha$ . The exterior power family has  $p_k(s) = \{p_1(s)\}^k$ , if  $p_1(u)$  is infinitely divisible. For integral  $k$  this  $p_k(s)$  is the Laplace transform of the sum of  $k$  independent copies of  $Z$ .

He showed that, for the interior power family,  $K_\alpha(v) = (1 - \alpha)v + \alpha K_1(v)$  (in an obvious notation). Thus  $\tilde{\lambda}_\alpha(v) = \alpha \tilde{\lambda}_1(v)$ . By contrast, the exterior power family yields  $\lambda_k(t) = \lambda_1(t)$ , but  $\tilde{\lambda}_k(v) = \tilde{\lambda}_1(v^{1/k})$ .

### 8.3.4 Effect of Censoring

Oakes and Jeong (1998) considered the efficacies of the preceding parametric and nonparametric test statistics when there is censoring. They assume that the distribution of the potential censoring time does not depend on the  $y$  or  $w$ . Suppose then that the potential censoring times  $c_i$  are independent and identically distributed with common survivor function  $Pr(C > c) = G\{(B(c)\}$  and density  $g(B)b_0(c)$ . We set  $D_i = 1(T_i \leq c_i)$  for the indicator of the event that the  $i$ -th failure is observed.

Supposing first that the values of  $w_i$  are observed, the log-likelihood in  $(\beta, \gamma)$  is

$$l(\beta) = \beta \sum D_i y_i + \gamma \sum D_i w_i + \sum D_i \ln b_0(t_1) - \sum \exp(\beta y_i + \gamma w_i) B_i.$$

The scores in  $(\beta, \gamma)$  are

$$\frac{\partial l}{\partial \beta} = \sum y_i (D_i - \theta_i z_i B_i) = U_1^{(c,p)}(\beta),$$

say,

$$\frac{\partial l}{\partial \gamma} = \sum w_i (D_i - \theta_i z_i B_i),$$

and the observed information matrix has elements

$$\begin{aligned} -\frac{\partial^2 l}{\partial \beta^2} &= \sum y_i^2 \theta_i z_i B_i \\ -\frac{\partial^2 l}{\partial \beta \partial \gamma} &= -\sum y_i w_i \theta_i z_i B_i \\ -\frac{\partial^2 l}{\partial \gamma^2} &= -\sum w_i^2 \theta_i z_i B_i \end{aligned}$$

as before.

To calculate the Fisher information at the null hypothesis  $\beta = 0$  we must take expectations over the distribution of  $B_i$ , viewed as a random variable. Since  $y$  does not influence the latent censoring times  $c_i$ ,  $\pi_i = E(\theta_i z_i B_i)$  is a function of  $w_i$  alone. In fact  $\pi_i = \Pr(T_i \leq c_i; w_i)$  is just the probability calculated under the null hypothesis that the  $i$ -th observation will be uncensored. This will typically depend on  $w_i$ , even if the  $c_i$  do not, since individuals with higher frailties are more likely to be uncensored.

The limiting unit Fisher information matrix, calculated at  $\beta = 0$ , has elements

$$\mathcal{I}_{\beta\beta}^{(c,p)} = n^{-1} \sum \pi_i y_i^2, \quad \mathcal{I}_{\beta\gamma}^{(c,p)} = n^{-1} \sum \pi_i y_i w_i = 0, \quad \mathcal{I}_{\gamma\gamma}^{(c,p)} = n^{-1} \sum \pi_i w_i^2.$$

Again,  $\beta$  and  $\gamma$  are asymptotically orthogonal and the limiting unit effective Fisher information for  $\beta$  is  $\mathcal{I}_1^{(c,p)} = n^{-1} \sum \pi_i y_i^2$ , whether or not  $\gamma$  is known. Moreover, adaptation still holds with respect to any regular parametrization of the baseline distribution.

Under the model (8.3), the parametric log-likelihood is

$$\sum y_i [D_i \ln\{-\theta_i p'(\theta_i B_i)\} + (1 - D_i) \ln\{p(\theta_i B_i)\}].$$

The corresponding score statistic evaluated at  $\beta = 0$  is

$$U_2^{(c,p)} = \sum y_i \left[ D_i \left\{ 1 + \frac{B_i p''(B_i)}{p'(B_i)} \right\} + (1 - D_i) \left\{ \frac{B_i p'(B_i)}{p(B_i)} \right\} \right].$$

The limiting unit Fisher information is

$$\mathcal{I}_2^{(c,p)} = \lim n^{-1} \sum y_i^2 E \left[ D \left\{ 1 + \frac{B p''(B)}{p'(B)} \right\}^2 + (1 - D) \left\{ \frac{B p'(B)}{p(B)} \right\}^2 \right].$$

Since  $D^2 = D$ ,  $(1 - D)^2 = 1 - D$  and  $D(1 - D) = 0$  the expectation equals

$$E \left[ D \left\{ 1 + \frac{B p''(B)}{p(B)} \right\}^2 \right] + E \left[ (1 - D) \left\{ \frac{B p'(B)}{p(B)} \right\}^2 \right].$$

To calculate the first expectation, we must use the incomplete density of an observed failure time, namely  $G(B)\{-p'(B)\}b_0(t)dt$ , whereas the second term requires use of the incomplete density of an observed censoring time, namely  $p(B)g(B)b_0(t)dt$ . The resulting expression for the limiting unit effective Fisher information is

$$\begin{aligned} \mathcal{I}_2^{(c,p)} &= \lim n^{-1} \sum y_i^2 \left[ \int \left\{ 1 + \frac{B p''(B)}{p'(B)} \right\}^2 \{-p'(B)G(B)\}dB \right. \\ &\quad \left. + \int \left\{ \frac{B p'(B)}{p(B)} \right\}^2 p(B)g(B)dB \right]. \end{aligned}$$

For the nonparametric test under the same model the limiting unit effective Fisher information is

$$\lim n^{-1} \mathcal{I}_2^{(c,np)} = \sum y_i^2 \int \lambda(t)^2 \{-p'(B)G(B)\} dB,$$

where  $\lambda(t)$  is given above. A little manipulation shows that, as in the uncensored case, the two expressions are equal.

The parametric test statistic ignoring the  $z_i$  is

$$U_3^{(c,p)} = \sum y_i \{D_i + \ln p(B_i)\}.$$

We find after an easy calculation that

$$\begin{aligned} \text{cov}(U_2^{(c,p)}, U_3^{(c,p)}) &= \sum y_i^2 \text{cov} \left[ D_i + \ln p(B_i), D_i \left\{ 1 + \frac{B_i p''(B_i)}{p'(B_i)} \right\} \right. \\ &\quad \left. + (1 - D_i) \left\{ \frac{B_i p'(B_i)}{p(B_i)} \right\} \right] \\ &= \sum y_i^2 \int \lambda(t) \{-p'(B)G(B)\} dB. \end{aligned}$$

In many examples the optimal weight function  $\lambda(t)$  is monotone decreasing in  $t$ . An example where it is not the two-point distribution, for which  $\lambda(0) = \lambda(\infty) = 1$ . Although censoring always decreases the (absolute) efficacy of any of the tests, it is easily seen that when  $\lambda(t)$  is monotone equal censoring always increases the ARE's  $e(3 : 1)$  and  $e(2 : 1)$ . Censoring tends to reduce the impact of the omitted covariate by removing the late events.

When  $G(B) = p(B)^\eta$  we speak of Koziol-Green censoring (Koziol and Green, 1976). This model, also called proportional censoring, implies that the observed time  $X = \min(T, C)$  and failure indicator  $D = 1(T \leq C)$  are independent. Under Koziol-Green censoring we find that

$$\begin{aligned} \{e(3 : 1)^{1/2} &= (1 + \eta) \int_0^1 v^\eta \tilde{\lambda}(v) dv, \\ e(2 : 1) &= (1 + \eta) \int_0^1 v^\eta \{\tilde{\lambda}(v)\}^2 dv. \end{aligned}$$

For the exterior power family  $p_k(B) = \{p_1(B)\}^k$  defined in Section 8.6 we have, in an obvious notation,

$$e_{k,\eta}(3 : 1) = e_{k(1+\eta),0}(3 : 1), \quad e_{k,\eta}(2 : 1) = e_{k(1+\eta),0}(2 : 1), \quad (8.9)$$

allowing the ARE's in the censored case to be calculated from the corresponding formulas for the uncensored case.

### 8.3.5 Some Numerical Examples

We now give some numerical results, plotted against the coefficient  $\gamma$  of the omitted covariate. Given the variance of  $W$ ,  $\gamma$  is usually determined in terms of a free parameter of the distribution of the frailty  $Z = \exp(\gamma W)$ , by the relation  $\gamma^2 \text{var}(W) = \text{var}(\ln Z)$ . To allow comparability across different frailty distributions we set  $\text{var}(W) = 0.25$ , the same as for a simple Bernoulli variable. For the gamma distribution with index  $\kappa$ ,  $\text{var}(\ln Z) = \psi'(\kappa)$ , the trigamma function. So the exponential distribution, with  $\kappa = 1$ , corresponds to  $\gamma = \sqrt{(2\pi^2/3)} = 2.6$ . For the positive stable distribution  $\text{var}(\ln Z) = \pi^2(1 - \alpha^2)/(6\alpha^2)$  (Hougaard, 1986a, 1986b), and for the inverse Gaussian distribution,  $\text{var}(\ln Z)$  can be expressed in terms of the derivatives of  $K_\nu(y)$ , a modified Bessel function (Abramowitz and Stegun, 1965, Chapter 10). Hougaard (1991, p. 699), translated into Oakes and Jeong (1998) notation, gives the expression

$$\text{var}(\ln Z) = \frac{\partial^2}{\partial \nu^2} \ln K_\nu(2\psi) |_{\nu=-1/2}.$$

We now investigate the form of the optimal weighting functions  $\lambda(t)$  for the two-point, gamma, positive stable, and inverse Gaussian frailty distributions.

1. **Two-Point Distribution:** If the unobserved covariate takes values  $W = 0$  and  $W = 1$  each with probability one half, we find that

$$\lambda(t) = 1 + \frac{(2e^\gamma - 1 - e^{2\gamma})B \exp\{-(1 + e^\gamma)B\}}{[e^{-B} + \exp(\gamma - Be^\gamma)][e^{-B} + \exp(-Be^\gamma)]}$$

(Morgan, 1986).

2. **Gamma Distribution:** If  $Z$  has a gamma distribution with index  $\kappa$ , so that  $p(s) = (1 + s)^{-\kappa}$ , a simple example of an exterior power family, then

$$\lambda(t) = 1 - \frac{B(1 + \kappa)}{1 + B} + \frac{B\kappa}{1 + B} = \frac{1}{1 + B}$$

and  $\tilde{\lambda}\{S(t)\} = S(t)^\rho$ , where  $\rho = 1/\kappa$ . This implies that the optimal non-parametric test with gamma frailties are just the  $G^\rho$  tests of Harrington and Fleming (1982) with  $\rho = 1/\kappa$ . Although much of the derivation is algebraically similar to theirs, the interpretation of these tests in terms of allowance for an omitted covariate appears to be new.

3. **Positive Stable Distribution:** For positive stable frailties with  $p(s) = \exp(-s^\alpha)$  we find  $\lambda(t) = \alpha$  as expected. The optimal weights are constant, reflecting the preservation of the proportional hazards property under this model. The simple logrank test is optimal among weighted logrank tests.

4. **Inverse Gaussian Distribution:** For the inverse Gaussian, we find that

$$\lambda(t) = 1 - \frac{B}{2(\psi + B)},$$

so that

$$\tilde{\lambda}\{S(t)\} = \frac{1}{2} + \frac{2\psi^2}{[2\psi - \ln S(t)]^2}.$$

For gamma frailties, the ARE of the simple logrank test to the adjusted logrank test is

$$e(3 : 1) = \left(\frac{\kappa}{1 + \kappa}\right)^2,$$

Harrington and Fleming (1982) give the ARE of the simple logrank test to the best  $G^\rho$  test under the model for which this is optimal. Oakes and Jeong (1998) formula gives

$$e(3 : 2) = \left(\frac{\kappa}{1 + \kappa}\right)^2 / \left(\frac{\kappa}{2 + \kappa}\right) = \frac{1 + 2\rho}{(1 + \rho)^2},$$

in agreement with their results with  $\rho = 1/\kappa$ . For example if the omitted covariate corresponds to an exponentially distributed hazard ratio, the ARE of the simple logrank test to the test including the covariate is only 25%. This can be increased to 33% by using the optimally weighted test for this situation, which happens to be the Wilcoxon test.

For the positive stable distribution we have, easily,

$$e(3 : 1) = e(2 : 1) = \alpha^2.$$

For the inverse Gaussian distribution the results can be expressed in terms of the exponential integral  $E_1(2\psi) = \int_{2\psi}^{\infty} (e^{-y}/y) dy$  (Abramowitz and Stegun, 1965, Chapter 5). Oakes and Jeong (1998) found that

$$e(3 : 1) = [1/2 + \psi - 2\psi^2 e^{2\psi} E_1(2\psi)]^2,$$

and

$$e(2 : 1) = \frac{1}{4} + \frac{7\psi}{6} - \frac{\psi^2}{6} + \frac{\psi^3}{3} - 2\psi^2 \left(1 + \frac{\psi^2}{3}\right) e^{2\psi} E_1(2\psi).$$

The ratio  $e(3 : 2) = e(3 : 1)/e(2 : 1)$  always lies between 0.96 and unity, its value at  $\psi \rightarrow 0$  and  $\psi \rightarrow \infty$ . For two-point distribution the ARE must be evaluated numerically, but this requires only univariate integration.

## 8.4 Test for Heterogeneity in Kidney Infection Data

McGilchrist and Aisbett (1991) reported data for the recurrence times (in days) of infections of 38 kidney patients from insertion of a catheter until it had to be removed owing to infection. In this data there are three covariates:

age, gender (1 for males and 2 for females) and type of disease ( $0 = GN$ ,  $1 = AN$ ,  $2 = PKD$ , and  $3 = \text{other}$ ). The survival time of the  $j$ -th recurrence ( $j = 1, 2$ ) in the  $i$ -th patient ( $i = 1, \dots, 38$ ) is denoted by  $t_{ij}$ . Given the above covariates  $y_{ij}$  and the unobserved frailties  $z_i$ , the hazard function is modeled as

$$\lambda(t_{ij}|y_{ij}, z_i) = z_i \lambda_0(t_{ij}) \exp(\beta' y_{ij}), \quad (8.10)$$

where  $\lambda_0(\cdot)$  stands for the baseline hazard function and  $\beta$  is the regression parameter. This model expresses the assumption that the frailties of the 38 different individuals are independent with the same frailty being assumed for the two recurrent times of the individuals. This data set has been used by several authors to investigate random effect models for survival data. For example, see Hougaard (2000), Ibrahim et al. (2001), and Therneau and Grambsche (2000).

In most articles the frailties,  $z_i$  are assumed to be i.i.d. which, however, is hard to justify in this example. Lee and Lee (2003) claim for the heterogeneity of the frailty distribution was based on noticing the differences between the two recurrent times for the individuals. They found that some patients have very large deviations and others have relatively small ones, which suggests that the 38 patients might have come from a heterogeneous population and should be classified into several groups to make a more correct inference. Therefore, Lee and Lee (2003) claim that this phenomenon is a result of the heterogeneity of the frailty distribution.

In order to check the heterogeneity of the frailty distribution, Lee and Lee (2003) assume that the  $z_i$  in Eq. (8.10) follow a lognormal distribution with mean zero and variance  $\sigma^2$ , provided all the frailties follow a common distribution. The hypotheses is to be tested as

$$H_0 : \sigma_i^2 = \sigma^2, \quad i = 1, 2, \dots, n \quad \text{versus} \quad H_1 : \sigma_i^2 \neq \sigma^2 \quad \text{at least for one } i.$$

Lee and Lee (2003) rearrange the estimated frailties,  $\hat{z}_i$ 's, according to the magnitude of the differences between the two recurrent times, regardless of censoring. More precisely, letting  $\Delta_i = |t_{i1} - t_{i2}|$ , the interval size, Lee and Lee (2003) assume  $z_i$  to be associated with its interval size  $\Delta_i$ , viz.,  $z_i := z_i(\Delta_i)$ , and rearrange the  $\hat{z}_i$ 's in order corresponding to the magnitude of  $\Delta_i$ . The censoring times are treated like event times for simplicity. Then apply the cusum of squares test based on the ordered  $\hat{z}_i$ 's and construct the cusum of squares test statistic:

$$T_n := \max_{1 \leq k \leq n} D_k := \frac{1}{\sqrt{n} \hat{k}_n} \max_{1 \leq k \leq n} \left| \sum_{t=1}^k \hat{z}_t^2 - \frac{k}{n} \sum_{t=1}^n \hat{z}_t^2 \right| \quad (8.11)$$

where

$$\hat{k}_n^2 = n^{-1} \sum_{t=1}^n \hat{z}_t^4 - \left( n^{-1} \sum_{t=1}^n \hat{z}_t^2 \right)^2.$$

Then we reject  $H_0$  if  $T_n$  is large.

Employ the cusum of squares test since it is simple to construct and useful not only for testing for a variance change but also for detecting the location where the change point occurs. Also, it is a distribution-free test and can be used for any underlying distributions. As a matter of fact, it is well known that for any i.i.d. r.v.'s  $x_i$  with a fourth moment,

$$S_n := \frac{1}{\sqrt{n}\hat{\tau}_n} \max_{1 \leq k \leq n} \left| \sum_{t=1}^k x_t^2 - \frac{k}{n} \sum_{t=1}^n x_t^2 \right| \quad (8.12)$$

where

$$\hat{\tau}_n^2 = n^{-1} \sum_{t=1}^n x_t^4 - \left( n^{-1} \sum_{t=1}^n x_t^2 \right)^2,$$

converges in law to the sup of a standard Brownian bridge due to Donsker's invariance principle (cf. Billingsley (1968)). Thus the critical values given any significance level are easily obtained from an existing table. For example, for the significance level  $\alpha = 0.05$ , the associated critical value is 1.358. For the details regarding the cusum of squares test, see Inclán and Tiao (1994) and the articles cited therein.

After calculation of  $T_n$ , Lee and Lee (2003) observe that there is one significant variance change in the frailties at  $\Delta_i = 122$ , that is, the 33th patient ( $k = 25$ ). This implies that the frailty tends to have a different variation depending on the interval size. Based on this result, they consider a new model to fit the data. It turns out that the new model explains the kidney infection data better than the original method.

#### 8.4.1 Models and Methods

In this section we report the result of analyzing the kidney infection data taking into consideration the result presented in previous section. Consider the following two models.

**Model I:** This model refers to the frailty model in Eq. (8.10), where the  $z_i$  are assumed to follow a lognormal distribution with mean zero and variance  $\sigma^2$ .

**Model II:** This is the model proposed based on the result of the variance change test. Assume that the frailties are classified into two groups, say, Group 1 ( $\Delta_i < 122$ ) and Group 2 ( $\Delta_i \geq 122$ ) and assume

$$\ln y_i \sim \begin{cases} N(0, \sigma_1^2) & \Delta_i < 122 \\ N(0, \sigma_2^2) & \Delta_i \geq 122 \end{cases}.$$

The parameters of the two models are estimated based on the BLUP (best linear unbiased prediction) estimation using an iterative Newton - Raphson method (cf. McGilchrist and Aisbett (1991)). Then compute the variance of the frailties from REML (restricted maximum likelihood) equations which allow ties in survival times.

Now, fitting Model II to the kidney infection data. From Table 8.2, we can see that Group 1 has no frailty effects; the estimated variance of 0.0003 is almost negligible. On the other hand, Group 2 has a variety of frail degrees among the patients. The result indicates that in Group 1, covariates are sufficient to explain individual heterogeneity and there is no need to consider the frailties. However, in Group 2 we have to consider random effects to accommodate extra individual variation. Note that although both Models I and II yield almost the same  $\hat{\beta}$ , Model II produces a smaller variance for the regression parameter. This enables us to say that Model II outperforms Model I, and Model II provides us with a clue as to the phenomenon of the various interval sizes.

Age $\hat{\beta}_1$	Sex $\hat{\beta}_2$	Disease		
		Type = 0 $\hat{\beta}_3$	Type = 1 $\hat{\beta}_4$	Type = 2 $\hat{\beta}_5$
Model I: $\hat{\sigma}^2 = 0.483$				
0.0052 (0.0147)	-1.6790 (0.4582)	0.1807 (0.5354)	0.3936 (0.5368)	-1.1400 (0.8099)
Model II: $\hat{\sigma}_1^2 = 0.0003$ , $\hat{\sigma}_2^2 = 1.494$				
0.0046 (0.0116)	-1.8252 (0.3999)	0.4530 (0.4493)	0.2444 (0.4184)	-1.0230 (0.7800)

TABLE 8.2: Regression estimates (standard error) from Model I and Model II

Age $\hat{\beta}_1$	Sex $\hat{\beta}_2$	Disease		
		Type = 0 $\hat{\beta}_3$	Type = 1 $\hat{\beta}_4$	Type = 2 $\hat{\beta}_5$
Model I: $\hat{\sigma}^2 = 0.363$				
0.0017 (0.0142)	-1.9744 (0.4400)	0.2329 (0.5166)	0.5625 (0.5203)	0.1881 (0.8069)
Model II: $\hat{\sigma}_1^2 = 3.839 \times 10^{-8}$ , $\hat{\sigma}_2^2 = 0.491$				
0.0021 (0.0118)	-1.8608 (0.3835)	0.3230 (0.4468)	0.3828 (0.4265)	-1.1579 (0.6971)

TABLE 8.3: Regression estimates (standard error) with patient 21 removed

Therneau and Grambsche (2000) argued that ignoring patient 21, the Cox's proportional hazards model without considering the frailty is well fitted to the kidney infection data. This means that any nonzero estimated frailties are entirely due to this observation. Lee and Lee (2003) also analyzed the data deleting patient 21. [Table 8.3](#) shows the result of this analysis.

Note that in Model I the estimate of  $\beta_5$  has a positive sign, whereas it had a negative sign before (but the other estimates keep their signs). Also, its absolute value changes from 1.1400 to 0.1881, which is a somewhat large change. This is a proof that patient 21 influences greatly the estimation of  $\beta_5$  in Model I. Meanwhile, in Model II, the estimates are not much affected by patient 21. This means that there are no affirmative points to justify viewing this observation as an outlier at least in estimating the regression parameter based on Model II.

However, one can see that the removal of the observation affected greatly the variance of estimated frailties. As seen in Table 8.3, the variance in both models I and II decreased. This phenomenon is guessable since the REML solution is always sensitive to small changes in data. Actually, without patient 21, the variance of estimated frailties in Group 1 changes from  $3.0 \times 10^{-4}$  to  $3.8 \times 10^{-8}$ , which indicates that assuming the frailty in Group 1 is even more meaningless than in the case with patient 21 included. This variance reduction effect is also seen in Group 2, but the variance is still significant, which supports the existence of frailties and is opposed to the result based on Model I.



# Chapter 9

---

## *Shared Frailty in Bivariate Exponential and Weibull Models*

9.1	Introduction .....	161
9.2	Bivariate Exponential Distributions .....	162
9.2.1	Marshall-Olkin (M-O): .....	162
9.2.2	Block-Basu (B-B): .....	163
9.2.3	Freund .....	163
9.2.4	Proschan-Sullo (P-S): .....	164
9.3	Gamma Frailty in BVW Models .....	165
9.3.1	Weibull Extension of BVE of Gumbel .....	165
9.3.2	Weibull Extension of BVE of Marshall-Olkin .....	169
9.3.3	Weibull Extension of BVE of Block-Basu .....	170
9.3.4	Weibull Extension of BVE of Freund and Proschan-Sullo .....	171
9.4	Positive Stable Frailty in BVW Models .....	173
9.4.1	Weibull Extension of BVE of Gumbel .....	173
9.4.2	Weibull Extension of BVE of Marshall-Olkin .....	175
9.4.3	Weibull Extension of BVE of Block-Basu .....	175
9.4.4	Weibull Extension of BVE of Freund and Proschan-Sullo .....	176
9.5	Power Variance Function Frailty in BVW Models .....	176
9.5.1	Weibull Extension of BVE Models .....	176
9.6	Lognormal and Weibull Frailties in BVW Models .....	178
9.6.1	Estimation of Parameters .....	178
9.7	Compound Poisson Frailty in BVW Models .....	180
9.8	Compound Poisson (with Random Scale) Frailty in BVW Models .....	180
9.9	Estimation and Tests for Frailty under BVW Baseline .....	181

---

### 9.1 Introduction

Bivariate exponential distributions are widely employed while discussing the parametric survival analysis of two components of a system. The medical literature considers paired organs like kidneys, eyes, lungs, breasts, etc., of an individual as a two components system, which work under interdependency circumstances. There are several text books that address the univariate survival analysis from the frequentist and Bayesian perspective, these include Lawless (1982), Cox and Oakes (1984), Lee and Wang (2003), Kalbfleisch and Prentice (2002), and Ibrahim et al. (2001). Although these books contain versatile analysis of univariate survival data, very few of them attempted the analysis of paired data. In fact, there are many bivariate distribution (see Kotz et al., 2000) that can be employed for the analysis of paired data.

Hanagal and Kale (1991a, 1991b, 1992), Hanagal (1992b), Hanagal and Ahmadi (2008a, 2008b, 2009) and Ahmadi and Hanagal (2008) have done extensive work on BVE models. They obtained the MLE of parameters, Bayesian estimate of the parameters in BVE models and developed some tests for testing independence of two components and tests for symmetry of two components in BVE models based on MLEs. Pena and Gupta (1990) considered Bayesian estimation of the parameters of BVE of Marshall-Olkin in two situations namely, when the components are in a series and parallel system. They obtained posterior mode in BVE of Marshall-Olkin (1967) using gamma Dirichlet distribution as prior distribution. Ahmadi and Hanagal (2008, 2009) obtained Bayesian tests of hypotheses of independence, symmetry, and regression parameters in several BVE models. Hanagal and Ahmadi (2008a, 2008b, 2009a, 2009b) obtained Bayesian estimation of parameters using MCMC and EM algorithms in several BVE models.

---

## 9.2 Bivariate Exponential Distributions

We present the following four BVE models proposed by Marshall-Olkin (1967), Block-Basu (1974), Freund (1961), and Proschan-Sullo (1974). In the following discussion we consider  $(T_1, T_2)$  as failure times corresponding to two components.

### 9.2.1 Marshall-Olkin (M-O):

Marshall-Olkin (1967) proposed a BVE with two important properties viz, loss of memory property (LMP) and exponential marginals. The survival function of the life times  $(T_1, T_2)$  of the two components is

$$\begin{aligned} S_{M-O}(t_1, t_2) &= P(T_1 > t_1, T_2 > t_2) \\ &= \exp[-\lambda_1 t_1 - \lambda_2 t_2 - \lambda_3 \max(t_1, t_2)], \quad t_1, t_2 > 0 \end{aligned}$$

where  $\lambda_1, \lambda_2, \lambda_3 > 0$ .

The marginals of  $T_1$  and  $T_2$  are exponential with failure rates  $(\lambda_1 + \lambda_3)$  and  $(\lambda_2 + \lambda_3)$ , respectively. The parameter  $\lambda_3 = 0$  implies that two components are independent. Let  $(t_{1i}, t_{2i})$ ,  $i = 1, \dots, n$  be a random sample of size  $n$  from BVE of Marshall-Olkin (1967) and let  $n_1(n_2)$  be the number of observations with  $T_1 < T_2$  ( $T_1 > T_2$ ). The distribution of  $(n_1, n_2)$  is trinomial  $(n; \lambda_1/\lambda, \lambda_2/\lambda)$  where  $\lambda = \lambda_1 + \lambda_2 + \lambda_3$ . The pdf of  $(T_1, T_2)$  is given by

$$f_{M-O}(t_1, t_2) = \begin{cases} \lambda_1(\lambda_2 + \lambda_3) \exp(-\lambda_1 t_1 - (\lambda_2 + \lambda_3)t_2) & 0 < t_1 < t_2 \\ \lambda_2(\lambda_1 + \lambda_3) \exp(-\lambda_2 t_2 - (\lambda_1 + \lambda_3)t_1) & 0 < t_2 < t_1 \\ \lambda_3 \exp(-\lambda t) & 0 < t_1 = t_2 = t \end{cases} \quad (9.1)$$

This model is not absolutely continuous with respect to Lebesgue measure on  $R^2$ .

### 9.2.2 Block-Basu (B-B):

Block-Basu (1974) proposed an absolutely continuous BVE as an alternative model to Marshall-Olkin (1967), which has a singular component on the diagonal arising out of possible simultaneous failures. The survival function of BVE of Block-Basu (1974) is given by

$$S_{B-B}(t_1, t_2) = \frac{\lambda}{\lambda_1 + \lambda_2} e^{-\lambda_1 t_1 - \lambda_2 t_2 - \lambda_3 \max(t_1, t_2)} - \frac{\lambda_3}{\lambda_1 + \lambda_2} e^{-\lambda \max(t_1, t_2)}, \quad (9.2)$$

where  $\lambda_1, \lambda_2, \lambda_3 > 0$ ,  $\lambda = \lambda_1 + \lambda_2 + \lambda_3$ . Here the marginal distribution of  $T_1$  and  $T_2$  are not exponential, but a weighted combination of two exponentials with scales  $(\lambda_i + \lambda_3)$  and  $\lambda$ ,  $i=1,2$  with weights  $1 + \lambda_3 / (\lambda_1 + \lambda_2)$  and  $-\lambda_3 / (\lambda_1 + \lambda_2)$ . The pdf is given by

$$f_{B-B}(t_1, t_2) = \begin{cases} \frac{\lambda \lambda_1(\lambda_2 + \lambda_3)}{(\lambda_1 + \lambda_2)} \exp(-\lambda_1 t_1 - (\lambda_2 + \lambda_3)t_2) & 0 < t_1 < t_2 \\ \frac{\lambda \lambda_2(\lambda_1 + \lambda_3)}{(\lambda_1 + \lambda_2)} \exp(-\lambda_2 t_2 - (\lambda_1 + \lambda_3)t_1) & 0 < t_2 < t_1 \end{cases}$$

### 9.2.3 Freund

Freund (1961) proposed BVE as a model for failure time distribution of a system with life-times  $(T_1, T_2)$  operating in the following manner. Initially  $T_1$  and  $T_2$  are independent exponentials with failure rates  $\lambda_1$  and  $\lambda_2$  respectively,  $\lambda_1, \lambda_2 > 0$ . The interdependence of the components is such that failure of one component changes the failure rate of other from  $\lambda_1$  to  $\lambda_{11}$  ( $\lambda_2$  to  $\lambda_{22}$ ). The BVE of Freund (1961) with its joint pdf is given by

$$f_F(t_1, t_2) = \begin{cases} \lambda_1 \lambda_{22} \exp(-\lambda_{22} t_2 - (\lambda_1 + \lambda_2 - \lambda_{22})t_1) & 0 < t_1 < t_2 \\ \lambda_2 \lambda_{11} \exp(-\lambda_{11} t_1 - (\lambda_1 + \lambda_2 - \lambda_{11})t_2) & 0 < t_2 < t_1 \end{cases} \quad (9.3)$$

where  $\lambda_i > 0$ ,  $\lambda_{ii} > 0$ ,  $i=1,2$ .

This pdf has prominent property that  $P(T_1 = T_2) = 0$  indicating absolute continuity of the distribution.

### 9.2.4 Proschan-Sullo (P-S):

Proschan-Sullo (1974) proposed a BVE model which is a combination of both Marshall-Olkin and Freund models and the two component system operate in the following manner. Initially  $T_1$  and  $T_2$  follow BVE of Marshall-Olkin (1967) with parameters  $\lambda_1$ ,  $\lambda_2$  and  $\lambda_3$ . When a component fails the failure rate of other component changes the failure rate from  $\lambda_1 + \lambda_3$  to  $\lambda_{11} + \lambda_3$  ( $\lambda_2 + \lambda_3$  to  $\lambda_{22} + \lambda_3$ ). The BVE of Proschan-Sullo[P-S(1)] (1974) with its pdf is given by

$$f_{P-S(1)}(t_1, t_2) = \begin{cases} \lambda_1 \eta'_2 \exp(-\eta'_2 t_2 - (\lambda - \eta'_2) t_1) & 0 < t_1 < t_2 \\ \lambda_2 \eta'_1 \exp(-\eta'_1 t_1 - (\lambda - \eta'_1) t_2) & 0 < t_2 < t_1 \\ \lambda_3 \exp(-\lambda t) & 0 < t_1 = t_2 = t \end{cases} \quad (9.4)$$

where

$$\eta'_1 = \lambda_{11} + \lambda_3, \eta'_2 = \lambda_{22} + \lambda_3$$

provided  $\eta'_1 \neq \lambda \neq \eta'_2$ .

In this case if  $\lambda_3 = 0$  the Proschan-Sullo model reduce to Freund model. When  $\eta'_1 = \lambda = \eta'_2$ , the corresponding pdf of Proschan-Sullo [P-S(2)] is given by

$$f_{P-S(2)}(t_1, t_2) = \begin{cases} \lambda_1 \lambda \exp(-\lambda t_2) & 0 < t_1 < t_2 \\ \lambda_2 \lambda \exp(-\lambda t_1) & 0 < t_2 < t_1 \\ \lambda_3 \exp(-\lambda t) & 0 < t_1 = t_2 = t. \end{cases} \quad (9.5)$$

The Table 9.1 shows some properties of four BVE models.

Properties	Models			
	M-O	B-B	F	P-S
Loss Memory Property	✓	✓	✓	✓
Absolute Continuity		✓	✓	
Marginal exp	✓			
$(X - Y   X > Y) \sim \text{exp}$	✓	✓	✓	✓
$(Y - X   Y > X) \sim \text{exp}$	✓	✓	✓	✓

TABLE 9.1: Some properties of bivariate exponential distributions, ‘exp’ means exponential distribution

Univariate exponential has two important properties: loss of memory property (LMP) and absolute continuity and naturally we anticipate marginal exponential for BVE models. None of the BVE models satisfy all the three



properties. So we sacrifice one of the three properties and expect them to satisfy at least two properties. The LMP is more important property among all the three and we retain it and try to compromise between absolute continuity and marginal exponentiality. All the four models satisfy LMP. Marshall-Olkin model satisfies marginal exponential but not absolute continuity. Block-Basu model satisfies absolute continuity but not exponential marginals. Freund model has similar properties like Block-Basu model plus it has one interesting physical property, i.e., failure of one component changes the failure of rate of the life time of the other component. Proschan-Sullo model has neither absolute continuity nor exponential marginals but the important physical property like Freund model plus the occurrence of simultaneous failures of the two components.

Gumbel (1960) has proposed three types of BVE models. We mention only the Weibull extension of these BVE models. Some of the Weibull extension of these BVE models are given by Hanagal (2004, 2005a, 2005b, 2006a, 2006c). When the shape parameters of these bivariate Weibull (BVW) models is equal to one, all the BVW models reduce to BVE models. We take the simple power transformation of these BVE variates in order to get BVW models. Hanagal (2009a) has presented different frailty models under different BVW baseline models

### 9.3 Gamma Frailty in BVW Models

We present gamma frailty model under the different bivariate Weibull baseline models in this section.

#### 9.3.1 Weibull Extension of BVE of Gumbel

The Weibull extension of BVE-type I of Gumbel (1960) with survival function is given by

$$\begin{aligned} S(t_1, t_2) &= P[T_1 > t_1, T_2 > t_2] \\ &= e^{-\lambda_1 t_1^{c_1} - \lambda_2 t_2^{c_2} - \delta \lambda_1 \lambda_2 t_1^{c_1} t_2^{c_2}}, \quad t_1, t_2 > 0 \end{aligned} \quad (9.6)$$

where  $\lambda_1, \lambda_2, c_1, c_2 > 0, 0 \leq \delta \leq 1$ .

Here the marginal distribution of  $T_1$  and  $T_2$  are distributed as Weibull with scale parameters  $\lambda_1$  and  $\lambda_2$  and shape parameters  $c_1$  and  $c_2$ , respectively. The parameter  $\delta$  corresponds to the dependence parameter in BVW model.

Now the conditional survival function of BVW given the frailty ( $Z=z$ ) is given by

$$S(t_1, t_2 | z) = e^{-z(\lambda_1 t_1^{c_1} + \lambda_2 t_2^{c_2} + \delta \lambda_1 \lambda_2 t_1^{c_1} t_2^{c_2})} \quad (9.7)$$

where  $Z$  follows gamma distribution given in (5.2).

Integrating over  $Z$ , we get unconditional survival function and is given by

$$S_\alpha(t_1, t_2) = \left[ 1 + \frac{\lambda_1 t_1^{c_1} + \lambda_2 t_2^{c_2} + \delta \lambda_1 \lambda_2 t_1^{c_1} t_2^{c_2}}{\alpha} \right]^{-\alpha}. \quad (9.8)$$

This is the bivariate Burr distribution, generalizing the Pareto power distribution. The marginal distributions are Burr distribution with survival functions given by

$$S_\alpha(t_i) = \left[ 1 + \frac{\lambda_i t_i^{c_i}}{\alpha} \right]^{-\alpha}, \quad i = 1, 2. \quad (9.9)$$

The unconditional bivariate survival function in (9.8) has two types of dependencies, one parameter ( $\delta$ ) is due to dependence parameter and the other parameter ( $\alpha$ ) is due to frailty. These two parameters are identifiable.

As we have seen in the univariate Weibull regression, the scale parameter of the univariate Weibull distribution can be expressed in terms of regression coefficients. If  $\lambda$  is the scale parameter of the Weibull distribution, then  $\lambda = e^{-\beta' y}$  or  $\lambda = e^{\beta' y}$  where  $\beta$  is the vector of regression parameters and  $y$  is the vector of regressors or covariates. In the similar manner, the scale parameters  $\lambda_1$  and  $\lambda_2$  can be expressed in terms of regression parameters in the following way

$$\begin{aligned} \lambda_1 &= e^{-(\beta_0' y_0 + \beta_1' y_1)} \\ \lambda_2 &= e^{-(\beta_0' y_0 + \beta_2' y_2)} \end{aligned} \quad (9.10)$$

where

$$\begin{aligned} \beta_0' &= (\beta_{01}, \dots, \beta_{0p}), & -\infty < \beta_0 < \infty \\ \beta_1' &= (\beta_{11}, \dots, \beta_{1q}), & -\infty < \beta_1 < \infty \\ \beta_2' &= (\beta_{21}, \dots, \beta_{2q}), & -\infty < \beta_2 < \infty \\ y_0' &= (y_{01}, \dots, y_{0p}) \\ y_1' &= (y_{11}, \dots, y_{1q}) \\ y_2' &= (y_{21}, \dots, y_{2q}). \end{aligned}$$

The exponent terms in the above expressions, we can take either positive or negative but in either case  $\lambda_1, \lambda_2 > 0$ .  $\beta_0' y_0$  corresponds to the term containing identical covariates for both components.  $\beta_1' y_1$  corresponds to the term containing covariates for first component and  $\beta_2' y_2$  corresponds to the term containing covariates for second component.

Now the survival function of BVW in terms of covariates is given by

$$S_\alpha(t_1, t_2) = \left\{ 1 + \frac{t_1^{c_1} e^{-(\beta'_0 y_0 + \beta'_1 y_1)} + t_2^{c_2} e^{-(\beta'_0 y_0 + \beta'_2 y_2)}}{\alpha} \right. \\ \left. + \frac{\delta t_1^{c_1} t_2^{c_2} e^{-(2\beta'_0 y_0 + \beta'_1 y_1 + \beta'_2 y_2)}}{\alpha} \right\}^{-\alpha}. \quad (9.11)$$

The another Weibull extension of BVE-type II of Gumbel (1960) with survival function is given by

$$S(t_1, t_2) = S_1(t_1)S_2(t_2)[1 + \delta(1 - S_1(t_1))(1 - S_2(t_2))], \quad t_1, t_2 > 0 \quad (9.12)$$

where  $-1 \leq \delta \leq +1$ .

The above survival function is derived from the given two independent marginal distributions. This relation is true for any bivariate distribution. Suppose  $T_i$ ,  $i = 1, 2$  follows Weibull( $\lambda_i, c_i$ ) distribution then BVW of  $(T_1, T_2)$  with survival function is given by

$$S(t_1, t_2) = (1 + \delta)e^{-\lambda_1 t_1^{c_1} - \lambda_2 t_2^{c_2}} - \delta e^{-2\lambda_1 t_1^{c_1} - \lambda_2 t_2^{c_2}} \\ - \delta e^{-\lambda_1 t_1^{c_1} - 2\lambda_2 t_2^{c_2}} + \delta e^{-2\lambda_1 t_1^{c_1} - 2\lambda_2 t_2^{c_2}}, \quad (9.13)$$

where  $t_1, t_2 > 0$ ,  $\lambda_1, \lambda_2, c_1, c_2 > 0$ ,  $-1 \leq \delta \leq +1$ .

The above survival function is the weighted combination of four bivariate Weibull distributions with weights  $(1 + \delta), -\delta, -\delta, \delta$ . Here the marginal distribution of  $T_1$  and  $T_2$  are distributed as Weibull with scale parameters  $\lambda_1$  and  $\lambda_2$  and shape parameters  $c_1$  and  $c_2$ , respectively. The parameter  $\delta$  corresponds to the dependence parameter in BVW model. When  $\delta = 0$ ,  $T_1$  and  $T_2$  are independent.

Now the conditional survival distribution of the above BVW given the frailty ( $Z = z$ ) is given by the weighted combination of the four conditional survival functions given the frailty, that is,

$$S(t_1, t_2 | z) = (1 + \delta)e^{-z(\lambda_1 t_1^{c_1} + \lambda_2 t_2^{c_2})} - \delta e^{-z(2\lambda_1 t_1^{c_1} + \lambda_2 t_2^{c_2})} \\ - \delta e^{-z(\lambda_1 t_1^{c_1} + 2\lambda_2 t_2^{c_2})} + \delta e^{-z(2\lambda_1 t_1^{c_1} + 2\lambda_2 t_2^{c_2})}, \quad (9.14)$$

where  $Z$  follows gamma distribution given in (5.2).

Integrating over  $Z$ , we get unconditional survival function and is given by

$$S_\alpha(t_1, t_2) = (1 + \delta) \left[ 1 + \frac{\lambda_1 t_1^{c_1} + \lambda_2 t_2^{c_2}}{\alpha} \right]^{-\alpha} - \delta \left[ 1 + \frac{2\lambda_1 t_1^{c_1} + \lambda_2 t_2^{c_2}}{\alpha} \right]^{-\alpha} \\ - \delta \left[ 1 + \frac{\lambda_1 t_1^{c_1} + 2\lambda_2 t_2^{c_2}}{\alpha} \right]^{-\alpha} + \delta \left[ 1 + \frac{2\lambda_1 t_1^{c_1} + 2\lambda_2 t_2^{c_2}}{\alpha} \right]^{-\alpha}. \quad (9.15)$$

The above bivariate survival function is the weighted combination of the four bivariate Burr distributions with weights  $(1 + \delta), -\delta, -\delta, \delta$ . The marginal distributions are Burr distribution with survival functions given by

$$S_\alpha(t_i) = \left[ 1 + \frac{\lambda_i t_i^{c_i}}{\alpha} \right]^{-\alpha}, \quad i = 1, 2. \quad (9.16)$$

The parameters  $\lambda_1$  and  $\lambda_2$  here also expressed in terms of the regression parameters and covariates as we did in the Eq. (9.10).

The another Weibull extension of BVE-type III of Gumbel (1960) with survival function is given by

$$S(t_1, t_2) = e^{-(\lambda_1 t_1^{c_1/\delta} + \lambda_2 t_2^{c_2/\delta})^\delta}, \quad t_1, t_2 > 0 \quad (9.17)$$

where  $\lambda_1, \lambda_2, c_1, c_2 > 0, 0 \leq \delta \leq 1$ .

Here the marginal distribution of  $T_1$  and  $T_2$  are distributed as Weibull with scale parameters  $\lambda_1$  and  $\lambda_2$  and shape parameters  $c_1$  and  $c_2$ , respectively. The parameter  $\delta$  corresponds to the dependence parameter in BVW model.

Now the conditional survival distribution of the above BVW given the frailty ( $Z = z$ ) is given by

$$S(t_1, t_2 | z) = e^{-z(\lambda_1 t_1^{\frac{c_1}{\delta}} + \lambda_2 t_2^{\frac{c_2}{\delta}})^\delta}, \quad t_1, t_2 > 0 \quad (9.18)$$

where  $Z$  follows gamma distribution given in (5.2).

Integrating over  $Z$ , we get unconditional survival function and is given by

$$S_\alpha(t_1, t_2) = \left[ 1 + \frac{(\lambda_1 t_1^{\frac{c_1}{\delta}} + \lambda_2 t_2^{\frac{c_2}{\delta}})^\delta}{\alpha} \right]^{-\alpha}. \quad (9.19)$$

The marginal distributions are Burr distribution with survival functions given by

$$S_\alpha(t_i) = \left[ 1 + \frac{\lambda_i t_i^{c_i}}{\alpha} \right]^{-\alpha}, \quad i = 1, 2. \quad (9.20)$$

The parameters  $\lambda_1$  and  $\lambda_2$  here also expressed in terms of the regression parameters as we did in the Eq. (9.10). The survival function in terms of covariates is given by

$$S_\alpha(t_1, t_2) = \left[ 1 + \frac{(t_1^{c_1/\delta} e^{-(\beta'_0 z_0 + \beta'_1 z_1)} + t_2^{c_2/\delta} e^{-(\beta'_0 z_0 + \beta'_2 z_2)})^\delta}{\alpha} \right]^{-\alpha}. \quad (9.21)$$

### 9.3.2 Weibull Extension of BVE of Marshall-Olkin

Hanagal (2004) proposed BVW with survival function given by

$$S(t_1, t_2) = e^{-\lambda_1 t_1^c - \lambda_2 t_2^c - \lambda_3 t_{(2)}^c} \quad (9.22)$$

where  $t_{(2)} = \max(t_1, t_2)$ ,  $\lambda_1, \lambda_2, \lambda_3, c > 0$ .

The BVW model has important properties. Both marginal distribution of BVW are univariate Weibull. More specifically speaking, we can state that

$$(T_1, T_2) \sim BVW(\lambda_1, \lambda_2, \lambda_3, c) \implies \begin{aligned} T_1 &\sim \text{Weibull}((\lambda_1 + \lambda_3), c) \\ T_2 &\sim \text{Weibull}((\lambda_2 + \lambda_3), c) \end{aligned}$$

and  $T_{(1)} \equiv \min(T_1, T_2) \sim \text{Weibull}((\lambda_1 + \lambda_2 + \lambda_3), c)$ , where  $(\lambda_1 + \lambda_3)$ ,  $(\lambda_2 + \lambda_3)$ , and  $(\lambda_1 + \lambda_2 + \lambda_3)$  are the scale parameters of  $T_1$ ,  $T_2$  and  $T_{(1)}$  respectively and  $c$  is the common shape parameter. This BVW model is not absolutely continuous with respect to Lebesgue measure in  $R^2$ . It has singularity on the diagonal  $T_1 = T_2$ . Here the parameter  $\lambda_3$  corresponds to the dependence between the two variables  $(T_1, T_2)$  and  $\lambda_3 = 0$  implies  $T_1$  and  $T_2$  are independent. The probability of simultaneous failures, that is,  $P[T_1 = T_2]$  is  $\lambda_3/(\lambda_1 + \lambda_2 + \lambda_3)$  which is the correlation between  $T_1$  and  $T_2$ . The probability in the remaining two regions are  $P[T_1 < T_2] = \lambda_1/(\lambda_1 + \lambda_2 + \lambda_3)$  and  $P[T_1 > T_2] = \lambda_2/(\lambda_1 + \lambda_2 + \lambda_3)$ .

Now the conditional survival function of BVW given the frailty ( $Z = z$ ) is given by

$$S(t_1, t_2 | z) = e^{-z(\lambda_1 t_1^c + \lambda_2 t_2^c + \lambda_3 t_{(2)}^c)} \quad (9.23)$$

where  $Z$  follows gamma distribution given in (5.2).

Integrating over  $Z$ , we get unconditional survival function and is given by

$$S_\alpha(t_1, t_2) = \left[ 1 + \frac{\lambda_1 t_1^c + \lambda_2 t_2^c + \lambda_3 t_{(2)}^c}{\alpha} \right]^{-\alpha}. \quad (9.24)$$

This is the bivariate Burr distribution, generalizing the Pareto power distribution. The marginal distributions and  $T_{(1)} = \min(T_1, T_2)$  are Burr distribution with survival functions given by

$$\begin{aligned} S_\alpha(t_i) &= \left[ 1 + \frac{(\lambda_i + \lambda_3)t_i^c}{\alpha} \right]^{-\alpha}, \quad i = 1, 2 \\ S_\alpha(t_{(1)}) &= \left[ 1 + \frac{(\lambda_1 + \lambda_2 + \lambda_3)t_{(1)}^c}{\alpha} \right]^{-\alpha}. \end{aligned} \quad (9.25)$$

The unconditional bivariate survival function in (9.24) has two types of dependencies, one is due to simultaneous failures and the other is due to frailty. Now we develop a regression model for the two component system. The scale

parameters  $\lambda_1$  and  $\lambda_2$  can be expressed in terms of regression parameters and covariates as in Eq. (9.10) and the parameter  $\lambda_3$  expressed as follows.

$$\lambda_3 = e^{-(\beta'_0 y_0 + \beta'_1 y_1 + \beta'_2 y_2)}. \quad (9.26)$$

**Note:** From here onwards,  $\lambda_1$ ,  $\lambda_2$ , and  $\lambda_3$  are expressed in terms of covariates from Eq. (9.10) and (9.26) when one wants to express survival function in the form of regression model.

### 9.3.3 Weibull Extension of BVE of Block-Basu

Block-Basu (1974) proposed an absolutely continuous BVE as an alternative model to Marshall-Olkin (1967) which has a singular component on the diagonal arising out of possible simultaneous failures. Marshall-Olkin is not absolutely continuous with respect to Lebesque measure in  $R^2$ , there are some situations when this model is not appropriate. The survival function of BVE of Block-Basu (1974) is given by

$$S(x_1, x_2) = \frac{\lambda}{\lambda_1 + \lambda_2} e^{-\lambda_1 x_1 - \lambda_2 x_2 - \lambda_3 x_{(2)}} - \frac{\lambda_3}{\lambda_1 + \lambda_2} e^{-\lambda x_{(2)}} \quad (9.27)$$

where  $x_{(2)} = \max(x_1, x_2)$ ,  $\lambda_1, \lambda_2, \lambda_3 > 0$ ,  $\lambda = \lambda_1 + \lambda_2 + \lambda_3$ .

Taking the transformation  $T_1 = X_1^c$  and  $T_2 = X_2^c$ ,  $c > 0$ , we get BVW and its survival function is given by

$$\begin{aligned} S(t_1, t_2) &= P[T_1 > t_1, T_2 > t_2] \\ &= \frac{\lambda}{\lambda_1 + \lambda_2} e^{-\lambda_1 t_1^c - \lambda_2 t_2^c - \lambda_3 t_{(2)}^c} - \frac{\lambda_3}{\lambda_1 + \lambda_2} e^{-\lambda t_{(2)}^c} \end{aligned} \quad (9.28)$$

where  $t_{(2)} = \max(t_1, t_2)$ ,

Here the marginal distribution of  $T_1$  and  $T_2$  are now not Weibull but a weighted combination of two Weibull with scales  $(\lambda_i + \lambda_3)$  and  $\lambda$ ,  $i=1,2$  with weights  $1 + \lambda_3 / (\lambda_1 + \lambda_2)$  and  $-\lambda_3 / (\lambda_1 + \lambda_2)$ .

The above model turns out to be the absolutely continuous part of BVW of Hanagal (1996, 2004) and also a special case of BVW of Hanagal (2006a). The independence of two components corresponds to  $\lambda_3 = 0$ .

Now the conditional survival function of BVW given the frailty ( $Z = z$ ) is given by the weighted combination of the conditional survival functions given the frailty, that is,

$$S(t_1, t_2 | z) = \frac{\lambda}{\lambda_1 + \lambda_2} e^{-z(\lambda_1 t_1^c + \lambda_2 t_2^c + \lambda_3 t_{(2)}^c)} - \frac{\lambda_3}{\lambda_1 + \lambda_2} e^{-z(\lambda t_{(2)}^c)} \quad (9.29)$$

where  $Z$  follows gamma distribution given in (5.2).

Integrating over  $Z$ , we get unconditional survival function and is given by

$$S_\alpha(t_1, t_2) = \frac{\lambda}{\lambda_1 + \lambda_2} \left[ 1 + \frac{\lambda_1 t_1^c + \lambda_2 t_2^c + \lambda_3 t_{(2)}^c}{\alpha} \right]^{-\alpha} - \frac{\lambda_3}{\lambda_1 + \lambda_2} \left[ 1 + \frac{\lambda t_{(2)}^c}{\alpha} \right]^{-\alpha}. \quad (9.30)$$

The above bivariate survival function is weighted combination of bivariate Burr distribution and univariate Burr distribution with weights  $[1 + \lambda_3/(\lambda_1 + \lambda_2)]$  and  $-\lambda_3/(\lambda_1 + \lambda_2)$ , respectively. The marginal distributions are weighted combinations of two Burr distributions with weights  $[1 + \lambda_3/(\lambda_1 + \lambda_2)]$  and  $-\lambda_3/(\lambda_1 + \lambda_2)$ . The distribution of  $T_{(1)} = \min(T_1, T_2)$  is Burr distribution with survival function given by

$$S_\alpha(t_{(1)}) = \left[ 1 + \frac{(\lambda_1 + \lambda_2 + \lambda_3)t_{(1)}^c}{\alpha} \right]^{-\alpha}. \quad (9.31)$$

The unconditional bivariate survival function in (9.30) has two types of dependencies, one is due to the dependence parameter  $\lambda_3$  and the other is due to frailty.

### 9.3.4 Weibull Extension of BVE of Freund and Proschan-Sullo

Freund (1961) proposed BVE as a model for failure time distribution of a system with life-times  $(X_1, X_2)$  operating in the following manner. Initially  $X_1$  and  $X_2$  are independent exponential with failure rates  $\lambda_1$  and  $\lambda_2$  respectively,  $\lambda_1, \lambda_2, > 0$ . The interdependence of the components is such that failure of a component changes the failure rate of other component from  $\lambda_1$  to  $\lambda_{11}$  ( $\lambda_2$  to  $\lambda_{22}$ ). The BVE of Freund (1961) with its joint pdf is given by

$$f(x_1, x_2) = \begin{cases} \lambda_1 \lambda_{22} e^{-(\lambda_{22} x_2 - (\lambda_1 + \lambda_2 - \lambda_{22}) x_1)}, & 0 < x_1 < x_2 < \infty \\ \lambda_2 \lambda_{11} e^{-(\lambda_{11} x_1 - (\lambda_1 + \lambda_2 - \lambda_{11}) x_2)}, & 0 < x_2 < x_1 < \infty \end{cases} \quad (9.32)$$

where  $\lambda_1, \lambda_2, \lambda_{11}, \lambda_{22} > 0$ .

Proschan-Sullo (1974) proposed BVE which is the combination of both Marshall-Olkin and Freund models and the two component system operate in the following manner. Initially  $X_1$  and  $X_2$  follow BVE of Marshall-Olkin (1967). When a component fails the failure of a component changes the failure rate of other component from  $\lambda_1 + \lambda_3$  to  $\lambda_{11} + \lambda_3$  ( $\lambda_2 + \lambda_3$  to  $\lambda_{22} + \lambda_3$ ). The BVE of Proschan-Sullo (1974) with its pdf is given by

$$f(x_1, x_2) = \begin{cases} \lambda_1 (\lambda_{22} + \lambda_3) e^{-(\lambda_{22} + \lambda_3) x_2 - (\lambda_1 + \lambda_2 - \lambda_{22}) x_1}, & 0 < x_1 < x_2 < \infty \\ \lambda_2 (\lambda_{11} + \lambda_3) e^{-(\lambda_{11} + \lambda_3) x_1 - (\lambda_1 + \lambda_2 - \lambda_{11}) x_2}, & 0 < x_2 < x_1 < \infty \\ \lambda_3 e^{-(\lambda_1 + \lambda_2 + \lambda_3) x}, & 0 < x_1 = x_2 = x < \infty \end{cases} \quad (9.33)$$

where  $\lambda_1, \lambda_2, \lambda_3, \lambda_{11}, \lambda_{22} > 0$ .

Taking transformation  $T_1 = X_1^c$  and  $T_2 = X_2^c$ ,  $c > 0$  we get bivariate Weibull model (BVW) which was introduced by Hanagal (2005) with pdf given by

$$f(t_1, t_2) = \begin{cases} \lambda_1(\lambda_{22} + \lambda_3)c^2(t_1 t_2)^{c-1} e^{-(\lambda_{22} + \lambda_3)t_2^c - (\lambda_1 + \lambda_2 - \lambda_{22})t_1^c}, & 0 < t_1 < t_2 < \infty \\ \lambda_2(\lambda_{11} + \lambda_3)c^2(t_1 t_2)^{c-1} e^{-(\lambda_{11} + \lambda_3)t_1^c - (\lambda_1 + \lambda_2 - \lambda_{11})t_2^c}, & 0 < t_2 < t_1 < \infty \\ \lambda_3 e^{-(\lambda_1 + \lambda_2 + \lambda_3)t^c}, & 0 < t_1 = t_2 = t < \infty. \end{cases} \quad (9.34)$$

Re-parameterize  $\lambda_{11} = \phi_1 \lambda_1$ ,  $\lambda_{22} = \phi_2 \lambda_2$  and rewrite the above pdf as

$$f(t_1, t_2) = \begin{cases} \lambda_1(\lambda_2 \phi_2 + \lambda_3)c^2(t_1 t_2)^{c-1} e^{-(\lambda_2 \phi_2 + \lambda_3)t_2^c - (\lambda_1 + \lambda_2 - \lambda_2 \phi_2)t_1^c}, & 0 < t_1 < t_2 < \infty \\ \lambda_2(\lambda_1 \phi_1 + \lambda_3)c^2(t_1 t_2)^{c-1} e^{-(\lambda_1 \phi_1 + \lambda_3)t_1^c - (\lambda_1 + \lambda_2 - \lambda_1 \phi_1)t_2^c}, & 0 < t_2 < t_1 < \infty \\ \lambda_3 e^{-(\lambda_1 + \lambda_2 + \lambda_3)t^c}, & 0 < t_1 = t_2 = t < \infty. \end{cases} \quad (9.35)$$

As we know in BVE of Proschan-Sullo (1974), the marginals are weighted combinations of two exponential distributions. Here in the BVW also, the marginals are weighted combinations of two Weibull distributions with same weights. The  $\min(T_1, T_2)$  is Weibull with scale parameter  $(\lambda_1 + \lambda_2 + \lambda_3)$  and shape parameter  $c$ . When  $\lambda_3 = 0$ , the BVW in Eq. (9.35) reduces to BVW of Hanagal (2004) and when  $\phi_1 = \phi_2 = 1$ , it reduces to BVW of Hanagal (2006a). When  $\phi_i = 1$ ,  $i = 1, 2$  and  $\lambda_3 = 0$  then  $T_1$  and  $T_2$  are independent. The probabilities in the three regions are given by

$P[T_1 < T_2] = \lambda_1/(\lambda_1 + \lambda_2 + \lambda_3)$ ,  $P[T_1 > T_2] = \lambda_2/(\lambda_1 + \lambda_2 + \lambda_3)$  and  $P[T_1 = T_2] = \lambda_3/(\lambda_1 + \lambda_2 + \lambda_3)$ .

The survival function of this BVW is given by

$$S(t_1, t_2) = \begin{cases} \frac{\lambda_2(1-\phi_2)e^{-(\lambda_1+\lambda_2+\lambda_3)t_2^c}}{\lambda_1+\lambda_2(1-\phi_2)} \\ + \frac{\lambda_1 e^{-(\lambda_1+\lambda_2(1-\phi_2))t_1^c - (\lambda_2 \phi_2 + \lambda_3)t_2^c}}{\lambda_1+\lambda_2(1-\phi_2)t_1^c}, & 0 < t_1 \leq t_2 \\ \frac{\lambda_1(1-\phi_1)e^{-(\lambda_1+\lambda_2+\lambda_3)t_1^c}}{\lambda_2+\lambda_1(1-\phi_1)} \\ + \frac{\lambda_2 e^{-(\lambda_2+\lambda_1(1-\phi_1))t_2^c - (\lambda_1 \phi_1 + \lambda_3)t_1^c}}{\lambda_2+\lambda_1(1-\phi_1)}, & 0 < t_2 \leq t_1. \end{cases} \quad (9.36)$$

Now the conditional survival function of BVW given the frailty ( $Z = z$ ) is given by the weighted combination of the conditional survival functions given

the frailty, that is,

$$S(t_1, t_2 | z) = \begin{cases} \frac{\lambda_2(1-\phi_2)e^{-z(\lambda_1+\lambda_2+\lambda_3)t_2^c}}{\lambda_1+\lambda_2(1-\phi_2)} \\ + \frac{\lambda_1e^{-z[(\lambda_1+\lambda_2(1-\phi_2))t_1^c+(\lambda_2\phi_2+\lambda_3)t_2^c]}}{\lambda_1+\lambda_2(1-\phi_2)}, & 0 < t_1 \leq t_2 \\ \frac{\lambda_1(1-\phi_1)e^{-z(\lambda_1+\lambda_2+\lambda_3)t_1^c}}{\lambda_2+\lambda_1(1-\phi_1)} \\ + \frac{\lambda_2e^{-z[(\lambda_2+\lambda_1(1-\phi_1))t_2^c+(\lambda_1\phi_1+\lambda_3)t_1^c]}}{\lambda_2+\lambda_1(1-\phi_1)}, & 0 < t_2 \leq t_1 \end{cases} \quad (9.37)$$

where  $Z$  follows gamma distribution given in (5.2).

Integrating over  $Z$ , we get unconditional survival function and is given by

$$S_\alpha(t_1, t_2) = \begin{cases} \frac{\lambda_2(1-\phi_2)\left[1+\frac{(\lambda_1+\lambda_2+\lambda_3)t_2^c}{\alpha}\right]^{-\alpha}}{\lambda_1+\lambda_2(1-\phi_2)} \\ + \frac{\lambda_1\left[1+\frac{(\lambda_1+\lambda_2(1-\phi_2))t_1^c+(\lambda_2\phi_2+\lambda_3)t_2^c}{\alpha}\right]^{-\alpha}}{\lambda_1+\lambda_2(1-\phi_2)}, & 0 < t_1 \leq t_2 \\ \frac{\lambda_1(1-\phi_1)\left[1+\frac{(\lambda_1+\lambda_2+\lambda_3)t_1^c}{\alpha}\right]^{-\alpha}}{\lambda_2+\lambda_1(1-\phi_1)} \\ + \frac{\lambda_2\left[1+\frac{(\lambda_2+\lambda_1(1-\phi_1))t_2^c+(\lambda_1\phi_1+\lambda_3)t_1^c}{\alpha}\right]^{-\alpha}}{\lambda_2+\lambda_1(1-\phi_1)}, & 0 < t_2 \leq t_1. \end{cases} \quad (9.38)$$

The above bivariate survival function is the weighted combination of the bivariate Burr distribution and univariate Burr distribution with same weights as in BVE of Proschan-Sullo (1974). The unconditional bivariate survival function in (9.38) has three types of dependencies, one is due to simultaneous failures and second is due to load on one component due to the failure of another component and the third is due to frailty.

Substituting  $\lambda_3 = 0$  in all the above expressions we get the corresponding expressions for the Weibull extension of BVE of Freund (1961).

## 9.4 Positive Stable Frailty in BVW Models

We present the positive stable frailty model under the different bivariate Weibull baseline models in this Section.

### 9.4.1 Weibull Extension of BVE of Gumbel

Now the conditional survival distribution of Weibull extension of the BVE type I of Gumbel (1960) given the frailty ( $Z = z$ ) is given by

$$S(t_1, t_2 | z) = e^{-z(\lambda_1 t_1^{c_1} + \lambda_2 t_2^{c_2} + \delta \lambda_1 \lambda_2 t_1^{c_1} t_2^{c_2})} \quad (9.39)$$

where  $\lambda_1, \lambda_2, c_1, c_2 > 0, 0 \leq \delta \leq 1$ , and  $Z$  follows positive stable distribution given in (5.11).

Integrating over  $Z$ , we get unconditional survival function and is given by

$$S_\alpha(t_1, t_2) = e^{-(\lambda_1 t_1^{c_1} + \lambda_2 t_2^{c_2} + \delta \lambda_1 \lambda_2 t_1^{c_1} t_2^{c_2})^\alpha}. \quad (9.40)$$

The above bivariate survival function has two types of dependencies, one is due to dependence parameter  $\delta$  and other is due to frailty. The marginal distributions are also Weibull and are given by

$$S_\alpha(t_i) = e^{-(\lambda_i t_i^{c_i})^\alpha}, \quad i = 1, 2. \quad (9.41)$$

The parameters  $\lambda_1$  and  $\lambda_2$  here also expressed in terms of the regression parameters as we did in the Eq. (9.10).

Now the conditional survival distribution of Weibull extension of the BVE type II of Gumbel (1960) given the frailty ( $Z = z$ ) is given by

$$\begin{aligned} S(t_1, t_2 | z) = & (1 + \delta)e^{-z(\lambda_1 t_1^{c_1} + \lambda_2 t_2^{c_2})} - \delta e^{-z(2\lambda_1 t_1^{c_1} + \lambda_2 t_2^{c_2})} \\ & - \delta e^{-z(\lambda_1 t_1^{c_1} + 2\lambda_2 t_2^{c_2})} + \delta e^{-z(2\lambda_1 t_1^{c_1} + 2\lambda_2 t_2^{c_2})}, \end{aligned} \quad (9.42)$$

where  $Z$  follows positive stable distribution given in (5.11).

Integrating over  $Z$ , we get unconditional survival function and is given by

$$\begin{aligned} S_\alpha(t_1, t_2) = & (1 + \delta)e^{-(\lambda_1 t_1^{c_1} + \lambda_2 t_2^{c_2})^\alpha} - \delta e^{-(2\lambda_1 t_1^{c_1} + \lambda_2 t_2^{c_2})^\alpha} \\ & - \delta e^{-(\lambda_1 t_1^{c_1} + 2\lambda_2 t_2^{c_2})^\alpha} + \delta e^{-(2\lambda_1 t_1^{c_1} + 2\lambda_2 t_2^{c_2})^\alpha}. \end{aligned} \quad (9.43)$$

The marginal distributions are also of Weibull form with survival functions given by

$$S_\alpha(t_i) = e^{-(\lambda_i t_i^{c_i})^\alpha}, \quad i = 1, 2. \quad (9.44)$$

Now the conditional survival distribution of Weibull extension of BVE type III of Gumbel (1960) given the frailty ( $Z = z$ ) is given by

$$S(t_1, t_2 | z) = e^{-z(\lambda_1 t_1^{\frac{c_1}{\delta}} + \lambda_2 t_2^{\frac{c_2}{\delta}})^\delta}, \quad t_1, t_2 > 0 \quad (9.45)$$

where  $\lambda_1, \lambda_2, c_1, c_2 > 0, 0 \leq \delta \leq 1$ , and  $Z$  follows positive stable distribution given in (5.11).

Integrating over  $Z$ , we get unconditional survival function and is given by

$$S_\alpha(t_1, t_2) = e^{-(\lambda_1 t_1^{\frac{c_1}{\delta}} + \lambda_2 t_2^{\frac{c_2}{\delta}})^{\delta\alpha}}. \quad (9.46)$$

The marginal distributions are also of Weibull form with survival functions given by

$$S_\alpha(t_i) = e^{-(\lambda_i t_i^{c_i})^\alpha}, \quad i = 1, 2. \quad (9.47)$$

### 9.4.2 Weibull Extension of BVE of Marshall-Olkin

Suppose  $Z$  has a positive stable distribution which we assume as a frailty distribution. Now the conditional survival function of Weibull extension of BVE of Marshall-Olkin (1967) given the frailty ( $Z = z$ ) is given by

$$S(t_1, t_2 | z) = e^{-z(\lambda_1 t_1^c + \lambda_2 t_2^c + \lambda_3 t_{(2)}^c)} \quad (9.48)$$

where  $Z$  follows positive stable distribution given in (5.11).

Integrating over  $Z$ , we get unconditional survival function and is given by

$$S_\alpha(t_1, t_2) = e^{-(\lambda_1 t_1^c + \lambda_2 t_2^c + \lambda_3 t_{(2)}^c)^\alpha}. \quad (9.49)$$

The above bivariate survival function has two types of dependencies, one is due to simultaneous failures and other is due to frailty. When  $\alpha = 1$ , the frailty distribution is degenerate at  $Y = 1$ .

The main advantage of this model is that the marginal distributions are also of Weibull and  $T_{(1)} = \min(T_1, T_2)$  is also of Weibull form and corresponding survival functions are given by

$$\begin{aligned} S_\alpha(t_i) &= e^{-(\lambda_i + \lambda_3)^\alpha t_i^{c\alpha}}, & i = 1, 2 \\ S_\alpha(t_{(1)}) &= e^{-(\lambda_1 + \lambda_2 + \lambda_3)^\alpha t_{(1)}^{c\alpha}}. \end{aligned} \quad (9.50)$$

### 9.4.3 Weibull Extension of BVE of Block-Basu

Now the conditional survival function of Weibull extension of BVE of Marshall-Olkin (1967) given the frailty ( $Z = z$ ) is given by

$$S(t_1, t_2 | z) = \frac{\lambda}{\lambda_1 + \lambda_2} e^{-z(\lambda_1 t_1^c + \lambda_2 t_2^c + \lambda_3 t_{(2)}^c)} - \frac{\lambda_3}{\lambda_1 + \lambda_2} e^{-z\lambda t_{(2)}^c} \quad (9.51)$$

where  $Z$  follows positive stable distribution given in (5.11).

Integrating over  $Z$ , we get unconditional survival function and is given by

$$S_\alpha(t_1, t_2) = \frac{\lambda}{\lambda_1 + \lambda_2} e^{-(\lambda_1 t_1^c + \lambda_2 t_2^c + \lambda_3 t_{(2)}^c)^\alpha} - \frac{\lambda_3}{\lambda_1 + \lambda_2} e^{-(\lambda t_{(2)}^c)^\alpha}. \quad (9.52)$$

The above bivariate survival function has two types of dependencies, one is due to dependence parameter  $\lambda_3$  and other is due to frailty. When  $\alpha = 1$ , the frailty distribution is degenerate at  $Z = 1$ .

The marginal distributions are of weighted combinations of two Weibull and  $T_{(1)} = \min(T_1, T_2)$  is of Weibull form and corresponding survival function of  $T_{(1)}$  is given by

$$S_\alpha(t_{(1)}) = e^{-(\lambda_1 + \lambda_2 + \lambda_3)^\alpha t_{(1)}^{c\alpha}}. \quad (9.53)$$

#### 9.4.4 Weibull Extension of BVE of Freund and Proschan-Sullo

Now the conditional survival function of Weibull extension of BVE of Proschan-Sullo (1974) given the frailty ( $Z = z$ ) is given by

$$S(t_1, t_2 | z) = \begin{cases} \frac{\lambda_2(1-\phi_2)e^{-z(\lambda_1+\lambda_2+\lambda_3)t_2^c}}{\lambda_1+\lambda_2(1-\phi_2)} \\ + \frac{\lambda_1e^{-z[(\lambda_1+\lambda_2(1-\phi_2))t_1^c+(\lambda_2\phi_2+\lambda_3)t_2^c]}}{\lambda_1(1-\phi_1)e^{-z(\lambda_1+\lambda_2+\lambda_3)t_1^c}}, & 0 < t_1 \leq t_2 \\ \frac{\lambda_1+\lambda_2(1-\phi_2)}{\lambda_2+\lambda_1(1-\phi_1)} \\ + \frac{\lambda_2e^{-z[(\lambda_2+\lambda_1(1-\phi_1))t_2^c+(\lambda_1\phi_1+\lambda_3)t_1^c]}}{\lambda_2+\lambda_1(1-\phi_1)}, & 0 < t_2 \leq t_1 \end{cases} \quad (9.54)$$

where  $Z$  follows positive stable distribution given in (5.11).

Integrating over  $Z$ , we get unconditional survival function and is given by

$$S_\alpha(t_1, t_2) = \begin{cases} \frac{\lambda_2(1-\phi_2)e^{-((\lambda_1+\lambda_2+\lambda_3)t_2^c)^\alpha}}{\lambda_1+\lambda_2(1-\phi_2)} \\ + \frac{\lambda_1e^{-[(\lambda_1+\lambda_2(1-\phi_2))t_1^c+(\lambda_2\phi_2+\lambda_3)t_2^c]^\alpha}}{\lambda_1(1-\phi_1)e^{-((\lambda_1+\lambda_2+\lambda_3)t_1^c)^\alpha}}, & 0 < t_1 \leq t_2 \\ \frac{\lambda_1+\lambda_2(1-\phi_2)}{\lambda_2+\lambda_1(1-\phi_1)} \\ + \frac{\lambda_2e^{-[(\lambda_2+\lambda_1(1-\phi_1))t_2^c+(\lambda_1\phi_1+\lambda_3)t_1^c]^\alpha}}{\lambda_2+\lambda_1(1-\phi_1)}, & 0 < t_2 \leq t_1. \end{cases} \quad (9.55)$$

Substituting  $\lambda_3 = 0$  in all the above expressions we get the corresponding expressions for the Weibull extension of BVE of Freund (1961).

## 9.5 Power Variance Function Frailty in BVW Models

We present the power variance function (PVF) frailty model under the different bivariate Weibull baseline models in this section.

#### 9.5.1 Weibull Extension of BVE Models

The unconditional survival function of BVW with PVF frailty distribution corresponding to BVE type I of Gumbel (1960) is given by

$$S_{\alpha,\theta}(t_1, t_2) = e^{-\theta\{[1+(\lambda_1t_1^{c1}+\lambda_2t_2^{c2}+\delta\lambda_1\lambda_2t_1^{c1}t_2^{c2})/\theta]^\alpha-1\}/\alpha}. \quad (9.56)$$

The unconditional survival function of BVW with PVF frailty distribution corresponding to BVE type II of Gumbel (1960) is given by

$$\begin{aligned}
S_{\alpha,\theta}(t_1, t_2) = & (1 + \delta)e^{-\theta\{[1+(\lambda_1 t_1^{c_1} + \lambda_2 t_2^{c_2})/\theta]^\alpha - 1\}/\alpha} \\
& - \delta e^{-\theta\{[1+(2\lambda_1 t_1^{c_1} + \lambda_2 t_2^{c_2})/\theta]^\alpha - 1\}/\alpha} \\
& - \delta e^{-\theta\{[1+(\lambda_1 t_1^{c_1} + 2\lambda_2 t_2^{c_2})/\theta]^\alpha - 1\}/\alpha} \\
& + \delta e^{-\theta\{[1+(2\lambda_1 t_1^{c_1} + 2\lambda_2 t_2^{c_2})/\theta]^\alpha - 1\}/\alpha}.
\end{aligned} \tag{9.57}$$

The unconditional survival function of BVW with PVF frailty distribution corresponding to BVE type III of Gumbel (1960) is given by

$$S_{\alpha,\theta}(t_1, t_2) = e^{-\theta\{[1+(\lambda_1 t_1^{c_1} + \lambda_2 t_2^{c_2})/\delta]^\alpha - 1\}/\alpha}. \tag{9.58}$$

The unconditional survival function of BVW with PVF frailty distribution corresponding to BVE of Marshall-Olkin (1967) is given by

$$S_{\alpha,\theta}(t_1, t_2) = e^{-\theta\{[1+(\lambda_1 t_1^c + \lambda_2 t_2^c + \lambda_3 t_{(2)}^c)/\theta]^\alpha - 1\}/\alpha}. \tag{9.59}$$

The unconditional survival function of BVW with PVF frailty distribution corresponding to BVE of Block-Basu (1974) is given by

$$\begin{aligned}
S_{\alpha,\theta}(t_1, t_2) = & \frac{\lambda}{\lambda_1 + \lambda_2} e^{-\theta\{[1+(\lambda_1 t_1^c + \lambda_2 t_2^c + \lambda_3 t_{(2)}^c)/\theta]^\alpha - 1\}/\alpha} \\
& - \frac{\lambda_3}{\lambda_1 + \lambda_2} e^{-\theta\{[1+\lambda t_{(2)}^c/\theta]^\alpha - 1\}/\alpha}.
\end{aligned} \tag{9.60}$$

The unconditional survival function of BVW with PVF frailty distribution corresponding to BVE of Proschan-Sullo (1974) is given by

$$S_{\alpha,\theta}(t_1, t_2) = \begin{cases} \frac{\lambda_2(1-\phi_2)e^{-\theta\{[1+\lambda t_2^c/\theta]^\alpha - 1\}/\alpha}}{\lambda_1 + \lambda_2(1-\phi_2)} \\ + \frac{\lambda_1e^{-\theta\{[1+(\lambda_1 + \lambda_2(1-\phi_2))t_1^c + (\lambda_2\phi_2 + \lambda_3)t_2^c]/\theta]^\alpha - 1\}/\alpha}}{\lambda_1 + \lambda_2(1-\phi_2)}, & t_1 \leq t_2 \\ \frac{\lambda_1(1-\phi_1)e^{-\theta\{[1+\lambda t_1^c/\theta]^\alpha - 1\}/\alpha}}{\lambda_2 + \lambda_1(1-\phi_1)} \\ + \frac{\lambda_2e^{-\theta\{[1+(\lambda_2 + \lambda_1(1-\phi_1))t_2^c + (\lambda_1\phi_1 + \lambda_3)t_1^c]/\theta]^\alpha - 1\}/\alpha}}{\lambda_2 + \lambda_1(1-\phi_1)}, & t_2 \leq t_1 \end{cases} \tag{9.61}$$

where  $\lambda = \lambda_1 + \lambda_2 + \lambda_3$ .

Substituting  $\lambda_3 = 0$  in all the above expressions we get the corresponding expressions for the Weibull extension of BVE of Freund (1961).

## 9.6 Lognormal and Weibull Frailties in BVW Models

Hanagal (2004) proposed bivariate Weibull (BVW) with survival function given by

$$S(t_1, t_2) = e^{-\lambda_1 t_1^c - \lambda_2 t_2^c - \lambda_3 t_{(2)}^c} \quad (9.62)$$

where  $t_{(2)} = \max(t_1, t_2)$ ,  $\lambda_1, \lambda_2, \lambda_3, c > 0$ .

The conditional survival function with fixed covariates and given frailties is given by

$$S(t_1, t_2 | \beta, u) = e^{-[\lambda_1 t_1^c + \lambda_2 t_2^c + \lambda_3 t_{(2)}^c] e^\eta} \quad (9.63)$$

where  $\eta = y' \beta + u$  (using the notations of Section 5.7).

### 9.6.1 Estimation of Parameters

The log-likelihood or h-likelihood is given by

$$\begin{aligned} h = \ln L(\mathbf{t}_1, \mathbf{t}_2, \beta, \mathbf{u}) &= \ln L_1(\mathbf{t}_1, \mathbf{t}_2 | \beta, \mathbf{u}) + \ln L_2(\mathbf{u}) \\ &= l_1(\mathbf{t}_1, \mathbf{t}_2 | \beta, \mathbf{u}) + l_2(\mathbf{u}). \end{aligned} \quad (9.64)$$

The estimation achieved through the use of the BLUP method is used as an initial step in the computation of both the ML and REML estimators of variance component. The BLUP method predicts frailty. BLUP estimators on the other hand refer to the estimated fixed effects. The BLUP method is based on the maximization of the sum of the above two components. The second component in (9.64) is the loglikelihood of frailty model.

Bivariate and multivariate failure time data have been often proved difficult to analyze. When modeling such data based on either proportional hazard, or analyzing with parametric models including the frailty component, the failure times distribution formed by integrating out the random frailty component loses the simple properties of the original hazard function formulation. The method suggested here preserves all the simple properties of baseline parametric model without integrating out the random frailty component. The expression (9.64) is the general formula for any frailty random variable  $U$ . As an illustration, we have assumed  $U$  follows normal distribution with mean zero and variance  $\sigma^2$ . The procedure is same if  $U$  follows any other distribution. The only difference is we have to estimate the corresponding parameters involved in the distribution of  $U$ , including the variance component,  $\sigma^2 = \text{Var}(U)$ . Lognormal frailty distribution is a one-parameter model, which has one nice property of infinite divisibility (of both the frailty and logarithm to the frailty). A further advantage is that a power transformation of the frailty still gives a lognormal distribution. In most cases, the frailty term is

assumed to have lognormal distribution (See McGilchrist and Aisbett, 1991; McGilchrist, 1993, 1994; Yau and McGilchrist, 1998; Noh, Ha and Lee, 2006).

If we do not incorporate frailty term, the variance component,  $\sigma^2$  will be added into error term. If the estimate of  $\sigma^2$  is very close to zero, frailty will not play much role in the model and if the estimate of  $\sigma^2$  is significantly higher than zero, then frailty plays important role in the regression model and also reduces the mean square error (MSE) of the regression parameters.

Note that if  $U$  is normal or log-gamma or log-Weibull, then  $Z = \exp(U)$  becomes lognormal or gamma or Weibull respectively, and the corresponding model is called lognormal or gamma or Weibull frailty model. The second component in (9.64), assuming  $U$  follows normal is

$$l_2(\mathbf{u}) = -\frac{1}{2} \left[ n \ln 2\pi\sigma^2 + \sum_{i=1}^n \frac{u_i^2}{\sigma^2} \right]. \quad (9.65)$$

The second component in (6.5.3), assuming  $U$  follows extreme value distribution is

$$l_2(\mathbf{u}) = -n \ln b - n\nu + (1/b) \sum_{i=1}^n u_i - e^{-\nu} \sum_{i=1}^n e^{u_i/b}. \quad (9.66)$$

The estimators derived from maximizing the h-likelihood will be called as maximum hierarchical likelihood (MHL) estimators. The procedure forms the basis of the estimation process. In the case that  $l_1$  depends on  $c$  and  $\beta$  only through a linear predictor  $x'\beta + z'u$ , where  $x$  and  $z$  are design matrices, the solution of the estimating equations is conveniently carried out by the Newton-Raphson procedure.

Corbeil and Searle (1976) suggest a modification of ML procedure, whereby the likelihood is factored into two parts. The first part is maximized to obtain ML estimates of the fixed effects and second part is free of the fixed effects. Maximizing the second component yields MHL and restricted maximum hierarchical likelihood (RMHL) estimators of  $\sigma^2$  given by

$$\hat{\sigma}_{MHL}^2 = \frac{\sum_{i=1}^n \hat{u}_i^2 + \text{tr}(A^{-1})}{n} \quad (9.67)$$

$$\hat{\sigma}_{RMHL}^2 = \frac{\sum_{i=1}^n \hat{u}_i^2 + \text{tr}(A)}{n} \quad (9.68)$$

where  $A = (-H^{-1})_{22}$ , a sub-matrix corresponding to frailties  $\mathbf{u}$ .

The estimated asymptotic variances of (9.67) and (9.68) are  $2(\hat{\sigma}_{MHL}^2)^2[n - (\hat{\sigma}_{MHL}^2)^{-1}\text{tr}(A^{-1}) + (\hat{\sigma}_{MHL}^2)^{-2}\text{tr}(A^{-2})]^{-1}$  and  $2(\hat{\sigma}_{RMHL}^2)^2[n - (\hat{\sigma}_{RMHL}^2)^{-1}\text{tr}(A) + (\hat{\sigma}_{RMHL}^2)^{-2}\text{tr}(A^2)]^{-1}$  respectively. These forms were used by Schall (1991) and McGilchrist (1993, 1994).

Then  $\beta$  and  $\mathbf{u}$  are re-estimated using either MHL or RMHL estimators of  $\sigma$  and so on until all parameters  $(\beta, \mathbf{u}, \sigma)$  converge. We get estimates of

$(\beta, \mathbf{u}, \sigma)$  based on MHL and RMHL depending upon the respective equations (9.67) and (9.68) used in the Newton-Raphson procedure. Hanagal (2008b) has obtained tests for lognormal frailty i.e., for testing the null hypothesis  $H_0: \sigma^2 = 0$  based on the asymptotic normal test statistics  $W_1 = \frac{\hat{\sigma}_{MHL}^2}{\hat{s}e(\hat{\sigma}_{MHL}^2)}$  or  $W_2 = \frac{\hat{\sigma}_{RMHL}^2}{\hat{s}e(\hat{\sigma}_{RMHL}^2)}$  where  $\hat{s}e(\hat{\sigma}_{MHL}^2)$  and  $\hat{s}e(\hat{\sigma}_{RMHL}^2)$  are estimated under  $H_0 \cup H_1$ . A similar procedure, based on studentizing the test statistic was developed by Hanagal and Kale (1991).

---

## 9.7 Compound Poisson Frailty in BVW Models

The conditional survival function of BVW of Hanagal (2004) given the frailty ( $Z = z$ ) is given by

$$S(t_1, t_2 | z) = e^{-z(\lambda_1 t_1^c + \lambda_2 t_2^c + \lambda_3 t_{(2)}^c)} \quad (9.69)$$

where  $Z$  follows compound Poisson distribution with parameters  $\alpha$ ,  $\delta$  and assuming  $E(Z) = 1$ .

Integrating over  $Z$ , we get unconditional survival function and is given by

$$S(t_1, t_2) = \exp \left\{ \frac{\alpha}{(1-\alpha)\delta} \left[ 1 - \left( 1 + \frac{\delta}{\alpha} (\lambda_1 t_1^c + \lambda_2 t_2^c + \lambda_3 t_{(2)}^c) \right)^{1-\alpha} \right] \right\}, \quad (9.70)$$

where  $\alpha, \delta \geq 0$ .

The unconditional bivariate survival function in (9.70) has two types of dependencies, one is due to simultaneous failures and the other is due to frailty. The parameters  $\lambda_1$ ,  $\lambda_2$ , and  $\lambda_3$  can be expressed in terms of regression parameters as we did earlier.

---

## 9.8 Compound Poisson (with Random Scale) Frailty in BVW Models

The conditional survival function of BVW given the frailty ( $Z=z$ ) is given by (See Eq. (9.69))

$$S(t_1, t_2 | z) = e^{-z(\lambda_1 t_1^c + \lambda_2 t_2^c + \lambda_3 t_{(2)}^c)} \quad (9.71)$$

where  $Z$  follows compound Poisson-gamma distribution with random scale and parameters  $\eta$ ,  $\theta$ ,  $\nu$  and assuming  $E(Z) = 1$ .

Integrating over  $Z$ , we get unconditional survival function and is given by

$$S(t_1, t_2) = \left( \frac{\theta}{\theta + 1 - \left( \frac{\nu}{\nu + \lambda_1 t_1^c + \lambda_2 t_2^c + \lambda_3 t_{(2)}} \right)^{\eta}} \right)^{\nu \theta / \eta}. \quad (9.72)$$

The unconditional bivariate survival function in (9.72) has two types of dependencies, one is due to simultaneous failures and the other is due to frailty.

---

## 9.9 Estimation and Tests for Frailty under BVW Baseline

For the bivariate life time distribution, we use univariate censoring scheme given by Hanagal (1992a, 1992b) because the individuals do not enter at the same time and withdrawal or death of an individual or termination of the study will censor both life times of the components. Here the censoring time is independent of the life times of both components. This is the standard univariate right censoring for both failure times  $T_1$  and  $T_2$ .

Suppose that there are  $n$  independent pairs of components under study and  $i$ -th pair of the components have life times  $(t_{1i}, t_{2i})$  and a censoring time  $(w_i)$ . The life times associated with  $i$ -th pair of the components are given by

$$\begin{aligned} (T_{1i}, T_{2i}) &= (t_{1i}, t_{2i}), & \max(t_{1i}, t_{2i}) < w_i \\ &= (t_{1i}, w_i), & t_{1i} < w_i < t_{2i} \\ &= (w_i, t_{2i}), & t_{2i} < w_i < t_{1i} \\ &= (w_i, w_i), & z_i < \min(t_{1i}, t_{2i}). \end{aligned} \quad (9.73)$$

Discarding factors which do not contain any of the parameters, we want to estimate the parameters in the proposed model. Now the likelihood of the sample of size  $n$  is given by

$$L = \left( \prod_{i=1}^{n_1} f_{1i} \right) \left( \prod_{i=1}^{n_2} f_{2i} \right) \left( \prod_{i=1}^{n_3} f_{3i} \right) \left( \prod_{i=1}^{n_4} f_{4i} \right) \left( \prod_{i=1}^{n_5} f_{5i} \right) \left( \prod_{i=1}^{n_6} \bar{F}_i \right) \quad (9.74)$$

where

$$\begin{aligned}
 f_{1i} &= \frac{\partial^2 S(t_{1i}, t_{2i}, t_{1i} < t_{2i} | \mathbf{y})}{\partial t_{1i} \partial t_{2i}}, & 0 < t_{1i} < t_{2i} < w_i \\
 f_{2i} &= \frac{\partial^2 S(t_{1i}, t_{2i}, t_{1i} > t_{2i} | \mathbf{y})}{\partial t_{1i} \partial t_{2i}}, & 0 < t_{2i} < t_{1i} < w_i \\
 f_{3i} &= -\frac{\partial S(t_{1i}, t_{2i}, t_{1i} = t_{2i} = t_i | \mathbf{y})}{\partial t_i}, & 0 < t_{1i} = t_{2i} = t_i < w_i \\
 f_{4i} &= -\frac{\partial S(t_{1i}, w_i | \mathbf{y})}{\partial t_{1i}}, & 0 < t_{1i} < w_i < t_{2i} \\
 f_{5i} &= -\frac{\partial S(w_i, t_{2i} | \mathbf{y})}{\partial t_{2i}}, & 0 < t_{2i} < w_i < t_{1i} \\
 \bar{F}_i &= S(w_i, w_i | \mathbf{y}), & 0 < w_i < \min(t_{1i}, t_{2i})
 \end{aligned}$$

$n_1, n_2, n_3, n_4, n_5$  and  $n_6$  are the number of observations observed to fail in the range space corresponding to  $f_{1i}, f_{2i}, f_{3i}, f_{4i}, f_{5i}$ , and  $\bar{F}_i$ , respectively.  $f_{1i}$  and  $f_{2i}$  are the conditional pdf with respect to Lebesgue measure in  $R^2$  and  $f_{3i}, f_{4i}$ , and  $f_{5i}$  are the conditional pdf with respect to Lebesgue measure in  $R^1$  in their respective regions. When there is absolute continuity in BVW model,  $n_3 = 0$  and  $f_{3i}$  does not exist.

The likelihood equations can be obtained by taking first order partial derivatives of the loglikelihood with respect to the parameters and equating to zero. Usually the likelihood equations are not easy to solve. It may not be possible to obtain maximum likelihood estimators (MLEs) by the Newton-Raphson procedure. Sometimes, the likelihood equations do not converge for the specified sample sizes in the Newton-Raphson procedure and the method of maximum likelihood (ML) fails to estimate all the parameters simultaneously. One can obtain estimates of the parameters by two stage MLE method or conditional MLE method (See Hanagal, 2005c, 2006d, 2008b). In the first stage, estimate the parameters  $\eta, \theta, \nu, c$  by ML method under the base line model by conditioning  $\beta = 0$  and then in the second stage, estimate the parameters  $\beta$  by ML method after substituting MLEs of the parameters of the baseline model obtained from the first stage. Then re-substitute the estimates of  $\beta$  in the first stage and estimate the parameters in baseline model. Continue this iterative procedure until the convergence meet in both stages.

Hanagal (2005d, 2006b, 2006d, 2006f) has obtained estimation of the parameters and test for regression coefficients under different BVW baseline with gamma frailty model having the pdf given in Eq. (5.2). Hanagal (2005c, 2006e) has obtained estimation of the parameters, test for frailty and test for regression coefficients under different BVW with positive stable frailty model having the pdf given in Eq. (5.11). Hanagal (2007c, 2009b) has obtained the estimation of the parameters, test for frailty and test for regression coefficients under BVW baseline with PVF frailty model. Hanagal (2008b, 2010c) has obtained the estimation of the parameters, test for frailty and test for regression coefficients under BVW baseline with lognormal and Weibull frailty models.

Hanagal (2010b) has obtained the estimation of the parameters and test for regression coefficients under BVW baseline with compound Poisson frailty model. Hanagal (2010d) has obtained the estimation of the parameters and test for regression coefficients under BVW baseline using compound Poisson frailty with random scale model.

# Chapter 10

---

## *Frailty Models Based on Lévy Processes*

10.1	Introduction .....	185
10.1.1	Biological Interpretation of Failure Rate .....	186
10.1.2	A Model for Random Failure Rate Processes .....	187
10.2	Lévy Processes and Subordinators .....	189
10.2.1	Standard Compound Poisson Process .....	190
10.2.2	Compound Poisson Process with General Jump Distribution .....	191
10.2.3	Gamma Processes .....	191
10.2.4	Stable Processes .....	191
10.2.5	PVF Processes .....	191
10.2.6	Special Cases .....	192
10.3	Proportional Hazards Derived from Lévy Processes .....	193
10.3.1	Example 10.1: Gamma Process .....	193
10.3.2	Example 10.2: Compound Poisson Process .....	194
10.4	Other Frailty Process Constructions .....	195
10.5	Hierarchical Lévy Frailty Models .....	196
10.5.1	Application to the Infant Mortality Data .....	200

---

### 10.1 Introduction

Gjessing et al. (2003) have given a very nice introduction of frailty models based on Lévy processes. We present some of their results in this chapter. The failure rate is theoretical and descriptive technique which plays a fundamental role in survival and event history analysis. Basic methods in these fields, such as the counting process approach and the Cox model, focus on failure rates. The purpose of the failure rate is to measure, locally, the risk of an event given the nonoccurrence of the event up to that time. It is an intuitive and attractive concept, and is also closely related to incidence rate in epidemiology. In a context of probability, the failure rate is a special case of an intensity process or the derivative of a Doob-Meyer compensator; see Andersen et al. (1994).

In practice, the failure rate must be estimated from data, based on the observation of a number of individuals over time. Then we are faced with the fact that individuals will be dissimilar. In a sense, we can say that each individual will have his own failure rate, and what is estimated on the basis of a number of individuals is same sort of average. It is well known that the average failure rate may be entirely different from those individuals. Even when all individuals have failure rates with the same functional form but

different levels determined by varying proportionally constants, the estimated failure rate will have an entirely different shape from that of the individuals. This has been pointed out in frailty theory; see Hougaard (2000). Gjessing et al. (2003) have given detailed description of frailty models based on Lévy processes.

### 10.1.1 Biological Interpretation of Failure Rate

In the biological literature, several attempts at interpreting shapes of failure rates and drawing biological conclusions from these may be found. Clarke et al. (2000) discussed the development of diseases involving neuronal degeneration (e.g. Parkinson's disease). In such diseases, the onset of clinical symptoms may be delayed for years or decades after the premature neuronal death has started. A discussion of whether this delay reflects a cumulative damage process, as whether the death of a neuron is basically a random event, the likelihood of which is constant over time. Presumably, the first hypothesis would lead to increasing failure of cell death, while the second one would give a constant failure rate. The cumulative damage model implies a progressive process in the organism whereby previous cell death creates damage that increases the likelihood of new cell death. The random event model, also called a 'one-hit' model, implies that there has been damages at one specific time, e.g., a mutation, which then increases the likelihood of cell death to a new fixed level.

For a number of diseases and animal models, Clarke et al. (2000) studied the survival of neurons. They found that the failure rate is generally constant, or sometimes decreasing but not increasing. From this they draw the conclusion that the cumulative damage model is incorrect and that the one-hit random model is the true one. Clarke et al. (2001) gave a further discussion of this hypothesis. A similar study for a type of mutant mice is presented by Triarhou (1998), with the same conclusion of a constant rate of cell death.

Another setting where biological interpretations of failure rates have been presented is in the understanding of sleep. Lo et al. (2002) studied the changes between sleep and wake states throughout the night. They found that the duration of sleep periods has, more or less, an exponential distribution, while the duration of wake states has a distribution with a decreasing failure rate. This is interpreted within a stochastic process context, but basically the implication is that sleep represents a random walk with no drift.

A common feature in these papers is that they draw biological conclusions from the shapes of failure rates. It is then important to understand how very different underlying models may lead to similar failure rates. In particular, an approximately constant failure rate will be a common phenomenon for many models due to convergence to quasi-stationarity. It therefore appears difficult to draw conclusions about the underlying process with any degree of certainty.

An added difficulty comes from the possibility of frailty variation, that is, heterogeneity in risk between different individuals, those individuals being

either individual cells or organisms. Hence, the development of risk at the level of an individual neuron, say, is very hard to deduce.

This criticism does not necessarily imply that the conclusions in the mentioned papers are wrong. But it points to the necessity of understanding the failure rate and how its various shapes can arise.

### 10.1.2 A Model for Random Failure Rate Processes

The standard approach for combining a model for the failure rate with an allowance for individual heterogeneity in risk is the frailty model

$$\mu_i(t) = Z_i h(t) \exp(\beta' y_i). \quad (10.1)$$

The deterministic failure rate  $h(t)$  is estimated (parametrically or non-parametrically) from data. The exponential term allows the parameter vector  $\beta$  to be estimated, giving the effect of the covariate vector  $y_i$  on the failure rate. The individual frailty  $Z_i$  is drawn from a distribution of non-negative random variables. The subscript  $i$  indicates that a new realization of  $Z$  is drawn for each individual. For more information on standard frailty models and many extensions, see Hougaard (2000) and Aalen (1994). The simplifying assumptions made in a standard frailty model enable us to estimate the true underlying failure of individuals, which is easier when multivariate survival data are available, as compared to univariate data. However, there are two important aspects of the simplifying assumptions. First, it is assumed that all individuals have proportional hazards. This is obviously an assumption of mathematical convenience, although the theory developing from it also seems to yield useful insights into biological phenomena. Second, the individual, frailty  $Z_i$  is determined at time zero and follows the individual throughout its entire life, resulting in the individual risk later in life being perfectly correlated with that at the beginning. It is clearly of interest to develop more flexible models.

A general view of the individual variation in failure rates would be to regard the failure of an individual as a stochastic process. This would give individual flexibility. However, since what is observable is still the average failure rate, there should be a tractable mathematical connection between the individual failure rates and the average one. One approach, which has been tried with some success by Yashin and Manton (1997), is to define the failure rate as a diffusion process by means of a stochastic differential equation. It is not unnatural to think that the risk of an individual may develop as some kind of diffusion. By introducing Ornstein-Uhlenbeck-type processes, we may model the fact that many biological parameters tend to stabilize around certain values, or what is called homeostasis.

Here we shall attempt a different approach, namely one based on Lévy processes. This has been suggested before and developed to some extent by Kebir (1991), but we shall give here a number of additional results. Like diffusion processes, failure rates driven by Lévy processes also yield some degree

of tractability. It is possibility to get explicit formulae for the relationship between individual and average failure rate, the frailty and failure rate of survivors may be calculated, and so on. A basic difference from diffusion processes is the jump nature of the non-negative Lévy processes applied here. However, it could be imagined that individual failure rate may increase in jumps, for instance by the onset of an acute disease.

In this chapter, we shall not go into statistical estimation. The object is to study a mathematically tractable framework. One type of result we shall focus on is the occurrence of quasi-stationarity in some models. This means that, even though individuals are constantly leaving the risk set, the distribution of the failure rate of the survivors may still converge to a limit. An effect of this will be that the average failure rate converges towards a constant level, which is sometimes seen in practice. In first-passage time models, quasi-stationarity is important in understanding the shape of the failure rate (See Aalen and Gjessing, 2001).

The extension of frailty theory proposed here is hoped to have some biological relevance. When looking at the risk of disease it will usually be the case that some of the risk is already acquired at birth, for instance by the genetic makeup of the individual. An additional risk then accumulates throughout life depending on life style, various life circumstances, and accidental events. Also, important biological parameters, such as blood pressure and cholesterol, typically increase throughout life. If the focus is on the development of some disease, even this may depend on observable marker processes, such as CD4 counts in HIV infection. Such marker processes yield stochastic hazards. See Slate and Turnbull (2000). There is presently a development in survival analysis towards joint modeling of survival on the one hand and marker or covariate processes on the other hand. Mathematical frameworks may be useful for such developments.

We shall consider here frailty distribution defined by non-negative Lévy process, which in this chapter is taken to mean a process with non-negative, independent, time-homogeneous increments, i.e., a subordinator. The Laplace transform of such a process  $Z = \{Z(t) : t \geq 0\}$  at time  $t$  is given by Lévy-Khintchin formula

$$L(s, t) = E \exp\{-sZ(t)\} = \exp\{-t\Phi(s)\} \quad (10.2)$$

where  $s \geq 0$  is the argument of the Laplace transform. The function  $\Phi(s)$  is called the Laplace exponent of the Lévy process. The family of Lévy processes contains a number of important special cases, such as compound Poisson processes, gamma processes, stable processes, etc. In fact, all non-negative Lévy processes are limits of compound Poisson processes. For background material on Lévy processes, see Bertoin (1996).

When  $Z(t)$  is a subordinator, it is natural to associate frailty with such a process, since frailty of an individual can frequently be thought of as increasing over time. However, we would want to extend this to more general situations. To consider processes with a varying ‘rate’, define the non-negative determin-

istic rate function  $r(t)$  with integral  $R(t) = \int_0^t r(u)du$ , and let  $Z(R(t))$  be the time-transformed subordinator. Conditional on  $Z$ , we define our basic failure rate processes  $\lambda(t)$  as

$$\lambda(t) = h(t) \int_0^t a(u, t-u)dZ(R(t)). \quad (10.3)$$

This process will be starting point for all models considered in this chapter. The function  $a(u, s)$  is a weight function determining the extent to which the effect of previous jumps in  $Z(t)$  influence the failure rate at time  $t$ . Having two arguments in  $a(u, s)$  allows for the weight depending on two time scales, namely the time-scale of the stochastic process (first argument), and time as in distance from the current time (second argument). The deterministic function  $h$  determines the ‘base’ level of the failure rate, although it could just as well be absorbed in  $a$ . Since most frailty distributions of the classical frailty models are distributions of Lévy processes, the general formulation here incorporates most of these classical models, as will be seen below. Notice that, as in the standard frailty model Eq. (10.1), each individual in the population will have its own realization of the underlying process  $Z$  and thus of the hazard process  $\lambda$ , although we will suppress the subscript.

An introduction to models similar to those considered here is given by Kebir (1991). An overview of hazard processes developing under randomness is given by Singpurwalla (1995). Some models of frailty using Lévy processes are discussed in Aalen and Hjort (2002) and Hjort (2003). Gjessing et al. (2003) extend the results of Kebir (1991).

We will start by giving a short review of Lévy processes with some examples on Section 10.2. In Section 10.3, we present the proportional hazards derived from Lévy processes. Section 10.4 discusses other frailty process constructions. In the last Section 10.5, hierarchical Lévy frailty models is discussed.

## 10.2 Lévy Processes and Subordinators

The full distribution of a non-negative Lévy process, i.e. a subordinator  $Z$ , is determined by its Laplace exponent  $\Phi$ , defined in (10.2). For this reason, many general results on subordinates are derived in terms of  $\Phi$ . The function  $\Phi(\cdot)$  is increasing and concave, its derivative  $\Phi'$  decreasing. We have the general representation

$$\Phi(s) = d + \int_0^\infty (1 - e^{-sx})\pi(dx),$$

where  $d$  is called the drift coefficient since, at time  $t$ , the distribution of  $Z(t)$  is shifted to the right by an amount  $td$ . The measure  $\pi(dx)$  has support on

$(0, \infty)$  and satisfies the condition  $\int_0^\infty (1 \Lambda x) \pi(dx) < \infty$ . It is called the Lévy measure of the process  $Z$ . Notice that, when  $\int_0^\infty \pi(dx) = \rho < \infty$ , i.e. when  $(\frac{1}{\rho})\pi$  is a probability measure on  $(0, \infty)$ , we can write

$$\Phi(s) = ds + \rho(1 - L_0(s)),$$

where  $L_0(s)$  is the Laplace transform of  $(1/\rho)\pi$ . For the derivatives we have

$$\Phi'(s) = d + \int_0^\infty x e^{-sx} \pi(dx),$$

$$\Phi''(s) = - \int_0^\infty x^2 e^{-sx} \pi(dx),$$

which both exist for  $s > 0$ . Clearly,  $\lim_{s \rightarrow \infty} \Phi'(s) = d$ . When  $s = 0$ , we obtain the relationships

$$EZ(t) = \Phi'(0)t = (d + \int_0^\infty x \pi(dx))t,$$

$$Var Z(t) = -\Phi''(0)t = \int_0^\infty x^2 \pi(dx)t,$$

which in some cases will be infinite.

### 10.2.1 Standard Compound Poisson Process

Distributions generated by a compound Poisson process have previously been used as frailty models, see Aalen (1992) and Aalen and Tretli (1999). They are important examples in the present framework of Lévy processes. A compound Poisson process is constructed as follows. A Poisson process of rate  $\rho$  is running on time-scale  $t$ , and to each jump there is associated a gamma random variable, independent of the Poisson process, with shape parameter  $\eta > 0$  and scale parameter  $\nu > 0$ . The compound Poisson process is the sum of the gamma random variables up to time  $t$ . The Laplace transform of the distribution of the process at time  $t$  is given by

$$L(s, t) = \exp\{-t\rho[1 - (\nu/(\nu + s))^\eta]\}.$$

Hence,

$$\Phi(s) = \rho\{1 - L_0(s)\} \tag{10.4}$$

where  $L_0(s) = (\nu/(\nu + s))^\eta$  is the Laplace transform of the gamma distribution of the jumps and

$$\pi(dx) = \rho \frac{\nu^\eta}{\Gamma\eta} x^{\eta-1} e^{-\nu x} dx.$$

### 10.2.2 Compound Poisson Process with General Jump Distribution

Let the gamma variables be replaced by positive random variable  $X$  with distribution  $\mu_X(dx)$  and Laplace transform  $L_0(s) = E \exp(-sx)$ . The Laplace transform of the process now takes the form  $L(s, t) = \exp\{-\rho t + \rho t L_0(s)\}$ , giving

$$\Phi(s) = \rho\{1 - L_0(s)\}.$$

The Lévy measure  $\pi(dx) = \rho\mu_X(dx)$  is proportional to the jump distribution  $\mu_X$ .

It is true in general that  $\Phi$  is bounded if and only if  $Z$  is compound Poisson. This is proved for the characteristic exponent by Bertoin (1996), but the argument is easily adopted to  $\Phi$ .

### 10.2.3 Gamma Processes

Now let the value of the process at time  $t$  be gamma distributed with shape parameter  $\rho t$  and scale parameter  $\nu$ . Then

$$L(s, t) = \left( \frac{\nu}{\nu + s} \right)^{\rho t} = \exp\{-\rho t[\ln(\nu + s) - \ln \nu]\},$$

so that  $\Phi(s) = \rho\{\ln(\nu + s) - \ln \nu\}$ . Here  $\pi(dx) = \rho e^{-\nu x}/x dx$ . This can be considered a borderline case of the standard compound Poisson process in Section 10.2.1 when  $\eta \rightarrow 0$  and  $\rho \rightarrow \infty$  in such a way that  $\rho t$  converges to a positive constant.

### 10.2.4 Stable Processes

Stable frailty distribution have been applied successfully by Hougaard (1986, 2000) who has pointed out that they may preserve proportional hazards. The Laplace transform takes the form  $L(s, t) = \exp(-t\alpha s^\beta)$ , where  $\beta$  is a parameter in  $(0, 1)$ . Hence,  $\Phi(s) = \alpha s^\beta$ , with Lévy measure

$$\pi(dx) = \frac{\alpha \beta}{\Gamma(1 - \beta)} x^{-1 - \beta} dx.$$

### 10.2.5 PVF Processes

The PVF (power variance function) distributions constitute a general class of distributions studied, for instance, see Hougaard (2000). A class of Lévy processes may be defined from the PVF distribution by letting

$$\Phi_{PVF}(s, \rho, \nu, n) = \rho \left\{ 1 - \left( \frac{\nu}{\nu + s} \right)^n \right\}, \quad (10.5)$$

with  $\nu > 0, \eta > -1$ , and  $\eta\rho > 0$ . This is a direct extension of the standard compound Poisson process defined in (10.5) above. The Lévy measure is

$$\pi(dx) = \rho \frac{\nu^\eta}{\Gamma(\eta)} x^{\eta-1} e^{-\nu x} dx.$$

When  $\eta > 0$ , this is the standard compound Poisson. For  $\eta = 0$ , we recognize the gamma process as a borderline case. If  $0 > \eta > -1$ , then  $\pi([0, \infty)) = \infty$ , so the process is no longer compound Poisson, but the special case of  $\eta = \frac{1}{2}$  is the inverse Gaussian process and, when  $\nu \rightarrow 0$  and  $\rho \rightarrow -\infty$  in such a way that  $-\rho\nu^\eta$  converges to positive value, the result is a stable process.

### 10.2.6 Special Cases

#### Moving average:

Let  $a(t, v) \equiv a(v)$  depend only on the second argument and let  $r(t) \equiv h(t) \equiv 1$  (any constant value of  $h$  can be absorbed in  $a$ , and any constant value of  $r$  can be absorbed in  $\Phi$ ). Since  $\lambda(t) = \int_0^t a(t-u)dZ(u) = \int_0^t a(v)dZ(t-v)$ ,  $\lambda$  is seen to be a moving average process (although not necessarily stationary). Define  $A(v) = \int_0^v a(u)du$ . Then  $b(u, t) = A(t-u)$  and

$$S(t) = \exp \left( - \int_0^t \Phi(A(v))dv \right) \quad \text{and} \quad \mu(t) = \Phi(A(t)).$$

Note that  $\mu(t)$  is increasing and, if  $a$  is decreasing, also concave. It is clear that, if either  $\Phi$  is bounded (i.e.  $Z$  is compound Poisson) or  $A(\infty) := \lim_{t \rightarrow \infty} A(t) < \infty$ , then  $\lim_{t \rightarrow \infty} \mu(t) = \Phi(A(\infty)) < \infty$ , so that the hazard converges to a limit, and it is reasonable to assume that there is a quasi-stationary distribution for the hazard of survivors.

#### Standard frailty model:

Assume that  $a(t, v) \equiv 1$ . Let  $r(t)$  be equal to  $\rho$  up to time  $T$  and 0 after this time, and assume that  $h(t)$  is equal to 0 up to time  $T$ . From the general model (10.3) it follows that the hazard process equals

$$\lambda(t) = h(t)Z(\rho T), \quad t \geq 0.$$

The population hazard rate is  $\mu(t) = \rho Th(t)\Phi'(H(t))$  for  $t \geq 0$ , where  $H(t) = \int_0^t h(s)ds$ . We recognize the hazard rate of the standard frailty model, where the frailty distribution is generated by a Lévy process, as are almost all common frailty distributions. For instance, the PVF distributions described in Hougaard (2000) are distributions of Lévy processes.

### Frailty equals instantaneous jump of Lévy process:

Assume that  $a(t, v)$  depends only on the argument  $v$ , and that it equals the Dirac delta function in the argument. Then

$$\mu(t) = r(t)\Phi(h(t)).$$


---

## 10.3 Proportional Hazards Derived from Lévy Processes

Aalen and Hjort (2002) derived proportional hazards from Lévy processes. Suppose individuals in group 1 have frailty  $Z_1$  with Laplace transform  $E \exp\{-sZ(u_1)\} = \exp\{-u_1\Phi(s)\}$ , while individuals in group 2 have frailty  $Z_2$  with Laplace transform  $E \exp\{-sZ(u_2)\} = \exp\{-u_2\Phi(s)\}$ . The basic hazard rates in both groups equals  $h(t)$ , with cumulative hazard denoted by  $H(t)$ . The population survival functions in the two groups (i.e., those observed in the population as opposed to the unobservable individual survival function) can then be written

$$S_1(t) = E \exp\{-Z_1 H(t)\} = \exp\{-u_1\Phi(H(t))\}.$$

Similarly  $S_2(t) = \exp\{-u_2\Phi(H(t))\}$ . This implies that the two population hazard rates are given by

$$\mu_1(t) = u_1\Phi'(H(t))h(t) \quad \text{and} \quad \mu_2(t) = u_2\Phi'(H(t))h(t). \quad (10.6)$$

One sees that these two hazard rates are proportional, with  $\nu$  as the proportionality parameter. Clearly, the special Laplace transform of a Lévy process produces this result.

Proportional hazard regression models may be derived by putting a regression structure on the parameter  $u$ . Let a vector covariate  $(y_1, \dots, y_p)$  be given. A Cox model will then arise if  $u$  satisfies

$$u = \exp(\beta_1 y_1 + \dots + \beta_p y_p)$$

for a coefficient vector  $\beta$ . One then assumes that the covariates only influence the model through the parameter  $u$ .

### 10.3.1 Example 10.1: Gamma Process

It follows from the above that if the two distributions are, e.g. two gamma distributions with the same scale, but different shape parameters, then the

population hazard rates will be proportional. The density of gamma process at ‘time’  $u$  is given by

$$f(z; u, \rho, \nu) = \Gamma(\rho u)^{-1} \nu^{\rho u} z^{\rho u - 1} \exp(-\nu z).$$

It is important to note the shapes of these two frailty distributions. Commonly in frailty theory (See, for instance, Aalen, 1994, for a review) one assumes that the frailty distribution of a high-risk and a low-risk group have the same shape, but with a different scale factor. Typically, these results in a decreasing relative population hazard. It might, however, be more likely that the frailty distributions of a high-risk group has a different shape from that of a low-risk group. For instance, considering the risk of lung cancer for smokers and non-smokers, one would not assume that the risk distribution of non-smokers is simply a scaled down version of the risk distribution for smokers. Biologically, the reason why risk varies among individuals would be different in non-smokers (genetical, environmental) from smokers (with amount of smoking playing an important role). Considering the frailty as being built up through a Lévy process gives one way to suggest reasonable and flexible models as to how the frailty distribution would develop with increasing risk.

### **10.3.2 Example 10.2: Compound Poisson Process**

A compound Poisson distribution is a mixture of an atom at zero (when there is no occurrence in the Poisson process) and a continuous distribution. The density of the continuous part at ‘time’  $u$ , when the Poisson rate is 1 and the jump distribution is gamma with scale parameter  $\nu$  and shape parameter  $\rho$ , can be written up directly as

$$f(z; u, \rho, \nu) = \exp(-\nu z - u) \frac{1}{z} \sum_{k=1}^{\infty} u^k \frac{(\nu z)^{\rho k}}{k! \Gamma(\rho k)}.$$

The probability at zero equals  $\exp(-u)$ .

The class of compound Poisson distributions is useful in frailty theory, see Aalen (1992) who also gives a number of plots to illustrate the shape that the continuous density may have. It has nice mathematical properties, and the point mass at zero corresponds to the inclusion of a non-susceptible subgroup of individuals. This is often biologically reasonable, since only a subgroup of individuals may have the genetic makeup which makes them vulnerable to certain diseases. An example of fitting a compound Poisson frailty model to the incidence of testicular cancer is given in Aalen and Tretli (1999).

#### **The squared coefficient of variation:**

The squared coefficient of variation, that is the variance divided by the

square of the expectation, is a useful measure of how much a frailty distribution is spread out. From (10.2) one may derive the squared coefficient of variation as

$$CV^2 = \frac{1}{u} \frac{\Phi''(0)}{\Phi'(0)^2},$$

provided that the variance of the distribution exists. Comparing the distribution for two values of  $u$ , say  $u_1$  and  $u_2$ , gives the relative squared coefficient of variation

$$RCV^2 = \frac{u_1}{u_2}.$$

This may be reasonable measure of relative spread also when the variance does not exist. For instance, for the stable distributions, changing  $u$  corresponds to a scale change of the distribution, which is otherwise independent of  $u$ . From (10.6) it is seen that  $RCV^2$  corresponds to the hazard ratio, giving this concept a simple and explicit interpretation in terms of frailty distributions.

This explicit correspondence between the hazard ratio and the relative spread of the frailty distributions may be useful in interpreting hazard ratios. It may also explain why one does not usually find hazard ratios which are very large; commonly significant ratios may have values of 2 or 3, but usually not much higher. To take an example, if the high-risk group, with the smallest spread, had a  $CV^2$  equal to 1, then the  $CV^2$  in the low-risk group would equal the  $RCV^2$ . An  $RCV^2$  equal to 4, say, would imply a  $CV^2$  equal to 4 in the low-risk group. This means that the standard deviation is twice the mean of the distribution. For non-negative distributions this implies that some of the probability mass is at high values, while much mass is quite close to zero. This again would point to risk factors affecting a part of the population much more strongly than the rest. Genetic risk factors, with major gene effects, might have such a property, while polygenic risks or those depending on continuous risk factors would usually not have such a large effect.

## 10.4 Other Frailty Process Constructions

Let the time of Lévy process be  $cu$  where  $c$  is a random variable. If  $c$  has a stable distribution with parameters  $\alpha$  and  $\beta$ , then the Laplace transform in (10.2) will transform into

$$E \exp\{-cu\Phi(s)\} = \exp(-\alpha u^\beta \Phi(s)^\beta).$$

Considering again two times  $u_1$  and  $u_2$ , the hazard ratio will be  $\left(\frac{u_1}{u_2}\right)^\beta$  and once more there is proportionality between the two groups. This result is closely connected to the nice result of Hougaard (2000) that stable frailty distributions preserve proportionality.

Note that the randomized Lévy process has dependent increments, showing that also processes with dependent increments may give proportionality.

In the construction used above the outcome of the frailty mechanism is of the type  $Z(u)$  for a Lévy process  $Z$ , with the process having run longer for high risk than for low risk groups. A frailty process construction with a different perspective is as follows:

Individuals are pictured as being continuously exposed to an unobserved cumulative damage type process, of the compound Poisson type  $Z(t) = \sum_{j \leq N(t)} G_j$  for  $t \geq 0$ . Here  $G_1, G_2, \dots$  are taken to be iid non-negative variables, interpreted as adding over time to the hazard level of the individual, while  $N(\cdot)$  is a Poisson process with cumulative hazard rate  $H(t) = \int_0^t h(s)ds$ , that is, its increments are independent and Poisson  $h(s)ds$ .

The connection to the person's survival prospects is to model  $S(t|\mathcal{H}_t) = P\{T \geq t|\mathcal{H}_t\}$ , the survival distribution given the full history of what has happened to the person up to time  $t-$ , as

$$S(t|\mathcal{H}_t) = \exp(-Z(t)) = \prod_{j \leq N(t)} \exp(-G_j)$$

with  $k = E \exp(-G_j)$ , it follows that the unconditional survival function must take the form

$$S(t) = E \exp\{-Z(t)\} = Ek^{N(t)} = \exp\{-(1-k)H(t)\}.$$

Note that even though the survival function is discontinuous given the jumps of the unobservable damage process, it becomes continuous marginally, with cumulative hazard rate  $(1-k)H(t)$  and hazard rate function  $\mu(t) = (1-k)h(t)$ . The point in the present context is that we once again have a broad scenario with proportional hazard rates, namely when different groups of individuals have Poisson process rates proportional to each other. For further discussion of such frailty process modeling constructions, including indications of applications to nonparametric Bayesian event history analysis, see Hjort (2002).

## 10.5 Hierarchical Lévy Frailty Models

Moger and Aalen (2006) developed hierarchical Lévy frailty models. We present some of their results here in this section. An important limitation of the shared frailty models is the fact that all members of the family have the same frailty. This can be inappropriate, since one would also expect some individual, variation due to non-shared genes and environmental factors, and different degrees of dependence for different types of relatedness, e.g. siblings, families, neighborhoods. In Moger et al. (2004), they introduced a frailty model

based on the compound Poisson distribution with random scale. By applying a PVF distribution to a scale parameter in the compound Poisson frailty model, one gets a model with variation on both family and individual level. Some further discussion of the properties of the model is found in Moger and Aalen (2005). Since the compound Poisson distribution is included in the family of Lévy frailty distributions, a small extension of the compound Poisson-PVF frailty model can be accomplished by using the more general Lévy frailty distributions for the individual heterogeneity. This yields a hierarchical Lévy model, and it will be discussed here. We will not give any details on the likelihood construction for the different models in this section. However, by using the same techniques as in Moger and Aalen (2005), this should be straightforward also for more complex models.

In Moger and Aalen (2005), the compound Poisson - PVF model was constructed by applying a PVF distribution on  $\rho$  in (10.5). The compound Poisson distribution models the heterogeneity on the individual level, where all individuals have independent frailties. The PVF distribution on  $\rho$  models the family heterogeneity, so that all individuals in a family share a common value of  $\rho$ , thus creating independence between relatives. Individuals from different families are independent. More generally, let  $Z_1$  be the frailty variable for the individual level. This variable will often have independent values for all individuals. Let  $Z_2, \dots, Z_k$  be the frailty variables for higher levels, which will typically be independent for some members of a family, but shared for others. The variable  $Z_k$  may have the same value for all individuals of a family. Let  $Z_i$  follow Lévy distribution with Laplace transform  $L_{Z_i}(s) = \exp(-\rho_i \Phi_i(s))$  and probability distribution  $f_{z_i}$ . Denote the total frailty by  $X$ . Consider only the variation at the bottom level, and let all the other levels be given. Add a new level of frailty by randomizing  $\rho_1$  by  $Z_2$ . The Laplace transform of  $X$  will be

$$L_X(s) = E(L_{z_1}(s)|z_2) = \int \exp(-\rho_1 \Phi_1(s)) f_{z_2}(\rho_1) d\rho_1 = \exp[-\rho_2 \Phi_2(\Phi_1(s))]. \quad (10.7)$$

This is more general version of the model in Moger and Aalen (2005). When combined with a parametric baseline hazard  $h(t)$  in (10.1), one may get large improvement in fit compared to the simpler shared frailty models, since these models use separate distributions for individual and family variation. For non-parametric  $h(t)$ 's, the model will be equivalent to a shared frailty model, since the individual heterogeneity will be subsumed in  $h(t)$ . However, these could perhaps be situations where one would like to model the individual frailty by a specific probability distribution, even when using a non-parametric baseline hazard. The model can be useful for family data on diseases that are hypothesized to be caused by strong, unknown genetic or environmental effects, for which it is impossible to collect covariates information, but for which there exist biological theories on how the disease mechanism works.

This can be hinting at using specific parametric distributions for modeling the baseline hazard and the individual and family heterogeneity in a frailty model, as discussed for testicular cancer in Moger et al. (2004).

The two level Lévy model applies to data on groups where the genetic or environmental association is expected to be equal for all individuals, for instance litters; siblings or brothers. To extend the model to more general pedigrees where subgroups of individuals are more closely correlated than others, add another level to the model by randomizing  $\rho_2$  by  $Z_3$ . This yields the Laplace transform

$$\begin{aligned} L_X(s) &= E[E(L_{z_1}(z)|Z_2, Z_3)] = \int \int \exp[-\rho_1 \Phi_1(s)] f_{z_2}(\rho_1) d\rho_1 f_{z_3}(\rho_2) d\rho_2 \\ &= \exp[-\rho_3 \Phi_3(\Phi_2(\Phi_1(s)))] \end{aligned} \quad (10.8)$$

and so on for further levels. Hence, the structure is generated by applying function iteration to the Laplace exponent. The model described by  $L_3(s)$  could be used on data with two levels of dependence, consisting example of families in a neighborhood. The distribution  $Z_3$  could then describe common environmental factors shared by all individuals in the neighborhood, while  $Z_2$  corresponds to factors which are shared by a family, but independent for different families. The distribution  $Z_1$  models individual environmental factors which are independent for all.

An interesting special case applies when the positive stable distributions are used, that is, when  $\Phi_1(s) = s^{\alpha_i}$  (the scale parameter of the distribution, usually called  $\delta$ , will play the role of  $\rho_i$ ). The Laplace transform of  $X$  is then

$$L_X(s) = \exp[-\rho_3 s^{\alpha_1 \alpha_2 \alpha_3}]$$

which again is the Laplace transform of a stable distribution. This result is presented in Hougaard (2000), p. 354-362, in the section on the multivariate stable model. Moreover, he suggests a trivariate model for the lifetimes  $(T_1, T_2, T_3)$  of a sibling group, where individuals 2 and 3 are monozygotic twins, and individual 1 is a singleton. Hence, sibling 2 is more strongly correlated to sibling 3 than to sibling 1. For the hierarchical Lévy frailty models, this can be constructed as follows. All siblings share the same value of  $Z_2$  (and hence of  $\rho_1$ ). Siblings 2 and 3 will have the same value of  $Z_1$ , whereas sibling 1 will have an independent value of  $Z_1$ . Let different subscripts denote independent values of the  $Z_i$ 's. The joint Laplace transform for the sibling group will then be

$$\begin{aligned} L(s_1, s_2, s_3) &= E[E(\exp[-z_1^1 s_1 - z_1^2 s_2 - z_1^3 s_3]|z_1, z_2)] \\ &= E(\exp[-\rho_1 \Phi_1(s_1) - \rho_1 \Phi_1(s_2 + s_3)]|z_2) \\ &= \exp[-\rho_2 \Phi_2(\Phi_1(s_1) + \Phi_1(s_2 + s_3))]. \end{aligned}$$

Since all siblings are independent given  $z_1$  and  $z_2$ . This gives the joint survival function

$$S(t_1, t_2, t_3) = \exp[-\rho_2 \Phi_2(\Phi_1(H(t_1) + \Phi_1(H(t_2) + H(t_3)))]$$

and generalizes the formula of Hougaard (2000), p. 356.

The model above has a nested dependence structure, where all individuals are correlated. One may also construct a simple genetic model, for data consisting of parents and children. Here, the parents are assumed to be independent, but the children are related both the parents and to each other. Let the parents have independent values of both  $Z_1$  and  $Z_2$ , whereas the children will have independent values of  $Z_1$ , but their value of  $\rho_1$  assigned by  $Z_2$  is determined by those of the parents. In a simple additive model, assume that it is the mean of the parent's values  $(\rho_1^1 + \rho_1^2)/2$ . Let  $s_1$  and  $s_2$  be the argument of the Laplace transform for the parents and  $s_3, \dots, s_k$  the arguments for the children. The joint Laplace transform of this model is given by

$$\begin{aligned} L(s_1, \dots, s_k) &= E \left( \exp \left[ -\rho_1^1 \Phi(s_1) - \rho_1^2 \Phi_1(s_2) - \sum_{j=3}^k \frac{(\rho_1^1 + \rho_1^2)}{2} \Phi_1(s_j) \right] | z_2 \right) \\ &= E \left( \exp \left\{ -\rho_1^1 [\Phi_1(s_1) + \frac{1}{2} \sum_{j=3}^k \Phi_1(s_j)] - \rho_1^2 [\Phi_1(s_2) + \frac{1}{2} \sum_{j=3}^k (s_j)] \right\} | z_2 \right) \\ &= \exp \left\{ -\rho_2 [\Phi_2(\Phi_1(s_1) + \frac{1}{2} \sum_{j=3}^k \Phi_1(s_j)) + \Phi_2(\Phi_1(s_2) + \frac{1}{2} \sum_{j=3}^k \Phi_1(s_j))] \right\}. \end{aligned}$$

A drawback with this model, is that the frailty distribution of the children will have a different variance than the parent's distribution.

In the baseline hazard  $h(t)$  in Eq. (10.1) includes a scale parameter, one often sets the expectation of the frailty distribution equal to one, to assure identifiability. Simple results are valid for the expectation and variance of the hierarchical Lévy frailty model, provided that they exist for the model in question. Assume that the expectation of the  $Z_i$ 's equals one when  $\rho_i = 1$ , that is, we have  $\Phi_i(0) = 1$  for all  $i$ . For the variable in Eq. (10.7), we have

$$E(X) = \Phi'_2(\Phi_1(0))\Phi'_1(0) = 1$$

and

$$\begin{aligned} \text{Var}(X) &= \Phi''_2(\Phi_1(0))(\Phi'_1(0))^2 + \Phi'_2(\Phi_1(0))\Phi''_1(0) \\ &= \Phi''_2(0) + \Phi''_1(0) = \text{Var } Z_1 + \text{Var } Z_2. \end{aligned}$$

By induction it follows that the random variable in (10.7) has expectation and variance  $E(X) = 1$ ,  $\text{Var}(X) = \text{Var}(Z_1) + \text{Var}(Z_2) + \text{Var}(Z_3)$  and similarly for higher levels. Hence, the variance of a hierarchical Lévy frailty variable can be decomposed into a sum coming from different sources, without affecting the expectation. This is very useful in a frailty context, where the expectation

often should be kept constant and just the variance be decomposed. As an example,  $\text{Var } Z_1$  can be interpreted as the frailty variance related to individual factors,  $\text{Var } Z_2$  is the frailty variance related to common genetic and environmental factors within a family, and  $\text{Var } Z_3$  is the frailty variance relating to common environmental factors in the neighborhood.

### 10.5.1 Application to the Infant Mortality Data

The Medical Birth Registry of Norway has recorded all births in Norway (population around 4.5 million) since 1967, from the 16th week of gestation onward. By the 31st December 1998, 1,986,576 births were recorded. Information on all deaths occurring during the first year of life registered by Statistics Norway is linked to the birth records. By use of the national identification number on the mothers, the births may be linked into siblings. We do not consider the father. The average size of the siblings is about two, slightly higher than the average number of children per woman in Norway, which is around 1.8. The proportion of women without any children by the age of 40, has increased from ca. 10% for the 1935-cohort to ca. 13% for the 1960-cohort (from Statistics Norway's Web pages).

The database includes some covariates, most of which are known to have an influence on infant mortality. These are birth weight, gestational age, infant's birth year, mother's birth year, length, mother's age, parity, and gender. The proportion of missing data is fairly small, ranging from 0.2% for birth weight and 2.4% for length, to 5.9% for gestational age and 6.0% for gender. There are no missing values for the other covariates. For missing data in the continuous covariates, the mean value are imputed in the analysis.

This application is an illustration of different aspects when working with hierarchical Lévy frailty models, more than finding the best model for the infant mortality data. All siblings with one or more cases are included in the sample. The control siblings are stratified according to family size before sampling, and exactly 5% are randomly sampled without replacement from each stratum. There are four strata for the control families, for siblings of size 1, 2, 3, and  $>4$ . This yields 45,750 siblings with 89,745 individuals in the control sample. Stratifying according to sibship size is important to get good precision in the estimated frailty and baseline hazard parameters. The precision of these parameters are mainly decided by the number of familial cases and the prevalence of the disease, and one gets a more precise estimate of the latter by the stratification. According to the results in Moger et al. (in revision), this should give an efficiency of almost 100% for the frailty and baseline hazard parameters, compared to a cohort analysis using the same model. The precision of the regression effects will naturally be much lower, perhaps around 70-75%, but this is sufficient as an illustration of the model. To account for the fact that a case-cohort sample is analyzed, sampling weights will enter the likelihood, yielding a standard pseudo-likelihood. Moger and Aalen (2006) used the compound Poisson (CP)-gamma model to analyze the

data and compared the results for some of the different covariate modeling options from a case-cohort analysis of the data by using a two-level Lévy model.



## Part III

# Bivariate Frailty Models for Survival Data



# Chapter 11

---

## *Bivariate Frailty Models and Estimation Methods*

11.1 Introduction .....	205
11.2 Bivariate Frailty Models and Laplace Transforms .....	206
11.3 Proportional Hazard Model for Covariate Effects .....	207
11.4 The Problem of Confounding .....	207
11.5 A General Model of Covariate Dependence .....	208
11.6 Pseudo-Frailty Model .....	210
11.7 Likelihood Construction .....	211
11.8 Semiparametric Representations .....	212
11.9 Estimation Methods in Bivariate Frailty Models .....	214
11.9.1 Two-Stage Estimation Method .....	214
11.9.2 Basegroup Estimation Method .....	214
11.9.3 Estimation Methods Based on the EM Algorithm .....	215
11.9.4 Profile Estimation for Frailty Models .....	217
11.9.5 Profile Estimation for Transformation Models .....	219
11.9.6 Conclusions .....	221

---

### 11.1 Introduction

Iachine (2001) discussed the estimation method for bivariate frailty models. Maximum likelihood is not the only possible estimation method for survival models with random effects (Korsgaard et al., 1998; Yau and McGilchrist, 1998; Hougaard 2000, p. 274). However, it has several appealing features which is why it is popular. First, the presence of asymptotic consistency and asymptotic normality properties for finite-dimensional parametric models is a good starting point for estimation and inference. For semiparametric models the results of Murphy (1995), Parner (1998), and Korsholm (1999) justify the semiparametric estimation and inference procedures in gamma-frailty models. For more general frailty models the work of Murphy and Vaart (2000) on profile likelihood may provide a foundation for semiparametric maximum likelihood estimation.

Second, maximum likelihood techniques form a good combination with random effect models describing individual characteristics, e.g. models of individual frailty or major gene models. The advantage becomes clear where performing a joint analysis of several datasets (e.g. identical (MZ) or fraternal (DZ) twins, twins, and singletons, etc). In case of maximum likelihood the

combination of such data is straightforward - one has to multiply the respective likelihood functions and use the same parameters for identical individual characteristics. An example of this technique is the combination of MZ and DZ twins using a correlated frailty model where one uses the same frailty variance for MZ and DZ twins but different correlation parameters (Yashin and Iachine, 1995a). Another example is the combination of data from the three Nordic twin registers (Iachine et al., 1998).

---

## 11.2 Bivariate Frailty Models and Laplace Transforms

In the following we consider bivariate frailty models. The methodology can be easily extended to multivariate situations. For details see Yashin and Iachine (1995c), Peterson et al. (1996), Korsgaard and Andersen (1998).

Let  $T_1, T_2$  be the lifespans of two twins and let  $Z_1, Z_2$  be their *individual* frailties satisfying:

$$P(T_i > t_i | Z_1, Z_2) = P(T_i > t_i | Z_i), \quad i = 1, 2. \quad (11.1)$$

Assume that the twin's lifespans  $T_1, T_2$  are conditionally independent given  $Z_1, Z_2$ .

$$\begin{aligned} S(t_1, t_2 | Z_1, Z_2) &= P(T_1 > t_1, T_2 > t_2 | Z_1, Z_2) \\ &= P(T_1 > t_1 | Z_1)P(T_2 > t_2 | Z_2) \\ &= e^{-Z_1 H(t_1)} e^{-Z_2 H(t_2)} \end{aligned} \quad (11.2)$$

where  $H(t) = \int_0^t h_0(u)du$  is the cumulative baseline hazard function. The marginal survival function  $S(t_1, t_2)$  can be obtained by averaging the conditional bivariate survival function over the joint distribution of  $Z_1, Z_2$ :

$$\begin{aligned} S(t_1, t_2) &= E[S(t_1, t_2 | Z_1, Z_2)] \\ &= E[e^{-Z_1 H(t_1)} e^{-Z_2 H(t_2)}] \\ &= L(H(t_1), H(t_2)) \end{aligned} \quad (11.3)$$

where  $L(s_1, s_2)$  is the bivariate Laplace transform of  $Z_1, Z_2$ . When  $Z_1$  and  $Z_2$  are independent, (11.3) reduces to

$$S(t_1, t_2) = L_{Z_1}(H(t_1))L_{Z_2}(H(t_2)).$$

When  $Z_1$  and  $Z_2$  are dependent, this model will become correlated frailty model. The model corresponding to the case  $Z_1 = Z_2 = Z$  is called shared frailty model (Clayton 1978, Hougaard 1986a) as given below.

$$\begin{aligned} S(t_1, t_2) &= E[e^{-Z[H(t_1) + H(t_2)]}] \\ &= L(H(t_1) + H(t_2)) \end{aligned}$$

In the following it will be assumed that the bivariate Laplace transform of the frailty distribution may be parameterized by a finite-dimensional parameter  $\theta$ . The cumulative hazard function  $H(t)$  is left unspecified.

The assumption of conditional independence plays an important role in genetic analysis of longevity. Under this assumption the frailty variable becomes the only mediator of genetic influence on survival. It means, that the underlying hazard represents only non-genetic influence on lifespan.

---

### 11.3 Proportional Hazard Model for Covariate Effects

The bivariate frailty models discussed above do not include any effects of observed covariates. In epidemiological studies a proportional hazards model for time-independent covariates is often used. Hence, let  $\mathbf{Y} = (Y_1, \dots, Y_m)^T$  be an observed covariate vector. Then

$$h(t|Z, \mathbf{Y}) = r(\mathbf{Y})Zh_0(t), \quad (11.4)$$

where  $r(\mathbf{y})$  is a non-negative parametric risk function which depends on the parameter  $\beta$  (e.g.  $r(\mathbf{y}) = e^{\beta^T \mathbf{y}}$ ).

Assuming that  $(\mathbf{Y}_1, \mathbf{Y}_2)$  is independent of  $(Z_1, Z_2)$  the bivariate survival function for the frailty model conditional on  $\mathbf{Y}_1, \mathbf{Y}_2$  can be expressed as:

$$S(t_1, t_2|\mathbf{Y}_1, \mathbf{Y}_2) = L(r(\mathbf{Y}_1)H(t_1), r(\mathbf{Y}_2)H(t_2)). \quad (11.5)$$


---

### 11.4 The Problem of Confounding

It was realized that a combination of frailty and covariate effects in a proportional hazard model

$$h(t|Z, \mathbf{Y}) = r(\mathbf{Y})Zh_0(t) \quad (11.6)$$

leads to the so-called “confounding problem” (Hougaard, 1985; Yashin et al., 2001; Iachine, 2001). This may be illustrated by the following setup. The bivariate frailty model

$$S_{\theta, \beta}(t_1, t_2|\mathbf{Y}_1, \mathbf{Y}_2) = L_\theta(r_\beta(\mathbf{Y}_1)H(t_1), r_\beta(\mathbf{Y}_2)H(t_2)) \quad (11.7)$$

is characterized by the parameter vector  $\beta$  of the risk function  $r(y)$  and by the parameter vector  $\theta = (\theta_1, \dots, \theta_m)$  of the bivariate frailty distribution. Assuming symmetry the univariate submodel of model (11.7) is given by:

$$S_{\theta, \beta}(t|\mathbf{Y}) = L_\theta(r_\beta(\mathbf{Y})H(t), 0) = L_\theta(0, r_\beta(\mathbf{Y})H(t)). \quad (11.8)$$

If the univariate model (11.8) is identifiable (which is the case when  $EZ < \infty$ ), analysis of univariate data alone allows to recover the respective parameter vectors  $\theta$  and  $\beta$ .

There have been several attempts to repair this problem. The approach developed by Hougaard (1987) avoids the identifiability property of the univariate model (11.8) by using a positive stable frailty distribution with infinite expectation. Yashin et al. (1995) partially eliminates the univariate identifiability by suggesting a correlated frailty model. In this model only the variance of frailty is identifiable from univariate data. The degree of lifespan dependence is controlled by correlation of frailty which requires bivariate data for identification.

The root of the problem, however, is in the proportional hazards assumption (11.6) between frailty and the covariate effect. A radical solution of the confounding problem is to assume a general dependence of the baseline hazard on the covariates (Iachine 2001):

$$h(t|Z, \mathbf{Y}) = Z h_0(t|\mathbf{Y}). \quad (11.9)$$

The specification of the general dependence of the baseline hazard on the covariates is not straightforward but approaches based on the semiparametric representation presented below may simplify the problem.

---

## 11.5 A General Model of Covariate Dependence

A general dependence of the baseline hazard on the observed covariates may be combined with an arbitrary dependence of the frailty distribution on the covariates (i.e. zygosity, marker genotypes, etc). The possibility of studying dependence between frailty and covariates is based on better identifiability properties of bivariate frailty models as compared to univariate ones. In general, univariate frailty models do not allow for identification of frailty-covariate dependence. The resulting model is described below.

Let  $T_1, T_2$  be the survival times and let  $Z_1, Z_2$  be the individual frailties of the two twins. Let  $\mathbf{Y}_1, \mathbf{Y}_2$  be the observed covariate vectors for twin 1 and 2 such that

$$L(s_1, s_2; \mathbf{Y}_1, \mathbf{Y}_2, \theta) | \theta \in \Theta \subseteq \mathbb{R}^p \quad (11.10)$$

is a parametric family of Laplace transforms of the conditional distribution of  $(Z_1, Z_2)$  given  $\mathbf{Y}_1, \mathbf{Y}_2$  where  $\theta$  is a finite-dimensional parameter. Let

$$\{L_i(s; \mathbf{Y}, \theta) | \theta \in \Theta \subseteq \mathbb{R}^p\}, \quad i = 1, 2 \quad (11.11)$$

be respective Laplace transform families for the conditional distributions of  $Z_i$  given  $\mathbf{Y}_i$ .

Assume that the conditional hazard given the covariates and frailty is given by (Iachine 2001):

$$h_i(t|Z_i, \mathbf{Y}_i) = Z_i h_{0i}(t|\mathbf{Y}_i), \quad i = 1, 2 \quad (11.12)$$

where  $h_{0i}(t|\mathbf{Y})$  is an unknown baseline hazard function. Under the conditional independence assumption the conditional bivariate survival function for  $T_1, T_2$  given the frailties and covariates may be written as:

$$S(t_1, t_2|Z_1, Z_2, \mathbf{Y}_1, \mathbf{Y}_2) = e^{-Z_1 H_1(t_1|\mathbf{Y}_1) - Z_2 H_2(t_2|\mathbf{Y}_2)} \quad (11.13)$$

where  $H_i(t|\mathbf{Y}) = \int_0^t h_{0i}(u|\mathbf{Y}) du$  is the cumulative baseline hazard. Averaging over the conditional distribution of  $(Z_1, Z_2)$  given  $(\mathbf{Y}_1, \mathbf{Y}_2)$  yields

$$S(t_1, t_2|\mathbf{Y}_1, \mathbf{Y}_2) = L(H_1(t_1|\mathbf{Y}_1), H_2(t_2|\mathbf{Y}_2); \mathbf{Y}_1, \mathbf{Y}_2, \theta). \quad (11.14)$$

The respective univariate conditional survival functions are given by

$$S_i(t|\mathbf{Y}) = L_i(H_i(t|\mathbf{Y}); \mathbf{Y}, \theta). \quad (11.15)$$

Since a Laplace transform  $L(s)$  is a monotonously decreasing function of  $s$  (when the distribution is not concentrated in 0) the inverse function  $L^{-1}(u)$  exists and satisfies:

$$H_i(t|\mathbf{Y}) = L_i^{-1}(S_i(t|\mathbf{Y}); \mathbf{Y}, \theta). \quad (11.16)$$

This relationship may be used to obtain the so-called *semiparametric representation* of the bivariate survival function (11.14).

$$S(t_1, t_2|\mathbf{Y}_1, \mathbf{Y}_2)$$

$$= L(L_1^{-1}(S_1(t_1|\mathbf{Y}_1); \mathbf{Y}_1, \theta), L_2^{-1}(S_2(t_2|\mathbf{Y}_2); \mathbf{Y}_2, \theta); \mathbf{Y}_1, \mathbf{Y}_2, \theta). \quad (11.17)$$

This representation may be simplified by defining a bivariate *copula* (Schweizer and Sklar, 1983), which describes the bivariate dependence structure of  $T_1, T_2$  given  $\mathbf{Y}_1, \mathbf{Y}_2$ :

$$C(s_1, s_2|\mathbf{Y}_1, \mathbf{Y}_2, \theta)$$

$$= L(L_1^{-1}(s_1; \mathbf{Y}_1, \theta), L_2^{-1}(s_2; \mathbf{Y}_2, \theta); \mathbf{Y}_1, \mathbf{Y}_2, \theta). \quad (11.18)$$

The bivariate survival function can then be expressed as

$$S(t_1, t_2|\mathbf{Y}_1, \mathbf{Y}_2) = C(S_1(t_1|\mathbf{Y}_1), S_2(t_2|\mathbf{Y}_2)|\mathbf{Y}_1, \mathbf{Y}_2, \theta). \quad (11.19)$$

The main idea behind this representation is that it might be easier to model  $S_i(t|\mathbf{Y})$  than  $H_i(t|\mathbf{Y})$  since it is possible to estimate  $S_i(t|\mathbf{Y})$  directly from the data on  $T$  and  $\mathbf{Y}$  whereas to obtain an estimate of  $H_i(t|\mathbf{Y})$  requires the knowledge of the unknown parameter  $\theta$ . This decomposition of the model serves as a motivation for the so-called two-stage estimation procedures originally suggested by Hougaard (1986a).

It is clear, that if no specific assumption are made about the structure of  $h_0(t|\mathbf{Y})$  the model (11.19) is not identifiable from univariate data and therefore does not suffer from the “confounding problem”.

---

## 11.6 Pseudo-Frailty Model

The semiparametric representation (11.19) still leaves one open question: how should one model the univariate survival functions  $S_i(t|\mathbf{Y})$ ? One possible solution is to introduce additional “pseudo-frailty” variables  $\tilde{Z}_i, i = 1, 2$  which would describe deviations of  $S_i(t|\mathbf{Y})$  from a proportional hazards model with respect to the covariates. In other words we have as in (11.5):

$$S_i(t|\mathbf{Y}) = \tilde{L}_i(r(\mathbf{Y})\tilde{H}(t); \tilde{\theta}) \quad (11.20)$$

where  $\tilde{L}_i(s; \tilde{\theta})$  is the Laplace transform of  $\tilde{Z}_i$  and  $\tilde{\theta}$  is a parameter vector. It is clear, that when  $L_i = \tilde{L}_i$  the model takes the form (11.5). On the other hand, if  $Var(\tilde{Z}_i) = 0$  (and  $E\tilde{Z}_i = 1$ ), the parameterization turns into

$$S_i(t|\mathbf{Y}) = \exp(-r(\mathbf{Y})\tilde{H}(t)). \quad (11.21)$$

In particular, the case  $r(\mathbf{Y}) = e^{\beta^t \mathbf{Y}}$  corresponds to the Cox regression model.

The “pseudo-frailty” parameterization (11.20) is a generalization of the marginal modeling approach discussed by Hougaard (1986a), Petersen (1998), and Scheike et al. (1999). This approach has also been used to analyze dependent survival data in situations where dependence is a nuisance parameter (Lin and Wei, 1989; Lee et al., 1992). One important advantage of the “pseudo-frailty” parameterization is that it permits the inclusion of model (11.5) as a particular case of model (11.14) when  $L_i = \tilde{L}_i, i = 1, 2$ . Consequently, this property allows us to unite the so-called conditional and the marginal approaches for modeling covariate dependence. The appropriateness of the conditional approach vs. the marginal approach may be verified by testing the hypothesis

$$H_0 : L_i = \tilde{L}_i, \quad i = 1, 2. \quad (11.22)$$

A “pseudo-frailty” model for  $S_i(t|\mathbf{Y})$  is not the only possible choice. A more advantageous approach could be based on accelerated failure time models

(Hougaard et al., 1994; Keiding et al., 1997; Hougaard, 1999a). The choice of the univariate parameterization has some important implications for estimation and inference.

---

## 11.7 Likelihood Construction

Let  $T_1, T_2$  be non-negative random variables representing lifespans of a pair of twins and let  $\mathbf{Y}_1, \mathbf{Y}_2$  be respective explanatory variables or covariates. In survival analysis models for bivariate survival times are frequently specified in terms of the conditional bivariate survival function:

$$S(x_1, x_2; \mathbf{y}_1, \mathbf{y}_2, \theta) = P_\theta(T_1 > x_1, T_2 > x_2 | \mathbf{Y}_1 = \mathbf{y}_1, \mathbf{Y}_2 = \mathbf{y}_2) \quad (11.23)$$

where  $\theta = (\theta_1, \dots, \theta_p)^T$  is a vector of parameters. This specification of the model in terms of the survival function simplifies the computational burden of handling right-censoring and truncation in a maximum likelihood framework.

In case of right censoring, the lifespans  $X_1, X_2$  are not observed. Instead we observe  $(T_1, \Delta_1), (T_2, \Delta_2)$ , where

$$\begin{aligned} T_i &= \min(X_i, U_i) \\ \Delta = I(X_i \leq U_i) &= \begin{cases} 1, & \text{if } X_i \leq U_i \\ 0, & \text{if } X_i > U_i \end{cases} \end{aligned} \quad (11.24)$$

for  $i = 1, 2$ . Here  $U_1, U_2$  are non-negative random variables called “censoring times”.

The problem of truncation may be described by introducing the so-called *entry times* or *truncation times*  $T_{0,1}, T_{0,2}$  which are non-negative random variables where  $T_{0,i}$  is the entry time of the  $i$ -th twin. It is only possible to observe the failure times (either censored or uncensored) of the two twins if they both occur after the respective entry times, i.e. the conditions

$$\begin{cases} \min(X_1, U_1) > T_{0,1} \\ \min(X_2, U_2) > T_{0,2} \end{cases}$$

must hold.

The contribution of a right censored and left truncated observation of survival times of the two twins to the conditional likelihood given the truncation times and covariates may be written as

$$L(t_1, \delta_1, t_2, \delta_2 | t_{0,1}, t_{0,2}, \mathbf{y}_1, \mathbf{y}_2) = (-1)^{\delta_1 + \delta_2} \frac{\frac{\partial^{\delta_1 + \delta_2}}{\partial t_1^{\delta_1} \partial t_2^{\delta_2}} S(t_1, t_2; \mathbf{y}_1, \mathbf{y}_2, \theta)}{S(t_{0,1}, t_{0,2}; \mathbf{y}_1, \mathbf{y}_2, \theta)} \quad (11.25)$$

provided that censoring and truncation are independent and non-informative

given the covariates. These assumptions are quite reasonable for Danish, Swedish, and Finnish twin survival data where truncation is conditionally deterministic given date of birth and where right censoring occurs mainly due to end of the follow up period, in which case it is also conditionally deterministic given the birthdate.

Expression (11.25) illustrates why the concept of modeling bivariate survival in terms of survival functions is important in analysis of right-censored and truncated data - the likelihood can be expressed in terms of the survival function and its partial derivatives. An alternative approach based on the modeling of bivariate probability density functions requires evaluation of integrals over the density function, a computationally much more time consuming procedure. One can able to compute  $S(t_1, t_2; \mathbf{y}_1, \mathbf{y}_2, \theta)$  and its partial derivatives efficiently in order to use them in an optimization procedure. In the case of frailty models, the calculation of the marginal survival function is based on the Laplace transform of the bivariate frailty distribution.

Maximization of the likelihood function composed of contributions of the form (11.25) is a far from trivial task. Fortunately there exist a number of well-studied methods for numerical optimization (e.g., Newton-Raphson method, Broyden-Fletcher-Goldfarb-Shanno method). Several of them are implemented in the *Maximum Likelihood Package of GAUSS Programming Language* by Aptech Systems, Inc. (1995) which was used to obtain maximum likelihood estimates in the analysis.

---

## 11.8 Semiparametric Representations

Consider the problem of estimation in the bivariate model which does not include any covariate effects. If the cumulative hazard function  $H(t)$  belongs to some parametric family then the estimation is straightforward.

However, it may be difficult to choose a parametric representation of the cumulative baseline hazard  $H(t)$  because it is unobserved. It turns out that it is possible to avoid the specification of the underlying hazard functions and instead parameterize the marginal survival functions  $S_1(t), S_2(t)$  by rewriting the frailty model in the so-called *semiparametric form*. It is frequently easier to find a good parametric representation for  $S_1(t), S_2(t)$  since its fit can be assessed by, for example, Kaplan-Meier estimates. To simplify the presentation let us assume that the marginal distributions of  $Z_1$  and  $Z_2$  are identical. In this case we have that  $S_1(t) = S_2(t) = S(t)$  and the latter is given by:

$$S(t) = S(t, 0) = L(H(t), 0) = L_Z(H(t)) \quad (11.26)$$

where  $L_Z(s)$  is the Laplace transform of the marginal frailty distribution.

When the frailty distribution is not degenerate (i.e., the distribution is not

concentrated in 0) the inverse function  $L_Z^{-1}(u)$  is well-defined. One can then express  $H(t)$  from (11.26) as follows:

$$H(t) = L_Z^{-1}(S(t)) \quad (11.27)$$

and then substitute (11.27) into the expression for the marginal bivariate survival function:

$$S(t_1, t_2) = L(L_Z^{-1}(S(t_1)), L_Z^{-1}(S(t_2))). \quad (11.28)$$

Equation (11.28) is called a *semiparametric representation* of a bivariate survival function.

Note that one cannot always find an analytic expression for the inverse function  $L_Z^{-1}(u)$ . In such cases the inverse can be calculated numerically at the expense of additional computing time and possible numerical stability problems. Fortunately, for gamma frailty distribution and a more general power variance family (Hougaard, 1986b) frailty distribution such analytic expressions do exist (Yashin et al., 1995; Yashin et al., 1999).

This is not the case for the major gene model (Begin et al., 2000) where the Laplace transform of frailty has the form

$$L_Z(s) = \sum_{i=1}^n p_i e^{-r_i s}. \quad (11.29)$$

The same problem reoccurs for the quadratic hazard model (Iachine and Yashin, 1999) where the Laplace transform is given by

$$L_Z(s) = (1 + 2s)^{-\frac{1}{2}} e^{-\frac{m^2}{2}(1 - \frac{1}{1+2s})}. \quad (11.30)$$

In both cases the respective inverse function is computed by solving (11.29) and (11.30) with respect to  $s$  using the bisection method.

The semiparametric representation (11.28) plays an important role in estimation. Yashin et al. (1993) showed that incorrect specification of the baseline hazard may lead to biased estimates of the association parameters. To deal with this problem they suggested to parametrize the marginal survival function instead of the baseline hazard. For this purpose the so-called *gamma-Makeham parametrization* (Yashin et al., 1994) of the univariate survival function was used. This parametrization has the following form

$$S(t; a, b, c, s) = (1 + s^2(\frac{a}{b}e^{bt} - 1) + ct))^{-\frac{1}{s^2}}.$$

It was subsequently used in genetic studies of Danish (Yashin and Iachine, 1995c), Swedish (Yashin et al., 1999) and Finnish (Iachine et al., 1998) twins.

## 11.9 Estimation Methods in Bivariate Frailty Models

### 11.9.1 Two-Stage Estimation Method

Another use of representation (11.28) is related to two-stage estimation procedures suggested by Hougaard (1986a). The idea of this approach is to estimate the univariate survival functions on the first stage (e.g., using the Kaplan-Meier method or a univariate frailty or accelerated failure time model in the presence of observed covariates). In the second stage, the estimated functions are substituted into (11.28) and the remaining association parameters are estimated as if the univariate survival functions were known and fixed. However, the estimate of the covariance matrix of the association parameters obtained at stage two needs correction in order to take into account the additional uncertainty related to the estimation of the univariate characteristics. This problem has been addressed by Shih and Louis (1995), Glidden (2000), and Andersen (2000).

Application of the two-stage estimation method to truncated bivariate survival data (e.g., Swedish or Finnish twin survival data) is more complicated. Under bivariate truncation the life spans of the first twin  $T_1$  are sampled from the conditional distribution of  $T_1$  given  $\{T_1 > t_0, T_2 > t_0\}$  and not from the conditional distribution of  $T_1$  given  $\{T_1 > t_0\}$  as in univariate survival analysis ( $t_0$  is the entry time). Consequently, an uncritical application of the traditional methods of univariate survival analysis (e.g., the Kaplan-Meier method) to truncated twin survival data on the first stage of the two-stage procedure may produce biased estimates of the univariate survival characteristics and therefore invalidate the results of stage two.

Another issue which require further investigation is the possible loss of efficiency associated with the two-stage procedure in comparison to single-stage approaches.

### 11.9.2 Basegroup Estimation Method

Consider the frailty model (11.5) which incorporates the effects of observed covariates by using a proportional hazards assumption. If the distribution of the explanatory variable has an atom  $\{\mathbf{Y} = \mathbf{y}_0\}$ , one might avod the specification of the underlying hazard  $H(t)$  by using an extension of the semiparametric representation method called the *basegroup estimation method*.

Application of this method requires that we can find a suitable parametrization of  $S_0(t) = S(t|\mathbf{Y} = \mathbf{y}_0)$  (e.g., the *gamma-Makeham* parametrization above), the survival function for the population *basegroup* that is represented by the atom of  $\mathbf{Y}$ 's distribution, since in this case it is possible to verify the fit of the basegroup survival function parametrization

$S_0(t)$ . Solving for  $H(t)$  yields:

$$H(t) = \frac{1}{r(\mathbf{y}_0)} L_Z^{-1}(S_0(t)) \quad (11.31)$$

where  $r(\mathbf{Y}_0)$  is function of regression coefficients and explanatory variables. Substituting (11.31) into the bivariate model we obtain

$$S(t_1, t_2 | \mathbf{Y}_1, \mathbf{Y}_2) = L \left( \frac{r(\mathbf{Y}_1)}{r(\mathbf{y}_0)} L_Z^{-1}(S_0(t_1)), \frac{r(\mathbf{Y}_2)}{r(\mathbf{y}_0)} L_Z^{-1}(S_0(t_2)) \right). \quad (11.32)$$

The method was applied by Yashin et al. (1996) and Yashin and Iachine (1997) to estimate the effects of birth year on survival of the Danish Twins and by Herskind et al. (1996) in the analysis of the Danish Twins survey data.

### 11.9.3 Estimation Methods Based on the EM Algorithm

The case of a covariate with an absolutely continuous distribution in model (11.5) may be handled by the application of the EM algorithm, that allows to obtain a semiparametric estimate for  $H(x)$ .

The EM algorithm is an iterative parameter estimation procedure used in the analysis of incomplete data to calculate maximum likelihood parameter estimates. Initial ideas of the approach are discussed by Yates (1933). Orchard and Woodbury (1972) suggested the missing information principle and discuss ideas of how to calculate the observed information matrix. Dempster et al. (1977) introduced the term “EM-algorithm”. The issues of convergence were discussed by Wu (1983) and Boyles (1983). Louis (1982) developed techniques for calculation of the observed information matrix. These and other methods and approaches of dealing with missing data were summarized by Little and Rubin (1987). Clayton and Cuzick (1985) suggested a modified version of the EM algorithm for parameter estimation in the bivariate shared frailty model with gamma distributed frailty and an observed covariate. Gill (1985) suggested an idea of developing the EM algorithm using Cox’s partial likelihood. This idea was implemented by Nielsen et al. (1992) and by Klein (1992) who applied this technique to Framingham family data using a version of the shared frailty model with gamma-distributed frailty. A version of the EM-algorithm for the analysis of the correlated gamma-frailty model was developed by Iachine (1995b) and Petersen et al. (1996) and tested on simulated data. Parner (1998) showed consistency and asymptotic normality of the nonparametric maximum likelihood estimator for the correlated gamma-frailty model with observed covariates. Yashin and Iachine (1996) suggested a version of the EM algorithm for the analysis of the quadratic hazard model.

In the following we briefly present the main ideas behind the EM algorithm in the context of bivariate frailty models. Let  $T$  be the observed data and let  $Z$  be the unobserved data such that the complete loglikelihood of  $(T, Z)$  is  $l(t, z; \theta)$  where  $\theta$  is a parameter vector. A single iteration of the EM-algorithm consists of the E-step and M-step:

**E-step:**

$$Q(\theta; \theta_i) = E(l(t, Z; \theta) | t, \theta_i) \quad (11.33)$$

**M-step:**

$$\theta_{i+1} = \operatorname{argmax}_{\theta} Q(\theta; \theta_i) \quad (11.34)$$

where  $\theta_i$  are the parameters from the previous iteration. Under some additional regularity conditions  $\theta_i$  converges to  $\hat{\theta}$ , the maximum likelihood estimate of  $\theta$  based on observed data.

In the case of a bivariate frailty model the observed information is represented by censored survival time  $(T_1, \Delta_1), (T_2, \Delta_2)$  as defined in (11.24) and covariates  $\mathbf{Y}_1, \mathbf{Y}_2$  and the unobserved information by the frailty variables  $Z_1, Z_2$ . Let  $g(z_1, z_2)$  is the pdf and  $L(s_1, s_2)$  the Laplace transform for  $Z_1, Z_2$ . Define:

$$L_{\delta_1, \delta_2}(s_1, s_2) = (-1)^{\delta_1 + \delta_2} \frac{\partial^{\delta_1 + \delta_2}}{\partial s_1^{\delta_1} \partial s_2^{\delta_2}} L(s_1, s_2) \quad (11.35)$$

the complete likelihood for  $(t_1, \delta_1, z_1, t_2, \delta_2, z_2)$  (i.e. as if frailty had been observed) is:

$$\prod_{i=1}^2 [r(\mathbf{y}_i) z_i h_0(t_i)]^{\delta_i} e^{-r(\mathbf{y}_i) z_i H(t_i)} g(z_1, z_2). \quad (11.36)$$

Averaging over the bivariate frailty distribution yields the likelihood for  $(t_1, \delta_1, t_2, \delta_2)$ :

$$\prod_{i=1}^2 [r(\mathbf{y}_i) h_0(t_i)]^{\delta_i} L_{\delta_1, \delta_2}(r(\mathbf{y}_1) H(t_1), r(\mathbf{y}_2) H(t_2)). \quad (11.37)$$

The probability density function of  $(Z_1, Z_2)$  given  $(T_1, \Delta_1, T_2, \Delta_2)$  can be obtained by an application of Bayes' formula:

$$\frac{\prod_{i=1}^2 z_i^{\delta_i} e^{-r(\mathbf{y}_i) z_i H(t_i)} g(z_1, z_2)}{L_{\delta_1, \delta_2}(r(\mathbf{y}_1) H(t_1), r(\mathbf{y}_2) H(t_2))}. \quad (11.38)$$

The logarithm of the complete likelihood is given by

$$l(t, \delta, z) = \sum_{i=1}^2 (\delta_i [\ln r(\mathbf{y}_i) + \ln z_i + \ln h_0(t_i)] - r(\mathbf{y}_i) z_i H(t_i)) + \ln g(z_1, z_2). \quad (11.39)$$

Assume that the parameters of the frailty distribution and the risk function are separated, i.e.:

- $\theta$  - parameters of  $Z_1, Z_2$ , i.e.  $g(z_1, z_2; \theta)$
- $\beta$  - parameters of  $r(\mathbf{y})$ , i.e.  $r(\mathbf{y}; \beta)$
- $H(t)$  - functional parameter.

It follows from (11.39) that the functional  $Q$  decomposes into two terms:

$$Q(\theta, \beta, H; \theta_i, \beta_i, H_i) = Q_1(\beta, H; \theta_i, \beta_i, H_i) + Q_2(\theta; \theta_i, \beta_i, H_i). \quad (11.40)$$

**E-Step:**

The first term  $Q_1$  involves computation of the conditional expectations:

$$E[Z_1|T_1, \delta_1, \mathbf{Y}_1, T_2, \delta_2, \mathbf{Y}_2] = \frac{L_{\delta_1+1, \delta_2}(r(\mathbf{y}_1)H(t_1), r(\mathbf{y}_2)H(t_2))}{L_{\delta_1, \delta_2}(r(\mathbf{y}_1)H(t_1), r(\mathbf{y}_2)H(t_2))} \quad (11.41)$$

for twin 1 and

$$E[Z_2|T_1, \delta_1, \mathbf{Y}_1, T_2, \delta_2, \mathbf{Y}_2] = \frac{L_{\delta_1, \delta_2+1}(r(\mathbf{y}_1)H(t_1), r(\mathbf{y}_2)H(t_2))}{L_{\delta_1, \delta_2}(r(\mathbf{y}_1)H(t_1), r(\mathbf{y}_2)H(t_2))} \quad (11.42)$$

for twin 2 (computation of  $E[\ln Z_i]$ ) is not required because it is a constant with respect to the parameters).

**M-step:**

It follows from (11.40) that to maximize  $Q$  one can maximize  $Q_1$  independently of  $Q_2$  because of the parameter separation. The term  $Q_1$  is similar to the likelihood of the Cox's regression model with logarithms of the expectations (11.41) and (11.42) as additional covariates with a known regression coefficient (i.e., as "offset terms"). Therefore, it can be maximized using Cox's partial likelihood producing  $\beta_{i+1}$  and  $H_{i+1}(t)$  in the form of the so-called *Breslow* estimate.

The term  $Q_2$  (which involves the conditional expectation of a log-probability density function) depends on the chosen frailty distribution and it seems that there is no simple relation between  $Q_2$  and the Laplace transform. For the correlated gamma-frailty model and the quadratic hazard model these expectations can be calculated (Iachine, 1995b; Yashin and Iachine, 1996). Consequently, the complete EM-algorithm is only appropriate for estimation in those frailty models (11.5) where it is feasible to evaluate the term  $Q_2$ .

#### 11.9.4 Profile Estimation for Frailty Models

It turns out that the ideas of EM algorithm may be extended to allow for semiparametric estimation in arbitrary frailty models (11.5) specified in terms of parametric Laplace transform families using the *profile likelihood* approach of Nielsen et al. (1992).

This extension is based on the fact that when the parameters of the frailty distribution are fixed an *incomplete EM-algorithm* (i.e., where the M-step consists of maximization of  $Q_1$  only) may be used to obtain an estimate of  $H(t)$  which may then be used to maximize the profile likelihood with respect to  $\theta$ . As shown above, this method involves only the Laplace transform of the frailty distribution and its partial derivatives. A more detailed description of the method is given below, where we in addition allow for dependence between frailty and covariates.

Let  $L(s_1, s_2; \mathbf{y}_1, \mathbf{y}_2, \theta)$  be the conditional bivariate Laplace transform of  $Z_1, Z_2$  given  $\mathbf{Y}_1 = \mathbf{y}_1, \mathbf{Y}_2 = \mathbf{y}_2$  where  $\theta$  is a vector of parameters. Note that model (11.5) involves the proportional hazard assumption (11.4).

As it was shown above, the incomplete EM-algorithm may be used to obtain a semiparametric estimate of  $H(t)$  assuming that  $\theta$  and  $\beta$  are known by iteratively solving the following equation with respect to  $\hat{H}(t)$ :

$$\hat{H}(t) = \sum_{t_{i,j} \leq t} \delta_{i,j} \times$$

$$\left[ \sum_{t_k, t \geq t_{i,j}} r(\mathbf{y}_{k,l}) \bar{Z}_k(r(\mathbf{y}_{1,l}) \hat{H}(t_{1,l}), r(\mathbf{y}_{2,l}) \hat{H}(t_{2,l}); \delta_{1,l} \delta_{2,l}, \mathbf{y}_{1,i}, \mathbf{y}_{2,i}, \theta) \right]^{-1} \quad (11.43)$$

where

$$\bar{Z}_k(s_1, s_2; \delta_1, \delta_2, \mathbf{y}_1, \mathbf{y}_2, \theta) = \frac{L_{\delta+I(k=1), \delta_2+I(k=2)}(s_1, s_2; \mathbf{y}_1, \mathbf{y}_2, \theta)}{L_{\delta_1, \delta_2}(s_1, s_2; \mathbf{y}_1, \mathbf{y}_2, \theta)}. \quad (11.44)$$

The iterative procedure is organized in the following way: first a suitable initial estimate is chosen (e.g., Nelson-Aalen estimate of the cumulative hazard function), next to obtain an updated estimate one has to apply the right-hand-side of (11.43) to the estimate from the previous iteration. The procedure is then repeated until convergence. Equation (11.43) may be viewed as a Breslow cumulative hazard estimate in a Cox regression procedure with  $\bar{Z}_k$  included as an offset term. It results from the M-step of the EM-algorithm and was first introduced for the gamma-frailty model (Nielsen et al., 1992; Murphy, 1994; Iachine, 1995b; Petersen et al., 1996) and the quadratic hazard frailty model (Yashin and Iachine, 1996).

This procedure yields an estimate of  $\hat{H}(t; \theta, \beta)$  for each fixed value of  $\theta, \beta$ . This allows us to define a profile likelihood that only depends on  $\theta, \beta$ :

$$L_p(\theta, \beta) = \prod_{i=1}^n \left[ r(\mathbf{y}_{1,i})^{\delta_{1,i}} \Delta \hat{H}(t_{1,i}; \theta, \beta)^{\delta_{1,i}} r(\mathbf{y}_{2,i})^{\delta_{2,i}} \Delta \hat{H}(t_{2,i}; \theta, \beta)^{\delta_{2,i}} \times L_{\delta_{1,i}, \delta_{2,i}}(r(\mathbf{y}_{1,i}) \hat{H}(t_{1,i}; \theta, \beta), r(\mathbf{y}_{2,i}) \hat{H}(t_{2,i}; \theta, \beta) \mathbf{y}_{1,i}, \mathbf{y}_{2,i}, \theta) \right] \quad (11.45)$$

where  $\Delta \hat{H}(t) = \hat{H}(t) - \hat{H}(t-)$ . It is this profile likelihood that has to be maximized to obtain estimates of other parameters of the model. For the gamma-frailty model it has been shown that these estimates have attractive asymptotic properties (Murphy, 1995; Parner, 1998) and that the likelihood can be used for inference in the usual way (Korsholm, 1999). For other models the asymptotic properties may be based on the results of Murphy and van der Vaart (2000) provided the necessary regularity conditions hold.

### 11.9.5 Profile Estimation for Transformation Models

The ideas of profile likelihood and incomplete EM algorithm can be further developed to derive a general semiparametric estimation method for a special class of bivariate survival models defined below. In particular, this method is appropriate for semiparametric estimation in model (11.17) where the univariate survival functions are parameterized using the “pseudo-frailty” model (11.20).

Assume that the bivariate survival function can be expressed in the following form:

$$S(t_1, t_2) = D_\theta(H_1(t_1), H_2(t_2)) \quad (11.46)$$

where  $D_\theta(h_1, h_2)$  is a bivariate survival function with parameter  $\theta$  and  $H_i(t)$  are some unknown “cumulative hazard” functions in the sense, that they are absolutely continuous, zero at  $t = 0$  and non-decreasing for all  $t \geq 0$ , i.e.  $H_i(t) = \int_0^t h_i(u) du$ . Here  $D_\theta$  represents the parametric part of the model assuming that  $\theta$  is finite-dimensional and  $H(t)$  represents the non-parametric or functional parameter part of the model. Clearly, the cumulative hazard functions define transformations of the time scale for  $t_1$  and  $t_2$ . The parametric part  $D_\theta$  could, in principle, depend on observed time-independent covariates, but this dependence is omitted to simplify the notation.

The contribution of a single observation to the partial likelihood function under independent censoring and truncation is given by:

$$\begin{aligned} L(\theta, H_1, H_2) &= (-1)^{\delta_1 + \delta_2} \frac{\frac{\partial^{\delta_1 + \delta_2}}{\partial t_1^{\delta_1} \partial t_2^{\delta_2}} S_\theta(t_1, t_2)}{S_\theta(t_{0,1}, t_{0,2})} \\ &= (-1)^{\delta_1 + \delta_2} \frac{D_\theta^{\delta_1, \delta_2}(H_1(t_1), H_2(t_2))}{D_\theta(H_1(t_{0,1}), H_2(t_{0,2}))} h_1(t_1)^{\delta_1} h_2(t_2)^{\delta_2}. \end{aligned} \quad (11.47)$$

If the cumulative hazard functions  $H_1$  and  $H_2$  are known or belong to some parametric family (e.g., Gompertz) the estimation of the parameter  $\theta$  is straightforward. Now assume that  $\theta$  is known and let us concentrate on the problem of estimating  $H_1$  and  $H_2$ .

Note that bivariate frailty models are a special case of (11.47) in which case  $D_\theta$  is the bivariate Laplace transform of  $(Z_1, Z_2)$ . As it is the case for the frailty model we note that a maximizer of (11.47) cannot be found in the original class of absolutely continuous cumulative hazards. The parameter space for  $H(t)$  is therefore extended to a class of piece-wise constant cumulative hazard functions with jumps at observed uncensored survival times:

$$H_i(t) = \sum_{t_{i,j} \leq t} \Delta H_{i,j}, \quad i = 1, 2. \quad (11.48)$$

Now it is the size of a finite number of jumps that has to be estimated. The semiparametric analog of the partial likelihood contribution (11.47) is

now proportional to

$$L(\theta, H_1, H_2) = (-1)^{\delta_1 + \delta_2} \frac{D_\theta^{\delta_1, \delta_2}(H_1(t_1), H_2(t_2))}{D_\theta(H_1(t_{0,1}), H_2(t_{0,2}))} \Delta H_1(t_1)^{\delta_1} \Delta H_2(t_2)^{\delta_2}. \quad (11.49)$$

The log partial likelihood of the data is given by:

$$\begin{aligned} l(\theta, H_1, H_2) &= \sum_{i=1}^n (\delta_{1,i} \ln \Delta H_{1,i} + \delta_{2,i} \ln \Delta H_{2,i}) \\ &+ \sum_{i=1}^n \ln(-1)^{\delta_{1,i} + \delta_{2,i}} D_\theta^{\delta_{1,i}, \delta_{2,i}}(H_1(t_{1,i}), H_2(t_{2,i})) \\ &- \sum_{i=1}^n \ln D_\theta(H_1(t_{0,1,i}), H_2(t_{0,2,i})). \end{aligned} \quad (11.50)$$

The stationary point of the loglikelihood satisfies  $\frac{\partial l}{\partial \Delta H_{r,k}} = 0$  where

$$\begin{aligned} \frac{\partial l}{\partial \Delta H_{r,k}}(\theta, H_1, H_2) &= \frac{\delta_{r,k}}{\Delta H_{r,k}} \\ &+ \sum_{i=1}^n \frac{\partial}{\partial h_r} \ln(-1)^{\delta_{1,i} + \delta_{2,i}} D_\theta^{\delta_{1,i}, \delta_{2,i}}(H_1(t_{1,i}), H_2(t_{2,i})) \frac{\partial H_r(t_{r,i})}{\partial \Delta H_{r,k}} \\ &- \sum_{i=1}^n \frac{\partial}{\partial h_r} \ln D_\theta(H_1(t_{0,1,i}), H_2(t_{0,2,i})) \frac{\partial H_r(t_{0,r,i})}{\partial \Delta H_{r,k}}. \end{aligned} \quad (11.51)$$

It follows from (11.48) that

$$\frac{\partial H_r(t)}{\partial \Delta H_{r,k}} = 1(t_{r,k} \leq t). \quad (11.52)$$

The estimate of  $\Delta H_{r,k}$  must therefore satisfy:

$$\frac{\delta_{r,k}}{\Delta H_{r,k}} + A_{r,k}(H_1, H_2) = 0 \quad (11.53)$$

where

$$\begin{aligned} A_{r,k}(H_1, H_2) &= \sum_{i:t_{r,k} \leq t_{r,i}} K_\theta^{\delta_{1,i}, \delta_{2,i}, r}(H_1(t_{1,i}), H_2(t_{2,i})) \\ &- \sum_{i:t_{r,k} \leq t_{0,r,i}} K_\theta^{0,0,r}(H_1(t_{0,1,i}), H_2(t_{0,2,i})) \end{aligned} \quad (11.54)$$

and

$$K_\theta^{\delta_1, \delta_2, r}(h_1, h_2) = \frac{\partial}{\partial h_r} \ln(-1)^{\delta_1 + \delta_2} D_\theta^{\delta_1, \delta_2}(h_1, h_2). \quad (11.55)$$

Since the cumulative hazard estimate only jumps at uncensored survival times the estimate of  $\Delta H_{r,k}$  solves

$$\Delta H_{r,k} = -\frac{1}{A_{r,k}(H_1, H_2)} \quad (11.56)$$

for  $\delta_{r,k} = 1$  and is zero when  $\delta_{r,k} = 0$ . The cumulative hazard estimate satisfies:

$$\hat{H}_i(t) = \sum_{t_{i,j} \leq t} -\frac{1}{A_{i,j}(\hat{H}_1^k, \hat{H}_2^k)}. \quad (11.57)$$

This relationship may be used to define an iterative procedure used to find the estimate of  $H_i$ . Let  $\hat{H}_1^k, \hat{H}_2^k$  be the estimates from the  $k$ -th iteration. The next estimate may be obtained as follows:

$$\hat{H}_i^{k+1}(t) = \sum_{t_{i,j} \leq t} -\frac{1}{A_{i,j}(\hat{H}_1^k, \hat{H}_2^k)}, \quad i = 1, 2. \quad (11.58)$$

The procedure is then repeated until convergence. After achieving convergence the profile likelihood may be constructed by substituting the obtained estimates  $\hat{H}_1^k, \hat{H}_2^k$  into (11.49).

It should be noted, however, that convergence properties of this procedure are not completely clear and require further investigation. The use of profile likelihood in semiparametric models has been investigated by Murphy and Vaart (2000), but its validity for model (11.46) remains to be shown (e.g., by the verification of respective regularity conditions).

The ‘pseudo-frailty’ model suggested above is a special case of the transformation model (11.46). In this case the parametric part of the model has the form

$$D_{\theta, \bar{\theta}}(h_1, h_2 | \mathbf{Y}_1, \mathbf{Y}_2) = L(L_1^{-1}(\tilde{L}_1(r(\mathbf{Y}_1)h_1; \tilde{\theta}); \mathbf{Y}_1, \theta), L_2^{-1}(\tilde{L}_2(r(\mathbf{Y}_2)h_2; \tilde{\theta}); \mathbf{Y}_2, \theta); \mathbf{Y}_1, \mathbf{Y}_2, \theta). \quad (11.59)$$

The idea behind this approach is to avoid the need for correction of the estimated covariance matrix of the parameter estimates required when applying the two-stage procedure and to deal with bivariate truncation.

### 11.9.6 Conclusions

Bivariate frailty models allow for a number of estimation strategies developed around the maximum likelihood methodology. The estimation methods rely on a parametric specification of the Laplace transform of the conditional frailty distribution given the covariates and allow for a number of semiparametric approaches for estimation of the baseline hazard.

It seems that methods based on semiparametric representation provide the

greatest flexibility while avoiding the “confounding problem”. However, two-stage approaches offer simple estimation techniques but at an additional cost of the required correction of the standard errors, which is often not a trivial task. Other methods of profile estimation for general transformation models require further investigation.

# **Chapter 12**

---

## ***Correlated Frailty Models***

12.1	Introduction .....	223
12.2	Correlated Gamma Frailty Model .....	225
12.3	Correlated Power Variance Function Frailty Model .....	225
12.4	Genetic Analysis of Duration .....	226
12.4.1	Correlation Coefficient .....	227
12.4.2	Six Genetic Models of Frailty .....	229
12.4.3	Danish Twins Survival Data .....	230
12.4.4	Results: Genetics of Frailty and Longevity .....	231
12.5	General Bivariate Frailty Model .....	233
12.5.1	Gamma Model .....	234
12.5.2	Lognormal Model .....	236
12.5.3	Estimation Strategies .....	237
12.6	Correlated Compound Poisson Frailty for the Bivariate Survival Lifetimes	238
12.7	Applications .....	240

---

### **12.1 Introduction**

Frailty models are becoming increasing popular in multivariate survival analysis. Shared frailty models in particular are often used despite their limitations. To overcome their disadvantages numerous correlated frailty models were established during the last decade. In the present study, we examine correlated frailty models, and especially the behavior of the parameter estimates when using different estimation strategies.

Shared frailty models explain correlations within groups (family, litter, or clinic) or for recurrent events facing the same individual. However, this approach does have limitations. First, it forces unobserved factors to be the same within the cluster, which is not generally acceptable. For example, sometimes it may be inappropriate to assume that both partners in a twin pair share all of their unobserved risks. Second, the dependence between survival times within the cluster is based on their marginal distributions. To see this, when covariates are present in a proportional hazard model with a gamma distributed frailty, the dependence parameter and the population heterogeneity are confounded (see Clayton and Cuzick, 1985), implying that the multivariate distribution of the life times can be identified from the marginal distributions of these life times (see Hougaard, 1986a). Elbers and Ridder (1982) show that this problem applies to any univariate frailty distribution with a finite mean. Third, in most cases, shared frailty will only induce a positive association

within the group. However, there are some situations in which survival times for subjects within the same cluster are negatively associated. For example, this applies to growth rates for animals in the same litter that have a limited food supply.

To avoid these limitations, correlated frailty models are being developed for the analysis of multivariate failure time data, in which associated random variables are used to characterize the frailty effect for each cluster. In twin pairs, for example, one random variable is assigned to twin 1 and another to twin 2, so that they are no longer constrained to have a common frailty. These two variables are associated and jointly distributed, therefore, knowing one of them does not automatically imply the other. Also, these two variables certainly can be negatively associated, which would then induce a negative association between survival times.

Correlated frailty models provide not only variance parameters of the frailties as in shared frailty models, but they also contain additional parameter for modeling the correlation between frailties in each group.

Frequently one is interested in construction of a bivariate extension of some univariate family distributions (e.g., gamma). For example, for the purpose of genetic analysis of frailty one might be interested in estimation of correlation of frailty. It turns out that it is possible to carry out such extension for the class of infinitely-divisible distributions (Iachine 1995a, 1995b). In this case an additional parameter representing the correlation coefficient of the bivariate frailty distribution is introduced.

**Theorem 12.1** *Let  $Z$  be an infinitely divisible frailty variable with Laplace transformation  $L_Z(s)$ . Let  $\rho \in [0, 1]$ , then there exist random variables  $Z_1, Z_2$  each with univariate Laplace transform  $L_Z(s)$  such that the Laplace transform of  $Z_1, Z_2$  is given by:*

$$L(s_1, s_2) = L_Z^\rho(s_1 + s_2) L_Z^{1-\rho}(s_1) L_Z^{1-\rho}(s_2). \quad (12.1)$$

*If  $Z$  has a variance the  $\text{Corr}(Z_1, Z_2) = \rho$ .*

The respective bivariate survival model is identifiable under mild regularity conditions on  $Z$  provided that  $\rho > 0$ . The case  $\rho = 1$  is known as the shared frailty model. Versions of shared frailty model were analyzed by Clayton and Cuzick (1985), Vaupel et al. (1992) in the case of gamma frailty and Hougaard (1987) in the case of positive stable frailty among others.

The presence of a parameter representing the correlation between the frailty variables is crucial for genetic analysis of frailty, in particular for heritability estimation, since standard twin analysis methods of quantitative genetics are based on the analysis of correlations for MZ and DZ twins (Falconer, 1965; Neale and Cardon, 1992). The idea of genetic analysis of frailty using twin data is presented in more detail in Yashin and Iachine (1995) and Iachine et al. (1998).

## 12.2 Correlated Gamma Frailty Model

The correlated gamma frailty model is based on a bivariate extension of the univariate gamma distribution using additive independent components, as described in the previous section. The marginal distributions of  $(Z_1, Z_2)$  satisfy  $Z_i \sim \text{Gamma}(\frac{1}{\sigma^2}, \frac{1}{\sigma^2})$ , i.e.,  $Z_i$  is gamma distributed with mean 1, and variance  $\sigma^2$  and Laplace transform

$$L(s; \sigma^2) = (1 + \sigma^2 s)^{\frac{1}{\sigma^2}}. \quad (12.2)$$

The idea of correlated frailty yields the following bivariate Laplace transform:

$$L(s_1, s_2; \sigma^2, \rho) = (1 + \sigma^2(s_1 + s_2))^{-\frac{\rho}{\sigma^2}} (1 + \sigma^2 s_1)^{-\frac{1-\rho}{\sigma^2}} (1 + \sigma^2 s_2)^{-\frac{1-\rho}{\sigma^2}}. \quad (12.3)$$

where  $\text{corr}(Z_1, Z_2) = \rho$ . Yashin and Iachine (1994) introduced a more general form of the correlated gamma frailty model by allowing for different variances of frailty  $\text{var}(Z_i) = \sigma_i^2$ ,  $i = 1, 2$ . The respective semiparametric representation for the bivariate survival function is given by

$$S(t_1, t_2) = S_1(t_1)^{1-\frac{\sigma_1}{\sigma_2}\rho} S_2(t_2)^{1-\frac{\sigma_2}{\sigma_1}\rho} (S_1(t_1)^{-\sigma_1^2} + S_2(t_2)^{-\sigma_2^2} - 1)^{-\frac{\rho}{\sigma_1\sigma_2}}, \quad (12.4)$$

where  $0 \leq \rho \leq \min(\frac{\sigma_1}{\sigma_2}, \frac{\sigma_2}{\sigma_1})$ . This model was first applied to the analysis of Danish twin survival data (Yashin et al., 1993; Yashin and Iachine, 1995a, 1995b, 1997). Later similar analyses have been performed for Swedish, Finnish, and Danish data (Yashin et al., 1999; Iachine et al., 1998). The properties of the correlated gamma frailty model were also studied by Pickles and Crouchley (1994).

## 12.3 Correlated Power Variance Function Frailty Model

To investigate the sensitivity of heritability estimates to the choice of the frailty distribution the correlated frailty model was extended by applying the ideas of correlated frailty to the power variance function (PVF) family of frailty distributions. The univariate version of this distribution was introduced to frailty modeling by Hougaard (1986b).

In the correlated PVF family model the marginal distributions of  $(Z_1, Z_2)$  satisfy  $Z_i \sim \text{PVF}(\alpha, \delta = (\frac{1-\alpha}{\sigma^2})^{1-\alpha}, \theta = \frac{1-\alpha}{\sigma^2})$ , i.e.,  $Z_i$  is PVF distributed with mean 1, variance  $\sigma^2$ , and Laplace transform

$$L(s; \alpha, \sigma^2) = \exp\left\{-\frac{1-\alpha}{\alpha\sigma^2}((1 + \frac{\sigma^2}{1-\alpha}s)^\alpha - 1)\right\}. \quad (12.5)$$

The bivariate Laplace transform for the correlated PVF frailty model is given by

$$\begin{aligned} L(s_1, s_2; \alpha, \sigma^2, \rho) &= \exp\left\{-\rho \frac{1-\alpha}{\alpha\sigma^2} \left((1 + \frac{\sigma^2}{1-\alpha}(s_1 + s_2))^\alpha - 1\right)\right\} \times \\ &\quad \exp\left\{-(1-\rho) \frac{1-\alpha}{\alpha\sigma^2} \left((1 + \frac{\sigma^2}{1-\alpha}s_1)^\alpha - 1\right)\right\} \times \\ &\quad \exp\left\{-(1-\rho) \frac{1-\alpha}{\alpha\sigma^2} \left((1 + \frac{\sigma^2}{1-\alpha}s_2)^\alpha - 1\right)\right\} \end{aligned} \quad (12.6)$$

where  $\text{corr}(Z_1, Z_2) = \rho$ .

The correlated frailty model with PVF frailty distribution is characterized by the bivariate survival function of the form:

$$\begin{aligned} S(t_1, t_2) &= S(t_1)^{1-\rho} S(t_2)^{1-\rho} \exp\left(\frac{\rho(1-\alpha)}{\alpha\sigma^2} \times \right. \\ &\quad \left[1 - ((1 - \frac{\alpha\sigma^2}{1-\alpha} \ln S(t_1))^{\frac{1}{\alpha}} + (1 - \frac{\alpha\sigma^2}{1-\alpha} \ln S(t_2))^{\frac{1}{\alpha}} - 1)^\alpha\right]\). \end{aligned} \quad (12.7)$$

This model has been used in the analysis of the Danish Twin survival data by Yashin et al. (1999). The results confirm the presence of genetic influence on frailty. Note that the bivariate gamma distribution (used in the correlated gamma frailty model) and the inverse Gaussian distribution are special cases of the three-parameter bivariate frailty distribution. This gives us the opportunity to compare the analysis of the same data using different frailty models. The results show that the gamma frailty model and the extended PVF model fit the Danish twin data equally well.

## 12.4 Genetic Analysis of Duration

In traditional applications of quantitative genetics to twin data the correlation coefficients  $\rho_{mz}$  and  $\rho_{dz}$  of the phenotypic trait calculated for MZ and DZ twins play an important role. Using these coefficients, the narrow sense heritability  $a^2$  may be estimated as:

$$a^2 = 2(\rho_{mz} - \rho_{dz}). \quad (12.8)$$

Note that  $a^2$  corresponds to the proportion of total variance of the phenotypic trait associated with additive genetic effects. The proportion of variance associated with common environmental effects,  $c^2$ , can be calculated as:

$$c^2 = 2\rho_{dz} - \rho_{mz}. \quad (12.9)$$

These results follow a simple model, where the phenotypic trait is represented as the sum of additive genetic, shared environmental, and unshared environmental effects (Falconer, 1990; McGue et al., 1993). All components are supposed to be independent.

Using this approach, heritability estimates can be calculated for the human life span (McGue et al., 1993). However, these estimates do not clarify the genetic structure of the longevity, which is a mean (not a variance) of the life span distribution. To study the role of genetic and environmental factors in human longevity, Yashin and Iachine (1994a) suggest an approach based on the genetic model of frailty rater than of life span. The approach exploits an important advantage of survival model with frailty: additive decomposition of frailty on genetic and environmental components induces the competing risks structure of the respective survival model. As a result, observed mortality becomes a sum of the two terms: one depends on genetic and another on environmental parameters. These parameters can be estimated from bivariate data. It is natural to interpret these terms as senescent (biological) and premature mortality components and to relate the biological limit of human longevity to survival distribution corresponding to the senescent component of normality. Below, we derive a semiparametric representation for the bivariate survival function in the correlated frailty model with the genetic decomposition of frailty.

#### 12.4.1 Correlation Coefficient

Let  $T_i$  and  $Z_i$  ( $i = 1, 2$ ) be the life spans and the frailties of the two individuals who are twins; i.e., their individual hazards are represented by the proportional hazards model  $\mu(Z_i, t) = Z_i \mu_0(t)$ ,  $i = 1, 2$ . We assume that  $T_1$  and  $T_2$  are conditionally independent given  $Z_1$  and  $Z_2$  and that  $Z_i = \sum_{j=1}^n Z_{ij}$  ( $i = 1, 2$ ), where  $Z_{ij}$  and  $Z_{im}$  are independent (for any  $j, m, j \neq m$ ), non-negative, gamma-distributed, random variables constructed as follows. Let  $Z_{1j} = Y_{0j} + Y_{1j}$  and  $Z_{2j} = Y_{0j} + Y_{2j}$ , where  $Y_{ij}, i = 0, 1, 2$  and  $j = 1, 2, \dots, n$  are independent, non-negative, gamma-distributed, random variables:  $Y_{0j} \sim G(k_{0j}, \lambda)$  and  $Y_{1j}, Y_{2j} \sim G(k_{1j}, \lambda)$ . It is obvious that  $Z_{1j}$  and  $Z_{2j}$  are correlated. Let  $\rho_j$  be their correlation coefficient and  $\sigma_{ij}^2$  be the variance of  $Z_{ij}$ . We assume that  $\sigma_{ij}^2 = \sigma_j^2, i = 1, 2$  and  $j = 1, 2, \dots, n$  and  $E(Z_i) = 1, i = 1, 2$ . A model with different variances of bivariate frailty distribution is discussed by Yashin and Iachine (1994). If  $\sigma_{0j}^2$  is the variance of  $Y_{0j}$ , then  $\rho_j = \sigma_{0j}^2 / \sigma_j^2$ . Thus, the variances of  $Z_1$  and  $Z_2$  coincide and can be represented as  $\sum_{j=1}^n \sigma_j^2 = \sigma_z^2$ . Denote  $\gamma_j = \sigma_j / \sigma_z$  and let  $\rho_z$  be the correlation coefficient of bivariate frailty distribution. It follows from the definition

of  $Z_i$  ( $i = 1, 2$ ) that

$$\rho_z = \frac{\sum_{j=1}^n \sigma_{0j}^2}{\sigma_z^2}.$$

Taking into account the definition of  $\rho_j$  and  $\gamma_j$ , we get

$$\rho_z = \sum_{j=1}^n \rho_j \gamma_j^2. \quad (12.10)$$

The parameters  $k_{ij}$ ,  $i = 0, 1$ ;  $j = 1, 2, \dots, n$ ; and  $\lambda$  are represented as functions of  $\rho_j$ ,  $\gamma_j$ , and  $\sigma_z^2$  as follows:

$$\lambda = \frac{1}{\sigma_z^2}, \quad k_{0j} = \rho_j \left( \frac{\gamma_j}{\sigma_z} \right)^2, \quad k_{1j} = (1 - \rho_j) \left( \frac{\gamma_j}{\sigma_z} \right)^2, \quad j = 1, 2, \dots, n. \quad (12.11)$$

Given  $Y_{ij}$  ( $i = 0, 1, 2$ ;  $j = 1, 2, \dots, n$ ), the conditional bivariate survival function  $S(t_1, t_2 | \{Y_{ij}, i = 0, 1, 2, \dots, n\})$  is given by:

$$\begin{aligned} S(t_1, t_2 | \{Y_{ij}, i = 0, 1, 2; j = 1, 2, \dots, n\}) \\ = \exp \left( - \sum_{j=1}^n (Y_{0j} + Y_{1j}) H(t_1) - \sum_{j=1}^n (Y_{0j} + Y_{2j}) H(t_2) \right) \end{aligned} \quad (12.12)$$

where  $H(t) = \int_0^t h_0(u) du$ . Averaging equation (12.12) we obtain the marginal bivariate survival function  $S(t_1, t_2)$ :

$$\begin{aligned} S(t_1, t_2) &= (1 + \sigma_z^2 (H(t_1) + H(t_2)))^{-\left(1/\sigma_z^2\right) \sum_{j=i}^n \rho_j \gamma_j^2} \\ &\times (1 + \sigma_z^2 H(t_1))^{-\left(1/\sigma_z^2\right) \sum_{j=i}^n (1 - \rho_j) \gamma_j^2} \\ &\times (1 + \sigma_z^2 H(t_2))^{-\left(1/\sigma_z^2\right) \sum_{j=i}^n (1 - \rho_j) \gamma_j^2}. \end{aligned} \quad (12.13)$$

Taking into account that  $H(t) = (S(t)^{\sigma_z^2} - 1)/\sigma_z^2$  we obtain semiparametric representation

$$S(t_1, t_2) = \frac{S(t_1)^{1-\rho_z} S(t_2)^{1-\rho_z}}{(S(t_1)^{-\sigma_z^2} + S(t_2)^{-\sigma_z^2} - 1)^{\rho_z/\sigma_z^2}} \quad (12.14)$$

where  $\rho_z = \sum_{j=1}^n \rho_j \gamma_j^2$ . Note that in applications parameters  $\rho_j$  are usually known. They are determined by the level of relationship between related individuals. The properties of the correlated frailty model are discussed by Yashin et al. (1993b) and Yashin and Iachine (1993a, 1993b). The semiparametric structure of equation (12.14) gives us an opportunity to use the two-step

procedure for parametric estimation. First, we can estimate univariate survival functions nonparametrically using, say, Kaplan-Meier estimator. Then, we replace univariate survival functions in the likelihood based on equation (12.14) by their estimators and apply the standard maximum likelihood procedure for estimation of  $\gamma_j$  ( $j = 1, 2, \dots, n$ ) and  $\sigma_z^2$ . We call this procedure “semiparametric” here. This term should not be mixed with “semiparametric estimation” related to Cox’s regression analysis of survival data. Methods of parametric estimation related to the correlated frailty model are discussed by Yashin and Iachine (1994).

### 12.4.2 Six Genetic Models of Frailty

Here we consider the application of model (12.14) with condition (12.10) to the statistical analysis of survival data for MZ and DZ twins. In particular, we calculate semiparametric estimates for the six genetic models of frailty which correspond to six different assumptions about its structure. We refer to the notations used in McGue et al. (1993) and in Yashin and Iachine (1994a) for such models. To be more specific, let  $A, D, I, C, E$ , and  $H$  be the five components of frailty which represent additive genetic effects, dominance genetic effects, epistatic genetic effects, common environmental effects, uncommon environmental effects, and total genetic effects, respectively, in additive decomposition of frailty. From the estimation point of view not more than three components should be represented in the model when MZ and DZ data are analyzed. More components can be considered if, in addition, data about adopted children and twins reared apart are available. In these notations an ACE model refers to the decomposition of frailty  $Z = A + C + E$ . An AE model refers to the decomposition  $Z = A + E$ , ADE, DE, DCE, and HE models are defined similarly. We use small letters  $a^2, d^2, i^2, c^2, e^2$  instead of  $\gamma_j$  ( $j = 1, 2, \dots, n$ ) to refer to the respective proportions of variance. For example, the relationship

$$1 = a^2 + c^2 + e^2 \quad (12.15)$$

correspond to the decomposition of variance in the ACE model of frailty. In Yashin and Iachine (1994a) they use the assumption that the correlation coefficient of the bivariate frailty distribution for this model admits decomposition:

$$\rho = \rho_1 a^2 + \rho_4 c^2 + \rho_5 e^2 \quad (12.16)$$

where  $a^2$  is the proportion of the variance associated with additive genetic effects, called narrow-sense heritability;  $c^2$  is the proportion of variance associated with shared environmental factors;  $e^2$  is the proportion of variance with nonshared environmental factors; and  $\rho_1, \rho_4$ , and  $\rho_5$  are correlations between additive genetic, common environmental, and uncommon environmental components for related individuals. (Similarly,  $\rho_2$  and  $\rho_3$  are correlation coefficients between dominant genetic and epistatic genetic components, respectively.)  $H^2$  is used for the broad-sense heritability coefficient  $H^2 = a^2 + d^2 + i^2$ . Respective

equations for proportions of variance and correlation coefficient in the HCE model are:

$$1 = H^2 + c^2 + e^2 \quad (12.17)$$

$$\rho = RH^2 + \rho_4 c^2 + \rho_5 e^2 \quad (12.18)$$

where  $R$  is the correlation coefficient between total genetic components of the phenotype. We use representation (12.14) for the marginal bivariate survival function with respective decompositions of the correlation coefficients of frailty. Standard assumptions of the quantitative genetics specify different value of  $\rho_i$  ( $i = 1, 2, \dots, 5$ ) and  $R$  for MZ and DZ twins. For MZ twins  $\rho_i = 1$  ( $i = 1, 2, 3, 4$ ),  $\rho_5 = 0$ ,  $R = 1$ . For DZ twins  $\rho_1 = 0.5$ ,  $\rho_2 = 0.25$ ,  $\rho_3 = m$ ,  $\rho_4 = 1$ ,  $\rho_5 = 0$ ,  $R = k$ . Here  $0 \leq m \leq 0.25$  and  $0 \leq k \leq 0.5$  are unknown parameters. Note that the one more important assumption is that the variances of the phenotypic traits for MZ and DZ twins are the same. This hypothesis was tested and confirmed with Danish twin data for both sexes.

### 12.4.3 Danish Twins Survival Data

Yashin and Iachine (1995a) presented the following application of correlated frailty model. To illustrate the approach to the genetic analysis of durations we use the survival times of Danish identical (MZ) and fraternal (DZ) male and female twins born between 1870 and 1900, who survived at age 30. For information about the Danish Twin Registry (see Hauge, 1981). All together we consider 470 male MZ twin pairs, 780 male DZ twin pairs, 475 female MZ twin pairs, and 835 female DZ twin pairs. For both male and female MZ twins life expectancy is higher than for DZ twins (72.38 and 75.30 vs. 71.96 and 73.88, respectively). For the same cohorts of the Danish population, life expectancy is 72 years for males and 73 years for females. Empirical estimates of correlation coefficients show higher association between life spans of MZ than of DZ twins. Pearson's correlation for MZ twins is almost three times higher than for DZ twins for both sexes. The mean absolute values of differences between the life spans are about 2 years smaller for MZ than for DZ twins. Statistical tests (the  $\chi^2$  test and the log-rank test, for example) support the hypothesis about the same chances of survival for twin individuals from different zygotic groups and ordinary individuals in the Danish population for both sexes after age 30. This similarity plays an important role in the justification of statistical procedures for analyzing integrated data for MZ and DZ twins. The likelihood-ratio test also supports the hypotheses about the similarity of respective distributions of the life span. Univariate distributions, however, cannot capture the association between the life spans of related individuals. For these purposes, bivariate distributions of independent life spans are needed. A useful model of such bivariate distribution is based on the correlated frailty concept. Genetic analysis of human longevity involves genetic decomposition of the individual frailty and estimation of the characteristics

of this decomposition using methods of survival analysis and quantitative genetics.

#### 12.4.4 Results: Genetics of Frailty and Longevity

Since the Danish survival data deal with individuals who survived to age 30, we need to adjust the bivariate survival equation (12.14) to consider conditional survival distribution. The results of the semiparametric estimation procedure for six models of frailty (*ACE*, *AE*, *ADE*, *DE*, *DCE*, and *HE*) are given by Yashin and Iachine (1995a) for male Danish twins in Table 12.1 and female Danish twins in [Table 12.2](#). Small ‘s’ after models in these tables indicates that the semiparametric procedure was used for parameter estimation.

##### Genetics of Frailty:

One can see from Table 12.1 that for male twins the ACE model converges to the AE model since the estimate of  $c^2$  tends to zero. (Standard errors for the ACE model are not shown in Table 12.1 since 0 is the boundary of the parametric space.) Note that the value of the likelihood function for the ADE model is higher than for the AE model. However, the AE model is better than the ADE model according to the likelihood ratio test. For females the value of ACE likelihood is higher than AE likelihood.

Model	$a^2$	$d^2$	$H^2$	k	$c^2$	$e^2$	LogLik	AIC
ACEs	0.479				0.000	0.521	-17876.57	2.00
(—)					(—)	(—)		
AEs	0.479					0.521	-17876.57	0.00
(0.111)						(0.111)		
ADEs	0.247	0.258				0.494	-17876.09	1.05
(0.333)	(0.355)					(0.292)		
DES		0.519				0.481	-17876.58	0.02
		(0.124)				(0.124)		
DCEs		0.423			0.082	0.494	-17876.09	1.05
		(0.169)			(0.115)	(0.136)		
HEs			0.506	0.372		0.494	-17876.09	1.05
			(0.122)	(0.147)		(0.122)		

TABLE 12.1: Genetic and environmental components of frailty for Danish male twins survived to age 30

However, the AE model is better according to the likelihood-ratio test. The ADE model converges to the AE model since the estimate of  $d^2$  tends to zero. (Standard errors are not shown for the same reason as before.) Since not all of our models are nested we used the Akaike Information Criteria (AIC) to compare different models. According to AIC the AE model gives the best fit to the bivariate data for both sexes. It is curious that in all six models

Model	$a^2$	$d^2$	$H^2$	k	$c^2$	$e^2$	LogLik	AIC
ACEs	0.405 (0.168)				0.111 (0.145)	0.484 (0.174)	-18878.66	1.47
AEs	0.525 (0.168)					0.475 (0.168)	-18878.92	0.00
ADEs	0.525 (—)	0.000 (—)				0.475 (—)	-18878.92	2.00
DES		0.509 (0.152)				0.491 (0.152)	-18881.88	5.92
DCEs		0.270 (0.139)			0.246 (0.152)	0.484 (0.173)	-18878.66	1.47
HES			0.516 (0.174)	0.607 (0.148)		0.484 (0.174)	-18878.66	1.47

TABLE 12.2: Genetic and environmental components of frailty for Danish female twins survived to age 30

the estimate of  $e^2$  (i.e., uncommon environmental component of variation in frailty) is close to 0.5 for both sexes. These results support the main finding of the previous study based on parametric estimation procedure: environmental factors determine not less than 50% of variability in individual frailty (Yashin and Iachine, 1994). Standard deviations of the parameter estimates given in Tables 12.1 and 12.2 do not take into account the errors in nonparametric estimations of the univariate survival functions, so they are not real standard errors. However, the estimates calculated in the parametric estimation procedure (Yashin and Iachine, 1994) are very close to the semiparametric estimates given in Tables 12.1 and 12.2 and have almost the same standard deviations.

### Genetics of Longevity:

The important feature of the suggested genetic model of bivariate survival is that it shows how genetic and environmental factors may influence longevity and survival. For example, using this model one can evaluate a lower bound for the biological limit of the human life span. To illustrate this idea let us consider the univariate survival model with the AE model of frailty, which gives the best fit to this data. The observed force of mortality in this model permits the following decomposition:

$$h(t) = P_1(t)h_0(t) + P_2(t)h_0(t)$$

where  $P_1(t) = a^2/(1 + \sigma^2 H(t))$  and  $P_2(t) = e^2/(1 + \sigma^2 H(t))$ . It seems natural to associate the first term of this decomposition with the genetic (senescent) and the second with the environmental (premature) components of mortality. Hence, the univariate survival function may be represented as

$$S(t) = S(t)^{a^2} S(t)^{e^2}.$$

This decomposition has an important property: it separates genetic and environmental influences on human survival, which can be further analyzed. For example, this representation suggests an idea of estimating the biological limit of human longevity using only the genetic component  $S_G(t) = S(t)^{a^2}$  of the survival function. Straightforward calculations show that for the estimated value of heritability is closed to 0.5 the values of male and female biological longevity are about 80 and 83 years, with standard deviations of 14 and 14.8 years, respectively. Thus, according to this model, male and female twins born between 1870 and 1900 have only 8 and 10 years potential, respectively, to increase longevity.

---

## 12.5 General Bivariate Frailty Model

Wienke et al. (2005) discussed bivariate gamma frailty and bivariate log-normal frailty models. Consider some bivariate observations, for example, the life spans of twins, or age at onset of a disease in spouses. Here Wienke et al. (2005) deal with single-event data (at most one event per individual). They assume that the frailties are acting multiplicatively on the baseline hazard function and that the observations in a pair are conditionally independent given the frailties. Hence, the hazard of individual  $j$  ( $j = 1, 2$ ) in pair  $i$  ( $i = 1, \dots, n$ ) has the form

$$h(t|Z_{ij}, Y_{ij}) = Z_{ij} h_0(t) e^{\beta Y_{ij}}, \quad (12.19)$$

where  $t$  denotes age,  $Y_{ij}$  a vector of observable covariates,  $\beta$  is a vector of unknown regression coefficients describing the effect of the covariates  $Y_{ij}$ ,  $h_0(t)$  is some baseline hazard function and  $Z_{ij}$  are unobserved (random) effects of frailties. Bivariate frailty models are characterized by the joint distribution of a two-dimensional vector of frailties  $(Z_{i1}, Z_{i2})$ . The form of the baseline hazard is important because all methods described below are parametrical. In principle, any parametric formula for a hazard rate is possible (e.g., Gompertz, Gompertz-Makeham, Weibull, exponential, piecewise constant). The methods reviewed in the following were developed mainly for some specific baseline hazard rate, e.g. exponential or piecewise constant. However, these methods are general and can be modified in order to incorporate any baseline hazard. A vast literature on human mortality suggests using the Gompertz hazard rate to describe mortality. Correlated frailty models with the Gompertz baseline hazard have been used quite frequently (see Yashin et al., 1995; Wienke et al., 2002; Yashin and Iachine, 1995; Wienke et al., 2001; and Iachine et al., 1998). For that reason and to save space, we investigate only bivariate frailty models that have the Gompertz baseline hazard rate:

$$h_0(t) = ae^{bt}. \quad (12.20)$$

Any method in this context is based on likelihood functions. In order to derive a marginal likelihood function, the facilitating assumption of the conditional independence of life spans given frailty is always used. Denote by  $\theta$  the vector of all frailty parameters of the model. Let  $\delta_{ij}$  be a censoring indicator for an individual  $j$  ( $j = 1, 2$ ) in pair  $i$  ( $i = 1, \dots, n$ ). Indicator  $\delta_{ij}$  is 1 if the individual has experienced the event of interest, and it is 0 otherwise. According to (12.19), the conditional survival function of the  $j$ -th individual in the  $i$ -th pair is

$$S(t|Z_{ij}, Y_{ij}) = e^{-Z_{ij}H_0(t)e^{\beta Y_{ij}}}, \quad (12.21)$$

where  $H_0(t)$  is the cumulative baseline hazard function. Given (12.20),

$$H_0(t) = \frac{a}{b}(e^{bt} - 1). \quad (12.22)$$

The contribution of the  $j$ -th individual in the  $i$ -th pair of the conditional likelihood is given by

$$L(t_{ij}, \delta_{ij}|Z_{ij}, Y_{ij}) = (Z_{oj}h_0(t_{ij})e^{\beta Y_{ij}})^{\delta_{ij}} e^{-Z_{ij}H_0(t_{ij})e^{\beta Y_{ij}}}, \quad (12.23)$$

where  $t_{ij}$  stands for age at death or the censoring time of the individual. Then, assuming the conditional independence of life spans given frailty and integrating out the random effects, we obtain the marginal likelihood function:

$$\begin{aligned} L(t, \delta|Y) &= \prod_{i=1}^n \int \int_{R^2} (z_{i1}h_0(t_{i1})e^{\beta Y_{i1}})^{\delta_{i1}} e^{-z_{i1}H_0(t_{i1})e^{\beta Y_{i1}}} \\ &\quad \times (z_{i2}h_0(t_{i2})e^{\beta Y_{i2}})^{\delta_{i2}} e^{-z_{i2}H_0(t_{i2})e^{\beta Y_{i2}}} \\ &\quad \times f_z(z_{i1}, z_{i2}, \theta) dz_{i1}, dz_{i2}, \end{aligned} \quad (12.24)$$

where

$$\begin{aligned} t &= (t_1, \dots, t_n), t_i = (t_{i1}, t_{i2}), \delta = (\delta_1, \dots, \delta_n), \\ \delta_i &= (\delta_{i1}, \delta_{i2}), Y = (Y_1, \dots, Y_n), \end{aligned}$$

$Y_i = (Y_{i1}, Y_{i2})$  and  $f_Z(\cdot, \cdot|\theta)$  is the pdf of the corresponding frailty distribution.

### Remark:

In survival analysis it is very common that data are left truncated at some time points  $t_{ij}^*$  with  $t^* = (t_1^*, \dots, t_n^*)$ ,  $t_i^* = (t_{i1}^*, t_{i2}^*)$ . In this case the likelihood function in (12.23) looks like

$$L(t_{ij}, \delta_{ij}|Z_{ij}, Y_{ij}) = (Z_{oj}h_0(t_{ij})e^{\beta Y_{ij}})^{\delta_{ij}} e^{-Z_{ij}H_0(t_{ij})e^{\beta Y_{ij}}} e^{Z_{ij}H_0(t_{ij}^*)e^{\beta Y_{ij}}}. \quad (12.25)$$

The additional term accounts for the fact that only individuals are included into the data set who did not die before the commencement of the study. Similar changes are necessary in formula (12.24). Wienke et al. (2005) restrict to the case of non-truncated data.

### 12.5.1 Gamma Model

The gamma distribution (we use notation  $\Gamma(k, l)$  for the two parameter distribution with shape parameter  $k$  and scale parameter  $l$ ) is one of the most popular frailty distributions. Frailty cannot be negative. The gamma distribution is, along with the lognormal distribution, one of the most commonly used distributions to model variables that are necessarily positive. Furthermore, it turns out that the assumption that frailty at birth is gamma-distributed yields some useful mathematical results including:

- Frailty among the survivors at any time  $t$  is gamma-distributed with the same value of the shape parameter  $k$  as at birth. The value of the second parameter, however,  $k$  is now given by  $l(t) = l + H_0(t)$ , where  $H_0(t)$  denotes the cumulative baseline hazard function.
- Frailty among those who die at any age  $t$  is also gamma-distributed, with the same parameter  $l(t)$  as among those surviving to age  $t$  but with shape parameter  $k + 1$ .
- The Laplace transform of a gamma-distributed random variable  $Z \sim \Gamma(k, l)$  is of a very simple form:  $L_Z(s) = Ee^{-Zs} = (1 + \frac{s}{l})^{-k}$ .

To make sure that the model can be identified, it makes sense to use the parameter restriction  $EZ = 1$ , which results in  $k = l$  for the gamma distribution. Denoting the variance of the frailty variable by  $\sigma^2 = \frac{1}{l}$ , the univariate survival function is represented by

$$S(t) = \mathbf{L}(H_0(t)) = (1 + \sigma^2 H_0(t))^{-\frac{1}{\sigma^2}},$$

where  $H_0(t)$  denotes the cumulative baseline hazard function.

The correlated gamma-frailty model is developed by Yashin and Iachine (1994), Pickles and Crouchley (1994), and Petersen (1998) for the analysis of multivariate failure time data, in which two associated random variables are used to characterize the frailty effect for each cluster. For example, one random variable is assigned to twin 1 and another to twin 2 so that they are no longer constrained to having a common frailty as in the shared frailty model. To be more specific, let  $k_0, k_1$  be some real positive variables. Set  $l = k_0 + k_1$  and let  $Y_0, Y_1, Y_2$  be independent gamma-distributed random variables with  $Y_0 \sim \Gamma(k_0, l), Y_1 \sim \Gamma(k_1, l), Y_2 \sim \Gamma(k_1, l)$ . Consequently,

$$Z_1 = Y_0 + Y_1 \sim \Gamma(k_0 + k_1, l) \sim \Gamma(l, l), \quad (12.26)$$

$$Z_2 = Y_0 + Y_2 \sim \Gamma(k_0 + k_1, l) \sim \Gamma(l, l)$$

are the frailties of individual 1 and 2 in a pair. The bivariate survival function of this model is given by

$$S(t_1, t_2) = \begin{cases} S(t_1)^{1-\rho} S(t_2)^{1-\rho} [S(t_1)^{-\sigma^2} + S(t_2)^{-\sigma^2} - 1]^{\frac{\rho}{\sigma^2}}, & \text{if } \sigma^2, \rho > 0, \\ S(t_1)S(t_2), & \text{if } \sigma^2 = 0 \text{ or } \rho = 0, \end{cases} \quad (12.27)$$

where  $S(t)$  denotes the marginal univariate survival function, assumed to be equal for both partners in a twin pair and  $0 \leq \rho \leq 1$  holds. Furthermore, it holds that  $\rho = \text{corr}(Z_1, Z_2)$  and  $\sigma^2 = \mathbf{V}(Z_i)$ , ( $i = 1, 2$ ). For simplicity, we drop the dependence of the survival functions from observed covariates. Obviously, the shared gamma-frailty model by Clayton (1978) is a special case of (12.27) when  $\rho = 1$ . We will refer to model (12.27) as Model 1.

### 12.5.2 Lognormal Model

The lognormal model is much more flexible than the gamma model, because it is not based on the additive composition of the two frailties as used in (12.26). On the other hand, the lognormal distribution does not allow an explicit representation of the likelihood function, which requires more sophisticated estimation strategies. We assume that the two frailties of individuals in a pair are given by

$$\begin{pmatrix} Z_{i1} \\ Z_{i2} \end{pmatrix} \sim \text{Log } N \left( \begin{pmatrix} m \\ m \end{pmatrix}, \begin{pmatrix} s^2 & rs^2 \\ rs^2 & s^2 \end{pmatrix} \right), \quad (12.28)$$

where  $\text{Log } N$  denotes the (bivariate) lognormal distribution. Here  $m, s^2$  and  $r$  denote the mean, variance and correlation of the respective normal distribution. Mean, variance and correlation of the frailties are related to these parameters as follows:

$$\mu = \mathbf{E} Z_{ij} = e^{m + \frac{s^2}{2}}, \quad (12.29)$$

$$\sigma^2 = \mathbf{V}(z_{ij}) = e^{2m+s^2}(e^{s^2}-1), \quad (12.30)$$

$$\rho = \text{corr}(Z_{i1}, Z_{i2}) = \frac{e^{rs^2}-1}{e^{s^2}-1}. \quad (12.31)$$

Two different types of lognormal frailty models arise from two restrictions on the parameters of frailty distribution. First, one can use the restriction  $m = 0$ . This means that the logarithm of frailty has a mean of zero. In this case, a ‘standard’ individual has the logarithm of a hazard rate which is equal to  $\ln h_0(t)$ . Any individual in a population has the logarithm of the hazard rate distorted by some random variables  $W_{ij} = \ln Z_{ij}$ . This value is added to the ‘true’ logarithm of hazard rate  $\ln h_0(t)$  to provide the logarithm of the hazard rate of the individual. In this interpretation, it is natural to assume that the distortions  $W_{ij}$  have a normal distribution with a mean of zero. Such a model is called Model 2.

Second, following the usual definition of frailty by Clayton (1978) and Vaupel et al. (1979), one can use  $\mu = 1$  and in this case

$$m = \mathbf{E} \ln Z_{ij} = -\frac{1}{2}s^2, \quad (12.32)$$

$$s^2 = \mathbf{V}(\ln Z_{ij}) = \ln(1 + \sigma^2). \quad (12.33)$$

In this model a ‘standard’ individual has the hazard rate  $h_0(t)$ . Individual  $j$  in the  $i$ -th pair has the hazard rate of a ‘standard’ individual multiplied by the frailty  $Z_{ij}$ . The above restriction on  $\mu$  means that the average frailty in a population equals 1 (at the beginning of the followup). We refer to this model as Model 3.

### 12.5.3 Estimation Strategies

Parameter estimation in the gamma model is straightforward. The frailty term can be integrated out and an explicit representation of the unconditional bivariate survival function exists (12.27), which can be used to derive the likelihood function.

Several estimation methods for bivariate lognormal frailty models in consequence have been suggested to be used within a non-Bayesian framework. Various modifications of the maximum likelihood procedure are applicable to the bivariate frailty models. Ripatti and Palmgren (2000) derived an estimating algorithm based on the penalized partial likelihood (PPL). Xue and Brookmeyer (1996) suggested a modified EM algorithm for the bivariate lognormal frailty models. Sastry (1997) developed the modified EM algorithm for the multiplicative two-level gamma frailty model. The same method can be applied to bivariate lognormal frailty models. Ripatti and Palmgren (2002) present yet another method to deal with EM-like algorithms in a bivariate lognormal frailty model.

Wienke et al. (2005) use numerical integration procedures. Integrals over the univariate and multivariate normal distributions can be approximated in different ways. One possibility is to use Gauss-Hermite quadratures by Naylor and Smith (1982) and Smith et al. (1987). Similar ideas are employed in various applications of random effect models in event history analysis by Lillard (1993), Lillard et al. (1995), and Panis and Lillard (1995). The methods are implemented in the aML software package (aML version 1, see Lillard and Panis, 2000). Both methods were used to estimate parameters of the bivariate lognormal frailty models for simulated data.

Several studies on the application of Bayesian methods to multivariate frailty models exist. An example of a Bayesian approach to the gamma frailty model is found in Bolstad and Manda (2001). Gibbs’ sampling scheme for the bivariate lognormal frailty model with an exponential baseline hazard is given in Xue and Ding (1999). Korsgaard et al. (1998) present a Bayesian inference in the lognormal frailty model with a semiparametric hazard.

## 12.6 Correlated Compound Poisson Frailty for the Bivariate Survival Lifetimes

Moger and Aalen (2005) developed correlated compound Poisson frailty model for the survival time of the two individuals in a family. Let  $Z_1$  and  $Z_2$  be the frailty variables of two individuals in a family with joint distribution  $f_{Y_1, Y_2}(y_1, y_2)$ . Let their marginal distribution given  $\rho$ ,  $f_{Y_1}(y_1|\rho)$  and  $f_{Y_2}(y_2|\rho)$ , be independent identically distributed compound Poisson with parameters  $\eta$  and  $\nu$ . The parameter  $\rho$ , which is common for both  $Z_1$  and  $Z_2$ , is assumed to be PVF distributed with parameters  $\alpha$ ,  $\epsilon$  and  $\theta$ . The joint discrete part of  $(Z_1, Z_2)$  is

$$P(Z_1 = 0, Z_2 = 0) = E_\rho[\exp(-2\rho)] = L_\rho(2).$$

The joint density of the continuous part of the distribution can be found analogously to (5.35), by using the  $\rho$ , in the distributions  $f_{Z_1|\rho}(z_1|\rho)$  and  $f_{Z_2|\rho}(z_2|\rho)$  to get derivatives of  $L_\rho(s)$ . It is given by

$$\begin{aligned} f_{Z_1, Z_2}(z_1, z_2; \eta, \nu, \alpha, \theta, \epsilon) &= \frac{1}{z_1 z_2} \exp[-\nu(z_1 + z_2)] \sum_{n=2}^{\infty} (-1)^n L_\rho^{(n)}(2) \\ &\quad \times \sum_{k=1}^{n-1} \frac{n u^{n\eta} z_1^{(n-k)\eta} z_2^{k\eta}}{\Gamma((n-k)\eta) \Gamma(k\eta) (n-k)! k!}, \end{aligned}$$

where  $L_\rho^{(n)}(s)$  is defined in (5.36). Since the marginal distributions are compound Poisson given  $\rho$ , the joint distribution has an interesting feature: it is possible to have two related individuals where one has zero frailty and the other has a positive frailty. The probability is given by

$$P(Z_1 = 0, Z_2 > 0) = E_\rho[\exp(-\rho) - \exp(-2\rho)] = L_\rho(1) - L_\rho(2).$$

In some situations, this is an aspect that may make the model fit better than a shared frailty model. Also, it is interesting for the interpretation. For instance, testicular cancer is hypothesized to be caused by some sort of damage in foetal life (Henderson et al., 1988). This damage could be due to genetics, mothers or pregnancies. If there is a mother effect, it may not be natural with the possibility of  $Z_1 = 0$  and  $Z_2 > 0$ .

By using (5.37), one easily finds their joint Laplace transform

$$L_{Z_1, Z_2}(s, t) = L_\rho \left( 2 - \left( \frac{\nu}{\nu + s} \right)^\eta - \left( \frac{\nu}{\nu + t} \right)^\eta \right),$$

which in the case of PVF distributed  $\rho$  is

$$L_{Z_1, Z_2}(s, t) = \begin{cases} \exp\left(-\frac{\epsilon}{\alpha}\left\{\left[\theta + 2 - \left(\frac{\nu}{\nu+s}\right)^\eta - \left(\frac{\nu}{\nu+t}\right)^\eta\right]^\alpha - \theta^\alpha\right\}\right) \\ \quad \text{if } \alpha \leq 1, \alpha \neq 0, \\ \left[\frac{\theta}{\theta+2-\left(\frac{\nu}{\nu+s}\right)^\eta-\left(\frac{\nu}{\nu+t}\right)^\eta}\right]^\epsilon \quad \text{if } \alpha = 0. \end{cases} \quad (12.34)$$

Note that the univariate Laplace transform in (5.38) appears by setting  $t = 0$ .

By noting that  $Cov(Z_1, Z_2) = COV(E(Z_1|\rho), E(Z_2|\rho))$  and using (5.24), the correlation coefficient between frailties of two individuals in a family obtained by Moger and Aalen (2005) is

$$\text{Corr}(Z_1, Z_2) = \frac{\eta(1-\alpha)}{\theta + \eta(1-\alpha+\theta)} \quad \text{if } \theta > 0. \quad (12.35)$$

The parameter  $\theta$  determines the degree of correlation. Since none of the moments exist (when  $\theta = 0$ ), the correlation coefficient cannot be used as a measure of dependence for the compound Poisson-positive stable distribution. For values of  $\theta$  close to zero, the correlation between two related individuals is approaching one. It is evident that the correlation has to be larger than zero, so the model can not handle negative dependencies.

Let  $T_1$  and  $T_2$  be the lifetimes of the two individuals which are independent. The survival function of  $(T_1, T_2)$  given the two dependent frailties  $(Z_1, Z_2)$  is given by

$$S_{T_1, T_2|Z_1, Z_2}(t_1, t_2|z_1, z_2) = \exp(-H_1(t_1)z_1 - H_2(t_2)z_2). \quad (12.36)$$

The unconditional survival function of  $(T_1, T_2)$  is obtained by integrating  $(Z_1, Z_2)$  out

$$S_{T_1, T_2}(t_1, t_2) = E(\exp(-H_1(t_1)Z_1 - H_2(t_2)Z_2))$$

$$= \begin{cases} \exp\left(-\frac{\epsilon}{\alpha}\left\{\left[\theta + 2 - \left(\frac{\nu}{\nu+M_1(t_1)}\right)^\eta - \left(\frac{\nu}{\nu+M_2(t_2)}\right)^\eta\right]^\alpha - \theta^\alpha\right\}\right) \\ \quad \text{if } \alpha \leq 1, \alpha \neq 0, \\ \left[\frac{\theta}{\theta+2-\left(\frac{\nu}{\nu+M_1(t_1)}\right)^\eta-\left(\frac{\nu}{\nu+M_2(t_2)}\right)^\eta}\right]^\epsilon \quad \text{if } \alpha = 0. \end{cases} \quad (12.37)$$

Let  $(T_1, T_2)$  are independent Weibull distributions with  $W(\lambda_1, c_1)$  and  $W(\lambda_2, c_2)$  respectively, where  $\lambda_i$ 's scale parameters and  $c_i$ 's are shape parameters of Weibull distributions. The survival function of  $T_i$  is

$$S_{T_i}(t_i) = \exp(-\lambda_i t_i^{c_i}). \quad (12.38)$$

Now the unconditional survival function of  $(T_1, T_2)$  with correlated compound Poisson frailties is given by

$$S_{T_1, T_2}(t_1, t_2) = \begin{cases} \exp\left(-\frac{\epsilon}{\alpha}\left\{\left[\theta + 2 - \left(\frac{\nu}{\nu+\lambda_1 t_1^{\epsilon_1}}\right)^{\eta} - \left(\frac{\nu}{\nu+\lambda_2 t_2^{\epsilon_2}}\right)^{\eta}\right]^{\alpha} - \theta^{\alpha}\right\}\right) \\ \quad \text{if } \alpha \leq 1, \alpha \neq 0, \\ \left[\frac{\theta}{\theta + 2 - \left(\frac{\nu}{\nu+\lambda_1 t_1^{\epsilon_1}}\right)^{\eta} - \left(\frac{\nu}{\nu+\lambda_2 t_2^{\epsilon_2}}\right)^{\eta}}\right]^{\epsilon} \quad \text{if } \alpha = 0. \end{cases} \quad (12.39)$$

In order to solve the identifiability problem, we assume a mean of 1 for the frailty distributions. For the gamma distribution, this can be achieved by setting  $\theta = \epsilon$ . In the shared PVF model,  $E(Z) = 1$  is achieved by setting  $\epsilon = \theta^{1-\alpha}$ . The shared frailty models are compared to a compound Poisson model where  $\rho$  is gamma distributed, yielding a compound Poisson-gamma model. To secure a unit mean for the frailty, we get  $\epsilon = \nu\theta/\eta$ .

Hanagal (2010a) assumed the distribution of frailty as compound Poisson-gamma (with  $\alpha = 0$ ) distribution for the bivariate survival data. The bivariate survival function based on this frailty is given by

$$S(t_1, t_2) = \left[ \frac{\theta}{\theta + 2 - \left(\frac{\nu}{\nu+\lambda_1 t_1^{\epsilon_1}}\right)^{\eta} - \left(\frac{\nu}{\nu+\lambda_2 t_2^{\epsilon_2}}\right)^{\eta}} \right]^{\nu\theta/\eta}. \quad (12.40)$$

Hanagal (2010a) obtained the estimation of parameters in correlated compound Poisson frailty with Weibull distributions as baseline models based on the censored samples. He did a simulation study based 100 samples of sizes from bivariate Weibull baseline and obtained MLEs of the parameters.

## 12.7 Applications

Commenges and Gadda (1997) presented a family of tests based on correlated random effects models which provides a synthesis and generalization of tests for homogeneity. In these models, each subject has a particular random effect, but the random effects between subjects are correlated. They derive the general form of the score statistic for testing that the random effects have a variance equal to zero. They applied this result to both parametric and semi-parametric models. In both cases, they showed that under certain conditions the score statistic has an asymptotic normal distribution. They considered several applications of this theory, including overdispersion, heterogeneity between groups, spatial correlations and genetic linkage.

Zahl (1997) analyzed two data sets, one is based on malignant melanoma patients with cancer and another is based on cancer using correlated gamma frailty model. He used bivariate frailty variables  $Z_1 = X_1 + X_2$  and  $Z_2 = \alpha(X_1 + X_3)$ , where  $X_1, X_2$  and  $X_3$  are three independent gamma variables and  $\alpha$  is scaling parameter.

Viswanathan and Manatunga (2001) compared two frailty distributions, gamma and positive stable, based on the diagnostic plots. These plots are capable of differentiating between the two frailty models when strong association is present between two frailty models. They analyzed diabetic retinopathy data and a reasonable fit to the gamma frailty model is found.

Giard et al. (2002) suggested that multivariate frailty model can be used in the genetic analysis of the ageing process as a whole, simplified to consisting of the states healthy, disabled, and deceased. They evaluated simultaneously the relative magnitude of genetic and environmental influences on frailty variables corresponding to the period of good health and to the life span. The frailty variables can be interpreted as susceptibility to illness or death. The model can be applied to data on groups of related individuals (twins, siblings, a litter). One of the major advantages of this model is that it allows one to include groups of individuals where some or all members of the group are already deceased at the time of observation. They discussed the estimation procedures and analyzed twin data on prostate cancer.

Wienke et al. (2003) suggested a cure-mixture model to analyze bivariate time-to-event data using correlated gamma frailty model. They obtained the model for left-truncated and right censored data, and accounts for heterogeneity, as well as for an insusceptible (cure) fraction in the study population. They obtained estimation procedures and applied it to breast cancer incidence data for 5857 Swedish female monozygotic and dizygotic twin pairs from the cohort of the Swedish Twin Registry. They estimated the size of the susceptible fraction and the correlation between the frailties of the twin partners.

Finkelstein and Esaulova (2006) considered correlated frailty model under the assumption of conditional independence of components. They also discussed two examples based on their results.

Wienke et al. (2006) used three correlated frailty models to analyze bivariate survival data by assuming gamma, lognormal and compound Poisson distributed frailty. All approaches allow to deal with right censored lifetime data and account for heterogeneity as well as for a non-susceptible (cure) fraction in the study population. In the gamma and compound Poisson model traditional ML estimation methods are used, whereas in the lognormal model MCMC methods are applied. Breast cancer incidence data of Swedish twin pairs illustrate the practical relevance of the models, which are used to estimate the size of the susceptible fraction and the correlation between the frailties of the twin partners.

Garibotti et al. (2006) applied Cox proportional hazards models to data from three-generation pedigrees in the Utah Population Database using two different frailty specification schemes that account for common environments

(shared frailty) and genetic effects (correlated frailty). In a model that includes measures of familial history of longevity and both frailty effects, they find that the variance component due to genetic factors is comparable to the one attributable to shared environments.

Congdon (2008) considered a model for survival data with a permanent survival fraction and non-monotonic failure rates and evaluate the gain in model fit, and effects on inference, from adding frailty. An application considers age at first maternity using data from the 2002 German General Social Survey, with permanent survival amounting to childlessness. Regressions are used to explain both the failure time of the event (here age at first maternity) and the permanent survival mechanism (susceptibility to undergo maternity or not). Additive correlated effects are included in the linked models defining these regressions and relate to two types of frailty: influences on the event rate itself and influences on the probability of susceptibility. A hierarchical Bayesian approach is adopted with likelihood conditional on bivariate random frailty effects, and a second stage prior defining the density of those effects. A Bayesian approach facilitates modeling with multivariate random effects whereas frequentist approaches based on marginal likelihoods with random effects integrated out using numerical methods may become infeasible or unreliable when there are many random parameters (Tutz and Kauermann, 2003; Kim et al., 2002). Monte Carlo Markov Chain (MCMC) methods are used for estimation via the WINBUGS package (Lunn et al., 2000), and generate samples from the posterior distribution without the form of the posterior density being known analytically (Gilks et al., 1996). This is useful in summarizing possibly non-normal densities relating to model functionals (e.g. modal ages at maternity) and in obtaining posterior probabilities relating to hypotheses on such functionals, e.g. that the modal age for women with low education years is lower than the modal age for women with extended education.

# Chapter 13

---

## Additive Frailty Models

13.1	Introduction	243
13.2	Modeling Multivariate Survival Data Using the Frailty Model	245
13.3	Correlated Frailty Model	246
13.4	Relations to Other Frailty Models	247
13.4.1	Shared Frailty Model	248
13.4.2	Over-Dispersion Model	248
13.4.3	Twin Model	248
13.4.4	Litter Model	249
13.4.5	Genetic Model	249
13.4.6	Adoption Model	250
13.4.7	Competing Risks	250
13.4.8	Example	251
13.5	Additive Genetic Gamma Frailty	255
13.5.1	Example 13.1	255
13.5.2	Example 13.2	256
13.5.3	Example 13.3	257
13.5.4	Application in Danish Adoptive Register Data	259
13.6	Additive Genetic Gamma Frailty for Linkage Analysis of Diseases	263
13.6.1	Genetic Frailties Defined by Multiple Unlinked Disease Loci	265
13.6.2	Expected Genetic Frailties over the Inheritance Vectors	266
13.6.3	Model for Age of Onset Data in a Family	268
13.6.4	Conditional Hazards Ratio for Sib Pairs	268
13.6.5	Estimation Methods and Test of Linkage	270
13.6.6	Breast Cancer Data Example	270

---

### 13.1 Introduction

A common assumption made in the analysis of effects of treatments or risk factors on survival is that individuals are conditionally independent given the observed covariates. Often, this assumption is perfectly plausible because of the way the study was designed. When dealing with repeated or multiple events within the same individual, the independence assumption can be questionable, and this should be taken into account in the estimation of treatment effects. In many basic genetic epidemiological studies, the aim is to assess genetic and environmental influences on some particular disease outcome, and addressing the dependence becomes the main objective of the model. Animal studies have as the natural unit the litter or perhaps whole pedigrees, whereas for humans, it is the family. In family, husband and wife share different environmental exposures, including certain lifestyle characteristics, perhaps diet

and exercise habits, and more exogenous environments such as air pollution and radiation level. Sharing a number of environmental exposures, it seems reasonable to suppose that husband and wife have related event times. With children in the family, not only some environment but also genes are shared, thus making it likely that the event times of the children are related, perhaps more so than the event times of their parents because siblings share more genes than their parents. This chapter focuses on the last situation with families involving individuals with varying degrees of association between their respective event times - an association that is partly accounted for by their familial relationship, i.e., mother-father, daughter-son, uncle-son, and partly by the environments that they as individuals share.

One important example that will serve to motivate and illustrate the statistical model proposed by Petersen (1998) is the study of genetic and environmental influences on premature death in adult adoptees, presented in Sorensen et al. (1988), born during the period 1924 through 1926 who were placed early in life with adoptive parents unrelated to them. The idea of the adoption design is that the adoptee's resemblance to the biological relatives is due to inheritance, whereas the resemblance to the adoptive relatives is attributed to the shared environment. The study thus allows for the assessment of both genetic and environmental components in the development of some disease. One of the conclusions reached from the study was that premature death in adults from infections such as pneumonia, tuberculosis, and bronchitis has a strong genetic background but that environmental exposures did not have a (significant) influence.

In this chapter, a class of multivariate frailty models is suggested where, conditionally on the multivariate frailty, the observations are independent. The models correspond to well-known variance components models for normally distributed data in the sense that association between event times is induced by letting individuals share some, but not all, of the variance components. However, instead of describing the life times (possibly after taking the logarithm) directly in a variance components model, frailty components are combined additively, which then act multiplicatively on the individual hazard rates. This both extends, in a rather straight-forward way, ordinary survival analysis with its emphasis on hazard modeling and incorporates well-known variance components models to account for the dependence between events of related individuals. The models allow for covariates, which makes it possible to examine whether the dependence can be explained by specific covariates.

### 13.2 Modeling Multivariate Survival Data Using the Frailty Model

In the usual frailty model theory, a single unobserved random component is introduced into the intensity function. This is done to model two different but related sources of variation in event time data. One source of variation is from unobserved individual covariates that are not included in the study either because of practical circumstances or because they are not known to be risk factors. This is the original use of the term frailty (see Vaupel, Manton, and Stallard, 1979), and such heterogeneity and selection effects are discussed in more detail by Vaupel and Yashin (1985) and Aalen (1994).

The other source of variation stems from unobserved common covariates, and when they are integrated out, dependence is generated between events (see, e.g., Clayton, 1978; Hougaard, 1986; Nielsen et al., 1992).

A general approach to multivariate event time data must account for variation due to both unobserved shared traits and unobserved nonshared (individual) traits.

In a model, referred to in the following as a shared frailty model, with a single unobserved random component with a finite mean frailty distribution and possibly covariates, the frailty parameter measures something besides dependence. Using a result by Elbers and Ridder (1982), this point has been stressed by Hougaard (1986), who suggests the positive stable distribution to remedy this deficiency in the shared frailty model.

A shared frailty model has another drawback, which is especially compelling in the twin example of Yashin, Vaupel, and Iachine (1995) and is illustrated by the data in Vaupel et al. (1992). They find in two separate analyses of Danish monozygotic (MZ) and dizygotic (DZ) twins a higher dependence parameter (the variance in a gamma distribution) and a steeper baseline hazard (conditional on the frailty) for monozygotic than for dizygotic twins. To understand how this can come about and why it is a problem, note that, with respect to their life time distribution, twins are, as individuals, like singletons from the general population. We would therefore like models for twin data to allow for both the same baseline hazard function for MZ and DZ twins and the same marginal survival function (univariate, not bivariate). Failure to meet these requirements will invalidate a generalization from twins to a general population. Because the marginal survivor function is found from the conditional by integration with respect to the frailty distribution, these two requirements imply that the dependence parameters for MZ and DZ twins are identical. But there are literally hundreds of studies indicating that MZ twins with respect to many different endpoints are more closely associated than DZ twins. This is a drawback of the frailty model in the context of modeling correlated event times and stems from the fact that individuals sharing the same frailty (and co-variates) have exactly the same risk. As a way of circumvent-

ing this problem, Yashin et al. (1995) suggests decomposing the frailty of each twin in a pair into a sum of two independent frailties, one of which is shared by both twins.

---

### 13.3 Correlated Frailty Model

This section defines a frailty model for event time data that corresponds to a variance components model encountered in the analysis of Gaussian data.

In the following, indices  $i, j$  and  $g$  will always refer to families, individuals, and strata, respectively. To avoid too many indices, we think of each individual as being uniquely identified by  $i$  and  $j$  such that  $g = g(i, j)$ . Furthermore, because the family is thought of as the independence unit in the sense that individuals from different families have independent event times, the definition of the model is carried out for only one family, implying that the index  $i$  can be omitted.

Let  $N = (N_j; j = 1, \dots, m)$  be a multivariate counting process that keeps track of correlated event times of  $m$  individuals. The aim is to specify intensity processes  $h = (h_j; j = 1, \dots, m)$  for the event times, where

$$h_j(t) = \mathbf{A}_j^T Z X_j(t) \alpha_g^\theta(t; \mathbf{Y}_j), \quad j = 1, \dots, m, \quad (13.1)$$

and  $X_j(t)$  is a nonnegative predictable process, i.e., known at time  $t-$ , usually just indicating whether the subject indexed by  $j$  is at risk for an event just before  $t$ . As before,  $Z$  is a  $k$  dimensional vector of unknown individual effects and  $\mathbf{A}_j = (a_{1j}, \dots, a_{kj})^T$  is a known  $k$ -dimensional design vector. More specifically, it is assumed that the design matrix  $\mathbf{A} = (a_{ij}; i = 1, \dots, k; j = 1, \dots, m)$  has the form

$$a_{ij} = \begin{cases} 1, & \text{individual } j \text{ has component } i \\ 0, & \text{individual } j \text{ does not have component } i. \end{cases}$$

The hazard function  $\alpha_g^\theta(t; Y_j)$  is written in a very general form, allowing for individual - specific covariates  $Y_j$ , regression and baseline hazard parameters  $\theta$ , and stratum-specific baseline hazards defined by the function  $g = g(j)$ . When, in the following, covariates are relevant to the exposition, one may think of them as acting on the intensity in a Cox-model manner. Other choices of relative risk functions, perhaps more relevant in a particular application, can similarly be handled but are not discussed further here. In the following, the hazard function will be denoted  $\alpha_j(t)$ . The random components  $Z_1, \dots, Z_k$  are assumed independent and gamma distributed with parameters  $(\nu_1, \eta), \dots, (\nu_k, \eta)$ , respectively. The common scale parameter restricts the mean frailty of individuals (with all  $a$  coefficients equal to one) to one and is imposed to make the hazard function identifiable. In applications, a

reparametrization in terms of the frailty variances  $\sigma_1^2, \dots, \sigma_k^2$  of  $Z_1, \dots, Z_k$ , respectively, is carried out. Between the two sets of parameters, there is a one-to-one correspondence given by

$$\nu_i = \frac{\sigma_i^2}{(\sigma_1^2 + \dots + \sigma_k^2)^2}, \quad i = 1, \dots, k. \quad (13.2)$$

The common scale parameter is given by the inverse total variance, i.e.,  $\eta = (\sigma_1^2 + \dots + \sigma_k^2)^{-1}$ .

Analogous to what is done with variance components for normally distributed variables, equation (13.1) is interpreted as the conditional intensity process of the event times given the frailties. Intensities, and also life times, of different individuals are correlated through the frailties but, given the frailties, the life times for different individuals are independent. Note that there is no requirement that the  $j$  or  $h$  indices are balanced over families. Families may be of different sizes and some strata may be represented in some families and not in others.

The correlated frailty model is a model in its own right because it clearly addresses, and does so in a rather straightforward way, the difficult question of how to model correlated event times data. Although the model is not based on a theory of how the genes and the environments act together to determine the lifespan of an individual, we can motivate a model of this type in two different ways.

An interpretation of the shared frailty model in terms of competing risks was given in Hougaard (1984), and this can similarly be given in the multivariate case where each individual, conceptually, has frailty components corresponding to different causes of death. In a simple example, individuals could share genes with some environment with others. Then we can think of individuals dying either because of the environments that they have been exposed to or because of their genetic makeup.

Another way of motivating the use of the proposed additive structure is that it arises as the first order approximation of a more complicated dependence structure. Motivating the additive structure in this way is likely to be relevant when interest is in the baseline hazard or in regression coefficients because information about the original frailty components cannot immediately be retrieved.

## 13.4 Relations to Other Frailty Models

Before going into more technical matters, we sketch some applications and show the connection to some previous studies in frailty models. Because the structure, inducing dependence between individuals within a family, is in the

frailties, we introduce  $Z^{(j)} = \mathbf{A}_j^T \mathbf{Z}$ , which is the  $j$ -th individual's frailty. So,  $(Z^{(1)}, \dots, Z^{(m)})^T = \mathbf{A}\mathbf{Z}$ , where  $\mathbf{A}$  is the design matrix for the entire family.

### 13.4.1 Shared Frailty Model

The shared frailty model is

$$h_{ij}(t) = Z_{i0} X_{ij}(t) \alpha_g(t), \quad i = 1, \dots, n; \quad j = 1, \dots, m_i; \quad g = 1, \dots, k. \quad (13.3)$$

With  $k = 4$ , corresponding to, e.g., father, mother, male, and female offspring, this approach would allow for an assessment of interfamily variability. The model (13.3) is a reformulation of the model used by Clayton (1978) to estimate familial tendency in heart attack incidence based on fathers and sons ( $k = 2$ ). Details concerning this model can be found in Andersen et al. (1993). Formulated in our notation, the model is, for each family, defined by

$$\mathbf{A} = \mathbf{1}_m,$$

where  $\mathbf{1}_m$  is an  $m \times 1$  vector of ones.

### 13.4.2 Over-Dispersion Model

The over-dispersion model is

$$h_{ij}(t) = Z_{ij} X_{ij}(t) \alpha(t), \quad i = 1, \dots, n; \quad j = 1, \dots, m_i. \quad (13.4)$$

Model (13.4) is a formulation of the original frailty model, used by Vaupel et al. (1979) to model heterogeneity due to unobserved covariates. Each subject has its own associated frailty component acting multiplicatively on the baseline hazard. Frailties (and life times) associated with different subjects are independent. Note that the notion of families in this set-up has no meaning because they are all size one. The indices  $ij$  are kept only to conform with previous notation. Here  $\mathbf{A}$  contains only a single element, which is one, i.e.,  $\mathbf{A} = (1)$ .

### 13.4.3 Twin Model

The twin model is

$$\begin{aligned} h_{i1}(t) &= (Z_{i0} + Z_{i1}) X_{i1}(t) \alpha_g(t) \\ h_{i2}(t) &= (Z_{i0} + Z_{i2}) X_{i2}(t) \alpha_g(t), \quad i = 1, \dots, n. \end{aligned} \quad (13.5)$$

This model, suggested by Pickles et al. (1994) and Yashin et al. (1995), applies to classical twin studies involving identical and nonidentical twins. The frailties represent shared genes and environments ( $Z_{i0}$ ) and nonshared genes

and environments ( $Z_{i1}, Z_{i2}$ ). The model allows for different baseline hazards for male and female twins. For this model,

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}.$$

### 13.4.4 Litter Model

The litter model is expressed as

$$\begin{aligned} h_{i1}(t) &= (Z_{i0} + Z_{i1})X_{i1}(t)\alpha(t) \\ &\vdots && \vdots && i = 1, \dots, n; j = 1, \dots, m_i \\ h_{im_i}(t) &= (Z_{i0} + Z_{im_i})X_{im_i}(t)\alpha(t). \end{aligned} \quad (13.6)$$

This represents an extension of the shared frailty model (13.3) along the lines of the twin model. As opposed to the shared frailty model, it allows for heterogeneity among subjects within litters. Apart from being a seemingly more reasonable model in a litter study context, it also has more operational value since it may be used to test the goodness-of-fit of a shared frailty model. The litter model, described in detail in Petersen, Andersen, and Gill (1996), has a design matrix of the form

$$\mathbf{A} = (\mathbf{1}_m, \mathbf{I}_m),$$

where  $\mathbf{I}_m$  is an identity matrix of dimension  $m$ .

### 13.4.5 Genetic Model

The genetic model is

$$\begin{aligned} h_{i1}(t) &= (Z_{i1} + Z_{i2})X_{i1}(t)\alpha_1(t) \\ h_{i2}(t) &= (Z_{i3} + Z_{i4})X_{i2}(t)\alpha_2(t) \\ h_{i3}(t) &= (Z_{i1} + Z_{i3})X_{i3}(t)\alpha_3(t), \quad i = 1, \dots, n. \end{aligned} \quad (13.7)$$

This model, studied in Korsgaard and Andersen (1998), involves families of three: mother ( $h_{i1}$ ), father ( $h_{i2}$ ), and child ( $h_{i3}$ ). Focus is on genetic effects only and, using standard results from quantitative genetics, it is hypothesized that correlation between frailties of different individuals are given by their family relationship. Thus, because mother and child (on the average) share half their genes, the correlation between the frailty of the mother ( $Z^{(M)} = Z_{i1} + Z_{i2}$ ) and the child ( $Z^{(C)} = Z_{i1} + Z_{i3}$ ) is one half. Imposing this kind of structure reduces the model to one with a single parameter. For this model,

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}.$$

### 13.4.6 Adoption Model

The adoption model is

$$\begin{aligned} h_{i1}(t) &= (Z_{i1} + Z_{i2})X_{i1}(t)\alpha_1(t) \\ h_{i2}(t) &= (Z_{i1} + Z_{i3} + Z_{i4})X_{i2}(t)\alpha_2(t) \\ h_{i3}(t) &= (Z_{i3} + Z_{i5})X_{i3}(t)\alpha_3(t), \quad i = 1, \dots, n. \end{aligned} \quad (13.8)$$

This is a model particularly suited for adoption data, where an adoptee ( $h_{i2}$ ) is separated from its biological parents, here just mother ( $h_{i1}$ ), at some, usually early, point in life and placed with an adoptive family, here adoptive mother ( $h_{i3}$ ). The adoptee and the adoptive mother share no genes or rather no more genes than two randomly chosen individuals from a general population. This makes it possible to assess both a genetic and an environmental component in the development of some disease. The motivation for this particular model comes from the conceptual division of frailties into genetic and environmental components, and a natural additive structure accounting for an association between life times due to genetic and environmental effects would be

$$\begin{aligned} Z^{(A)} &= G_s^A + G_{ns}^A + E_s^A + E_{ns}^A \\ Z^{(C)} &= G_s^B + G_{ns}^C + E_s^A + E_{ns}^C \\ Z^{(B)} &= G_s^B + G_{ns}^B + E_s^B + E_{ns}^B. \end{aligned}$$

Here, each individual has, for reasons of symmetry, four frailties, two interpreted as genetic ( $G$ ) and two as environmental ( $E$ ) contributions. The genetic makeup of an individual stems from that individual's biological mother (A) and father, and the adoptive child (C) therefore shares one of the two genetic frailties with its biological mother ( $G_s^B \sim Z_{i3}$ ). Similarly, the adoptee shares (some) environment with the adoptive mother (B), ( $E_s^A \sim Z_{i1}$ ) but none (ignoring prenatal environmental influences) with the biological mother. The remaining frailties are specific to each individual ( $G_s^A + G_{ns}^A + E_{ns}^A \sim Z_{i2}$ ,  $G_{ns}^C + E_{ns}^C \sim Z_{i4}$  and  $G_{ns}^B + E_s^B + E_{ns}^B \sim Z_{i5}$ ). Note that there is no assumption about the magnitude of correlations between genetic frailties.

For this model,

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{pmatrix}.$$

### 13.4.7 Competing Risks

The competing risks model is

$$h_i(t) = \{(Z_{i0} + Z_{i1})\alpha_{g_1}(t) + (Z_{i0} + Z_{i2})\alpha_{g_2}(t)\}Y_i(t), \quad i = 1, \dots, n. \quad (13.9)$$

A correlated frailty model in this form with only two causes of death was used by Zahl (1997) to assess the excess hazard for patients with malignant

melanoma and colon cancer. It was also used by Yashin and Iachine (1995) to estimate a biological limit of human longevity in the hypothetical situation where all environmental causes of death are removed.

for this model,

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}.$$

The basic independence unit, which in the general formulation is the family, is now each subject. Associated with each subject, indexed by  $i$ , is a number of processes (here just two), one for each cause of death. Each cause of death has its own associated baseline hazard function, in the example above  $\alpha_{g_1}(t)$  and  $\alpha_{g_2}(t)$ , corresponding to different strata for each individual in the family. Although conceptually different, the competing risks model yields the same likelihood expressions as the correlated frailty model proposed here when noting that we will always observe at most one event for each family since individuals can die only once.

One of the important topics in competing risks theory is the assumption of independence between the intensities (or the latent life times) associated with each cause of death. The correlated frailty model offers a way of relaxing this crucial independence assumption in a straightforward way.

Explicit details of four different approaches to estimation can be found in Petersen (1998).

### 13.4.8 Example

Petersen (1998) analyzed adoption data described by Sorensen et al. (1988). We present his results here in this example. The study is comprised of 960 families that included children born during the period 1924 through 1926 who were placed early in life with adoptive parents unrelated to them. An analysis of all combinations of two for which one was the adoptee and the other one of the parents, e.g., adoptee and biological mother, was carried out by means of a Cox model where crude classifications of the life lengths of one or both biological or adoptive parents were used as covariates in the analysis. This gave estimates of the relative risks of dying among adoptees with a dead parent as compared with the risk among adoptees with the parent still alive.

For the data used here, the follow-up period is extended to 1987, 5 more years than the original study. For simplicity, data are restricted to families with no missing values (919). At the time of follow-up, 164 biological mothers, 190 adoptive mothers, and 47 adoptees had died from some type of infections - pneumonia, tuberculosis, bronchitis, and a number of other less prevalent types. Note that the official death certificates, from which the causes of death were obtained, may include up to three causes of death, one of which is infection. Although there seems to be a fairly large number of events for the estimation, there are only very few concordant deaths, 10 for biological mothers and adoptees and 4 for adoptive mothers and adoptees.

Because the scientific interest is in premature death, it was decided to

censor life times at age 70, which is about the age of retirement in many countries. It also makes the interpretation of the results more comparable to those in Sorensen (1988). A more technical reason for this artificial censoring is that we expect to improve the properties of the estimators by ensuring that a suitably large number of individuals are at risk of death at all times (see Nielsen et al., 1992). As the data set is rather thin, it was decided to specify baseline hazards through the parametric marginal approach discussed by Petersen (1998). In the example, four marginal distributions are estimated because, apart from distinguishing between biological and adoptive mothers, male and female adoptees are allowed different marginal survival. All marginal distributions are, for convenience, chosen as Gompertz distribution with two parameters, i.e., with a hazard function in the form  $h(t) = a \exp(bt)$ . Altogether,  $4 \times 2$  marginal + 3 frailty = 11 parameters are estimated.

The marginal distribution are estimated using data censored at age 80 and not at age 70, which is the point of censoring for the estimation of association parameters. Using these marginal survival parameters as if they were known in the likelihood yields an approximate profile likelihood that is a function of the parameters of the frailty distributions only. The estimates of the parameters and the standard deviations of the frailty distributions are given in [Table 13.1](#). From Table 13.1, it is clear that the small positive effect of the shared environment ( $\sigma_{E_s}$ ) is non-significant ( $P=0.62$ ). A significance test of the shared genetic effect ( $\sigma_{G_s}$ ), which is the other primary study objective, turns out clearly significant ( $P=0.002$ ). The genetic effect may be explained by differences in the strength of the immune systems of the individuals - differences that, to some extent, are determined by genetic factors.

Note that all tests are usual two-sided likelihood ratio tests and that an argument could be made for testing one-sided because parameters in model (13.1) cannot be negative. The P-values should then be divided by two, and the confidence interval correspondingly would become more narrow. As conclusions using the two approaches are the same, the more conservative two-sided tests are kept. A secondary hypothesis of interest is whether infections could be due to inheritance alone, i.e., that there is no environmental influence. Note that this is not a test of the hypothesis that the third parameter equals zero but rather that it equals the genetic parameter, i.e.,  $H_0 : \sigma_{G_{ns}+E_{ns}} = \sigma_{G_s}$ , or equivalently that the correlation between the frailty components of the biological mother and the adoptee is 0.5. This hypothesis is tested in a model with no effect of the shared environment, which means that the data of the adoptive mother do not contribute power to this test. Estimating in a model with only the biological mother and the adoptee yields estimates 1.55 and 1.50 for  $\sigma_{G_s}$  and  $\sigma_{G_{ns}+E_{ns}(+E_s)}$ , respectively. Estimating the correlation between the frailty components of the biological mother and the adoptee yields 0.51 with a test-based 95% confidence interval [0.060,1]. A test of the hypothesis that the correlation is equal to 0.5 is comfortably accepted ( $P=0.99$ ), and one finds the estimates in [Table 13.2](#).

Parameter	Estimate	SE	95% CI
$\sigma_{E_s}$	0.59	0.64	[0,3.05]
$\sigma_{G_s}$	1.47	0.75	[0.75,5.63]
$\sigma_{G_{ns}+E_{ns}}$	1.01	4.74	[0, 13.47]

TABLE 13.1: Parameter estimates in correlated frailty model of death by infection

Parameter	Estimate	SE	95% CI
$\sigma_{G_s}$	1.56	0.42	[0.78,2.56]

TABLE 13.2: Parameter estimates in correlated frailty model with a purely genetic effect

The analysis have been performed, as in the original presentation of the data, assuming the same dependence structure for male and female adoptees. Splitting the data into two parts depending on the sex of the adoptee yields two sets of parameter estimates. For male adoptees, we found 1.07 (1.08), 2.94 (1.07), and 3.40 (1.92) for  $\sigma_{E_s}$ ,  $\sigma_{G_s}$ , and  $\sigma_{G_{ns}+E_{ns}}$ , respectively, where the numbers in parenthesis are the standard errors. For female adoptees we found 0.29 (1.68), 0.77 (0.55), and 0.00 (4.23) for  $\sigma_{E_s}$ ,  $\sigma_{G_s}$ , and  $\sigma_{G_{ns}+E_{ns}}$ , respectively. From this, it is observed that the result from the analysis of the male adoptees that drives the combined analysis, whereas the analysis of the female adoptees is almost void of information. At the risk of over-interpreting the results, one is struck by the fact that a purely genetic effect is stronger between mother and son than between mother and daughter. Such a sex difference could be the result of a genetic mechanism linked to the X-chromosome. However, a test of identical dependence structure for male and female adoptees is accepted ( $-2 \ln Q = 4.5$ , d.f. = 3, P = 0.21).

Another way of estimating the frailty parameter is by means of the two-stage method discussed by Petersen (1998). Estimation using this method addresses in particular the question about the influence of the admittedly crude approximation to the observed marginal survival. This sort of validation was carried out on two models denoted BM-AC and AM-AC, where BM-AC signified that it is the association between the biological mother and the adoptive child that is studied and AM-AC signifies that it is the association between adoptive mother and the adoptive child that is studied. The result of this type of validation is, for BM-AC model, 2.85 (the shared genes) and 5.23 (the individual specific environment and genes) compared with 1.55 and 1.50 using the Gompertz marginal approach. There is a considerable deviation in the point estimates using these two approaches, but this deviation reflects the great uncertainty with which especially the second parameter is estimated

as well as the positive correlation between the estimates. Setting the second parameter equal to the value found using Gompertz marginals (1.50) yields an estimate of 1.62 (compared with 1.55 for the Gompertz marginal) for the standard deviation of the genetic frailty component. This new point estimate of log-likelihood value is 0.27 lower than the maximum found at 2.85 and 5.23. Estimating under the hypothesis of a purely genetic effect, i.e., that the correlation between the frailty components of the biological mother and the adoptee is 0.5, yields an estimate of 1.66 with test-based confidence interval [0.80, 2.90], which is to be compared with [Table 13.2](#). Note that there is no theoretical justification for the confidence intervals of either method since the marginals are fixed in the estimation. But because this is the case for both methods, it does make sense to compare them. Confidence intervals that include the uncertainty with which the marginals are estimated are expected to be wider than the ones reported by Petersen (1998). Compared to the uncertainty with which the parameters are estimated, there is a good correspondence between the two methods, indicating that a rather crude fit of the marginal distribution, here with a two-parameter Gompertz distribution, yields essentially the same results as those obtained when estimating the marginal distribution nonparametrically.

The great uncertainty with which the distribution of the subject-specific frailty is estimated renders the point estimates of the frailty parameters almost useless unless large sample sizes are available. In terms of testing whether there is a genetic or shared environmental effect, this is not major concern because the association between the life times of the individuals are picked up by the shared frailty components. It is rather in the process of determining the role of inheritance, e.g., the natural hypothesis that the etiology of the disease is purely genetic, that the uncertainty becomes relevant. This is a hypothesis about correlation. If the mode of the inheritance is known, the problem becomes smaller. Note, for example, the great improvement in the confidence interval a factor two in the upper limit after having imposed the constraint that the individual specific parameter is equal to the genetic parameter (correlation 0.5, compare [Tables 13.1](#) and 13.2).

As a result of the great uncertainty with which the individual-specific parameter is estimated, one is compelled to consider other parametrizations or combined measures. A relevant measure might be the relative importance of the shared environment versus genes, where an obvious candidate is  $\sigma_{E_s}^2 / \sigma_{G_s}^2$ . Estimate of this ratio is 0.16 with 95% confidence interval [0, 9.2]. Note that the estimate of the fraction is a restatement of the previous findings from [Table 13.1](#), where shared genes account for six times more of the variability than the shared environment. The wide confidence interval is now more a reflection of the uncertainty with which the genetic and environmental frailty distributions are identified. However, the interval, being so wide, is unable to exclude the importance of the shared environment.

## 13.5 Additive Genetic Gamma Frailty

In the additive genetic gamma frailty model studied by Korsgaard and Andersen (1998), individual frailties are correlated as a result of an additive genetic model. In some bivariate cases (e.g. parent-offspring or full sibs), the genetic Frailty model is a correlated frailty model with correlation equal to one half.

### Definition:

Additive genetic gamma frailties (genetic frailties)  $Z_1, \dots, Z_n$  for  $n$  individual are linear combinations of independent gamma distributed random variables. The linear combinations are given by  $E(Z_i) = 1$ ,  $i = 1, \dots, n$ , and  $\text{var}(Z_1, \dots, Z_n) = A\sigma_Z^2$ , where  $A$  is the numerator relationship matrix. The  $ij$ -th off-diagonal element of  $A$  is the numerator of Wright's (1922) coefficient of relationship between individuals  $i$  and  $j$  and the  $i$ -th diagonal element is  $1 + f_i$ , where  $f_i$  is Wright's (1922) coefficient of inbreeding for individual  $i$  (Quass, 1976).

The following two examples can both be derived from the algorithm treating the general case. The first example relates to the data actually analyzed and the second example deals with inbreeding.

### 13.5.1 Example 13.1

In a small family with a father ( $F$ ), a mother ( $M$ ), and a child ( $C$ ), let  $U_1$  represent the part of the father's genome affecting frailty that is transmitted to the child, and  $U_2$  the corresponding part of the father's genome not transmitted to the child.  $U_3$  and  $U_4$  represent a similar partitioning of the mother's genome affecting frailty (Figure 13.1).

Assuming that the father and mother are unrelated, additive genetic gamma frailties for the father  $Z_F$ , mother  $Z_M$ , and child  $Z_C$  are given by

$$\begin{aligned} Z_F &= U_1 + U_2 \\ Z_M &= U_3 + U_4 \\ Z_C &= U_1 + U_3. \end{aligned}$$

$U_i$ 's will either be used to represent part of the genome affecting frailty or the corresponding random variable used in the construction of frailties. The meaning should be clear from the context.

Assuming that the  $U_i$ 's are i.i.d., with  $U_i \sim \Gamma(\lambda/2, \lambda)$ , so that  $E(U_i) = 1/2$  and  $\text{var}(U_i) = 1/(2\lambda)$ , then

$$E(Z_F) = E(Z_M) = E(Z_C) = 1$$

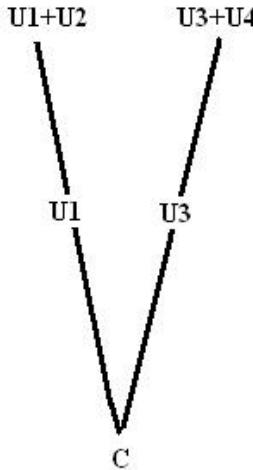


FIGURE 13.1: Father, mother, and child genetic diagram.

and

$$\text{var}(Z_F, Z_m, Z_C) = \begin{pmatrix} 1 & 0 & 1/2 \\ 0 & 1 & 1/2 \\ 1/2 & 1/2 & 1 \end{pmatrix} 1/\lambda = A\sigma_Z^2$$

as required with  $\sigma_Z^2 = 1/\lambda$ . The correlation between frailty of the father  $Z_F$  (or mother) and frailty of the child  $Z_C$  is  $1/2$  due to the fact that the two individuals share half their genes.

The above matrix  $A$  is the correlation matrix as well. This is not true when inbreeding is present, as shown in Example 13.2 ([Figure 13.2](#)).

### 13.5.2 Example 13.2

Consider four individuals 1, 2, 3, and 4. The individuals 1 and 2 are unrelated and have a common offspring 3. The offspring 4 is from the backcross of 3 with 2.

Additive genetic gamma frailties  $Z_1, Z_2, Z_3$ , and  $Z_4$  of the four individuals are given by

$$\begin{aligned} Z_1 &= U_1 + U_2 + U_3 \\ Z_2 &= V_1 + V_2 + V_3 + V_4 + V_5 + V_6 \\ Z_3 &= U_1 + U_2 + V_1 + V_2 + V_3 + V_4 \\ Z_4 &= U_1 + 2V_1 + V_2 + V_3 + V_5 \end{aligned}$$

where  $U_1, U_2, U_3, V_1, \dots, V_6$  are mutually independent.  $U_1$  and  $U_2$  are  $\Gamma(\lambda/4, \lambda)$  distributed and  $U_3 \sim \Gamma(\lambda/2, \lambda)$ .  $V_1, V_2, V_3$ , and  $V_4$  are  $\Gamma(\lambda/8, \lambda)$  distributed, and  $V_5$  and  $V_6$  are  $\Gamma(\lambda/4, \lambda)$  distributed. By this construction it is

easily seen that

$$E(Z_1) = E(Z_2) = E(Z_3) = E(Z_4) = 1$$

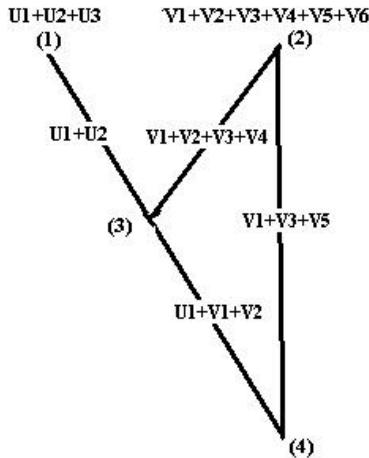


FIGURE 13.2: Father, mother, and 2 children genetic diagram.

and

$$\text{var}(Z_1, Z_2, Z_3, Z_4) = \begin{pmatrix} 1 & 0 & 1/2 & 1/4 \\ 0 & 1 & 1.2 & 3/4 \\ 1/2 & 1/2 & 1 & 3/4 \\ 1/4 & 3/4 & 3/4 & 5/4 \end{pmatrix} 1/\lambda = A\sigma_Z^2$$

as required, with  $\sigma_Z^2 = 1/\lambda$ .

Frailties constructed as the linear combinations given above can be justified genetically as follows: the offspring 3 of 1 and 2 received half its genome  $U_1+U_2$  from its father, 1, and half its genome  $V_1 + V_2 + V_3 + V_4$  from its mother, 2. Concerning 4, one-quarter of its genome,  $U_1$ , originates from 1 and three-quarters,  $2V_1 + V_2 + V_3 + V_5$ , originates from its mother, 2 (whereof  $V_1 + V_3 + V_5$  is transmitted from 2 directly and  $V_1 + V_2$  from 2 through its father, 3). By this 4 has inherited two copies of  $V_1$ , i.e.  $V_1$  is on both chromosomes, and genes at loci in this part are identically by descent.

### 13.5.3 Example 13.3

This section describes the preparations required to re-analyze data used in Nielsen et al. (1992) from the Danish adoptee register. Life length of 792 Danish adoptees born in the period 1924-1926, all of whom were alive at their

sixteenth birthday and life length of their biological parents are modeled by the additive genetic gamma frailty model. Data have a very simple family structure and are used to illustrate how the EM algorithm can be used to estimate the parameters in the additive genetic gamma frailty model. We proceed like Nielsen et al. (1992), but calculations in the additive genetic gamma frailty model are more complicated than for the shared frailty model.

Each of the 792 families, as well as the father and mother in each family, are assumed to be unrelated. The child is adopted very soon after birth and due to the fact that the child is adopted, it is likely that the parents did not live together. Therefore environmental correlations between any two lifetimes within family ought to be small. Modeling these data by the additive genetic gamma frailty model it is implicitly assumed that environmental correlations are absent. The model is:  $\mathbf{N} = (N_{ik} : i = 1, \dots, 792; k = 1, 2, 3)$  is a multivariate counting process with intensity process  $h$  given by

$$h_{ik}(t) = Z_{ik} Y_{ik}(t) \alpha_{g(i,k)}(t).$$

$N_{ik}(t)$  is one or zero according to whether individual  $(i, k)$  is observed dead at a time  $s \leq t$  or not.  $Y_{ik}(t)$  is an indicator function with value 1 if individual  $(i, k)$  is alive and under observation just before time  $t$ .  $\alpha_g$ 's are unknown base-line hazard functions stratified according to fathers, mothers, male and female offspring, respectively.  $A_g(\cdot)$  will denote integrated base-line hazard functions.  $Z_{ik}$ 's are additive genetic gamma frailties, for each family constructed as in Example 13.1.

$$\begin{aligned} Z_{i1} &= U_{i1} + U_{i2} \\ Z_{i2} &= U_{i3} + U_{i4} \\ Z_{i3} &= U_{i1} + U_{i3} \end{aligned}$$

$U_{ij}$ 's are i.i.d.  $\Gamma(\lambda/2, \lambda)$  distributed,  $i = 1, \dots, 792; j = 1, 2, 3, 4$ .  $i$  indicates family number, and  $k = 1, 2, 3$  corresponds to father, mother, and child, respectively. The function  $g$  is defined by

$$g(i, k) = \begin{cases} 1 & \text{if } k = 1 \\ 2 & \text{if } k = 2 \\ 3 & \text{if } (i, 3) \text{ is a male} \\ 4 & \text{if } (i, 3) \text{ is a female} \end{cases}$$

The observed, but incomplete, data are  $(\mathbf{N}, \mathbf{Y})$ , and the unobserved, but complete, data  $(\mathbf{N}, \mathbf{Y}, \mathbf{U})$ ,  $\mathbf{Y} = (Y(t))_{t \in R_+}$ , with  $\mathbf{Y}(t) = (Y_{ik}(t))_{i=1, \dots, n; k=1, 2, 3}$  the observed predictable process described above and  $\mathbf{N} = (\mathbf{N}(t))_{t \in R_+}$ , with  $\mathbf{N}(t) = N_{ik}(t)_{i=1, \dots, n; k=1, 2, 3}$ .  $\mathbf{U} = (U_{ij})_{i=1, \dots, n; j=1, 2, 3, 4}$ . The conditional intensity of  $\mathbf{N}$  given  $\mathbf{U} = \mathbf{u}$  is given by

$$h_{ik}(t) = z_{ik} Y_{ik}(t) \alpha_{g(i,k)}(t)$$

where  $z_{i1} = u_{i1} + u_{i2}$ ,  $z_{i2} = u_{i3} + u_{i4}$  and  $z_{i3} = u_{i1} + u_{i3}$ .

Specifying the intensity process of a counting process only specifies a partial likelihood (Andersen et al., 1988). In this case, where each counting process  $N_{ik}$  has one jump at the most, and under the assumption that conditional on  $\mathbf{U} = \mathbf{u}$ , censoring is independent (e.g., Arjas and Haara, 1984; Andersen et al., 1992) the partial conditional likelihood based on data  $(\mathbf{N}, \mathbf{Y})$  given  $\mathbf{U} = \mathbf{u}$  is given by

$$L_{\tau}^{N,Y|U}(\theta) = \prod_{i=1}^n \left[ \prod_{k=1}^3 \left( (z_{ik} Y_{ik}(t_{ik}) dA_{g(i,k)}(t_{ik}))^{\Delta N_{ik}(t_{ik})} \right. \right. \\ \left. \left. \exp \left\{ - \int_0^{\tau} z_{ik} Y_{ik}(s) dA_{g(i,k)}(s) \right\} \right) \right]$$

(Gill and Johansen, 1990).  $\theta = (\lambda, A_1(\cdot), A_2(\cdot), A_3(\cdot), A_4(\cdot))$  and  $t_{ik}$  is the observed survival time of individual  $(i, k)$  or the age at which it was censored.  $\tau$  denotes the end of the observation period (e.g.  $\tau = \infty$ ). The complete data partial likelihood  $L_{\tau}^{N,V,U}(\theta)$  is the product of  $L_{\tau}^{N,Y|U}(\theta)$  and the density of  $\mathbf{U}$ , that is

$$L_{\tau}^{N,Y|U}(\theta) = \prod_{i=1}^n \prod_{k=1}^3 \left[ z_{ik} Y_{ik}(t_{ik}) dA_{g(i,k)}(t_{ik})^{\Delta N_{ik}(t_{ik})} \right. \\ \times \exp \left\{ - \int_0^{\tau} z_{ik} Y_{ik}(s) dA_{g(i,k)}(s) \right\} \left. \right] \\ \times \prod_{i=1}^n \prod_{j=1}^4 \left[ \frac{\lambda^{\lambda/2}}{(\lambda/2)} u_{ij}^{\lambda/2-1} \exp(-\lambda u_{ij}) \right]$$

$L^{N,V,U}(\theta)$  is a full likelihood for  $\theta$  if conditional on  $\mathbf{U} = \mathbf{u}$  censoring is non-informative on  $\theta$  (e.g., Andersen et al., 1992). Explicit details of estimation procedures by EM algorithm are discussed by Korsgaard and Andersen (1998).

### 13.5.4 Application in Danish Adoptive Register Data

We present here the Danish adoptive register data for the analysis of additive frailty model. Since time is from birth of each individual, individuals are not at risk of being observed to die from age zero. Children belong to the data set conditional on whether they were alive at their sixteenth birthday and are considered to be at risk from that age. Mothers are considered to be at risk from delivery and fathers from conception, assumed to take place 280 days before delivery. This means that data beyond being right censored are left truncated. Results analyzing data unrestricted and data artificially censored at time  $t = 70$  are given in Table 13.3. In Table 13.3, second column, deaths denote the number of deaths among fathers/mothers/male offsprings/female offsprings and  $-2\ln Q$  is used to test goodness of fit of the model without frailty.

Lifetimes	Deaths	$\hat{\xi}$	$\ln L(\hat{\xi})$	$\ln L(0)$	$-2 \ln Q$	P
unrestricted	702/566/92/63	0	-9655.41	-9655.41	0	1
censored	337/274/92/63	0.26	-5506.41	-5507.41	2.01	0.15

TABLE 13.3: Estimation of frailty parameter in the additive frailty model

Analyzing lifetimes unrestricted,  $\hat{\xi} = 0$  is found to be the value of  $\xi = 1/\lambda$  that maximizes the profile log-likelihood of  $\xi$  for  $\xi \in R_+$ . Artificially censoring data at time  $t = 70$  gives  $\xi = 0.26$ , however, the  $p$ -value is low (0.1-0.2) but not significantly different from zero, the model without frailty. Because of the relatively small number of deaths in the group of offspring and because no offspring had the possibility to die after age 70, the conclusion that mortality does not seem to be heritable should be taken cautiously.

Nielsen et al. (1992) used the shared frailty model and analyzed among others the bivariate relationships: father and child, mother and child relationships, with lifetimes unrestricted and lifetimes artificially censored at age 70. A frailty parameter significantly different from zero was found only for the mother-child relationship, and then only for lifetimes artificially censored at age 70. Because a frailty parameter is significantly different from zero, in the shared frailty model, which can be due to environmental similarities as well as genetic similarities, they concluded that the risk of early death may be influenced by genetic background. If the additive genetic frailty model described in Example 13.3 is correct, results analyzing pairwise data on fathers (mothers) and children should be consistent with those obtained analyzing all of the fathers, mothers and children. The additive genetic frailty model for pairwise data is given by:

$$h_{ik}(t) = Z_{ik} Y_{ik}(t) \alpha_{g(i,k)}(t).$$

$i = 1, \dots, 792; k = 1, 2$ .  $k = 1$  represents the father (mother) and  $k = 2$  the child. Baseline hazard functions are  $\alpha_g$ ,  $g = 1, 2, 3$  according to fathers (mothers), male and female offsprings, respectively. The additive genetic gamma frailties are given by

$$Z_{i1} = U_{i1} + U_{i2}$$

$$Z_{i2} = U_{i1} + U_{i3}$$

where  $U_{i1}, U_{i2}, U_{i3}$  are i.i.d.  $\Gamma(\lambda/2, \lambda)$  distributed random variables. Results using the EM algorithm are given in Table 13.4. In Table 13.4, second column, deaths denote the number of deaths among fathers (mothers)/male offsprings/female offsprings and  $-2\ln Q$  is used to test goodness of fit of the model without frailty.

Lifetimes	Deaths	$\hat{\xi}$	$\ln L(\hat{\xi})$	$\ln L(0)$	$-2 \ln Q$	P
<u>Fathers and children</u>						
unrestricted	702/92/63	0.03	-5747.69	-5747.70	0.02	0.93
censored	337/92/63	0.25	-3494.51	-3494.96	0.89	0.35
<u>Mothers and children</u>						
unrestricted	566/92/63	0.58	-4942.12	-4944.16	4.08	0.035
censored	274/92/63	1.31	-3043.83	-3048.92	10.18	0.003

TABLE 13.4: Estimation of frailty parameter in the additive frailty model of fathers (mothers) and children

Analyzing fathers and children, the frailty parameter was not found to be significantly different from zero whether lifetimes were considered unrestricted or censored at age 70. The frailty parameter was found to be significantly different from zero analyzing mothers and children. These results are not consistent with those obtained analyzing all of the fathers, mothers, and children. The explanation could be that the model is wrong and/or that environmental correlations are present. The environmental and/or genetic similarities were further examined by analyzing pairwise data by the correlated frailty model described in Yashin et al. (1995) and Yashin and Iachine (1995) given by

$$h_{ik}(t) = Z_{ik} Y_{ik}(t) \alpha_{g(i,k)}(t).$$

Now frailties  $Z_{i1}$  and  $Z_{i2}$  are correlated with correlation coefficient  $\rho$ ,  $0 \leq \rho \leq 1$ . Frailties are given by

$$Z_{i1} = U_{i1} + U_{i2}$$

$$Z_{i2} = U_{i1} + U_{i3}$$

where  $U_{i1}, U_{i2}, U_{i3}$  are independent gamma random variables.  $U_{i1}$  is  $\Gamma(\alpha_1, \lambda)$  distributed, and both  $U_{i2}$  and  $U_{i3}$  are  $\Gamma(\alpha_2, \lambda)$  distributed. For reasons of identifiability, the constraint  $\alpha_1 + \alpha_2 = \lambda$  is imposed. The correlation  $\rho$  is equal to  $\alpha_1/\lambda$ . Again the EM algorithm was used to estimate the parameters; the results are given in Table 13.5.  $-2\ln Q$  is used to test goodness of fit of the additive genetic gamma frailty model.

The analysis of fathers and children revealed that the correlation  $\rho$  was not significantly different from 1/2, whether lifetimes were considered unrestricted or censored at age 70, and referring to the previous analysis, not significantly different from the model without frailty. Analyzing mothers and children, maximum was found for  $\rho = 0(0)$  and  $\xi = 20(30)$  analyzing lifetimes unrestricted (censored at age 70). The correlation  $\rho$  was found to be significantly different from 1/2.  $\rho = 0$  is the case with individually uncorrelated frailties. Without

Lifetimes	Deaths	$\hat{\rho}$	$\hat{\xi}$	$\ln L(\hat{\xi})$	$\ln L(0)$	$-2 \ln Q$	P
<u>Fathers and children</u>							
unrestricted	702/92/63	1	0.04	-5747.60	-5747.69	0.18	0.65
censored	337/92/63	1	0.16	-3494.21	-3494.51	0.6	0.45
<u>Mothers and children</u>							
unrestricted	566/92/63	0	20	-4923.34	-4942.12	37.56	0.0005
censored	274/92/63	0	30	-3003.30	-3043.83	81.06	0.0005

TABLE 13.5: Estimation of frailty parameter in the correlated frailty model of fathers (mothers) and children

covariates, this model is unidentifiable (mentioned in e.g., Nielsen et al., 1992) and the partial maximized likelihood ought to be constant as a function of  $\xi$ ,  $\xi \in R_+$ . During the above analysis, it was found that the likelihood does not satisfy this requirement. A little example that analytically shows that the partial likelihood has a “defect” is given in Korsgaard and Andersen (1995). The problem is not solved but further work is needed. Finally it is mentioned, how a test for goodness of fit of the additive genetic gamma frailty model analyzing data on all the fathers, mothers, and children could be performed. The frailty model with an arbitrary (positive) correlation structure between all of the fathers, mothers, and children is given by

$$h_{ik}(t) = Z_{ik} Y_{ik}(t) \alpha_{g(i,k)}(t)$$

with

$$Z_{i1} = U_i + V_{i1} + V_{i2} + W_{i1}$$

$$Z_{i2} = U_i + V_{i1} + V_{i3} + W_{i2}$$

$$Z_{i3} = U_i + V_{i2} + V_{i3} + W_{i3}.$$

Within family, the idea is:  $U_i$  is the part of frailty that all of the three individuals share,  $V_{i1}, V_{i2}$ , and  $V_{i3}$  are parts of frailties that the father and mother, father and child, and mother and child share, respectively, and  $W_{i1}, W_{i2}$ , and  $W_{i3}$  the parts of frailties that are individual for each father, mother and child, respectively. Assuming that all the  $U$ 's,  $V$ 's, and  $W$ 's are independent, and that  $U_i$  is  $\Gamma(\alpha, \lambda)$  distributed, and  $V_{ik}$  is  $\Gamma(\alpha_k, \lambda)$  distributed, and  $W_{ik}$  is  $\Gamma(\beta_k, \lambda)$  distributed,  $k = 1, 2, 3$ , then, for reasons of identifiability, imposing the constraint  $\alpha + \alpha_1 + \alpha_2 + \beta_1 = \alpha + \alpha_1 + \alpha_3 + \beta_2 = \alpha + \alpha_2 + \alpha_3 + \beta_3 = \lambda$ , we get

$$\text{var}(Z_{i1}, Z_{i2}, Z_{i3}) = \frac{1}{\lambda^2} \begin{pmatrix} \lambda & \alpha + \alpha_1 & \alpha + \alpha_2 \\ \alpha + \alpha_1 & \lambda & \alpha + \alpha_3 \\ \alpha + \alpha_2 & \alpha + \alpha_3 & \lambda \end{pmatrix}.$$

This is seen to be the shared frailty model if  $\alpha_1 = \alpha_2 = \alpha_3 = \beta_1 = \beta_2 = \beta_3 = 0$  and  $\alpha = \lambda$ . It is additive genetic gamma frailty model if  $\alpha = \alpha_1 = \beta_3 = 0$  and  $\alpha_2 = \alpha_3 = \beta_1 = \beta_2 = \lambda/2$ . If  $\alpha = \alpha_1 = \alpha_2 = \alpha_3 = 0$  and  $\beta_1 = \beta_2 = \beta_3 = \lambda$ ,

then it becomes the unidentifiable individual frailty model. The model without frailty is obtained if  $\lambda = \infty$ . And so on. In principle the additive genetic gamma frailty model can be used to analyze family structures of any complexities. The only deficiency being that the complexity of the computations grows exponentially with family size. In the example, we have 792 pedigrees, each with three individuals, and the total number of  $U$ 's required equals  $792 \times 4$ . In animal breeding applications one can easily imagine a situation with just a few or only one pedigree with thousands or even millions of related animals in each.

---

### 13.6 Additive Genetic Gamma Frailty for Linkage Analysis of Diseases

Linkage analysis of diseases with complex and unknown modes of inheritance remains problematic. Genetically complex diseases may exhibit one or more of the following properties: incomplete penetrance, involvement of several genes, interaction effects between these genes, and heterogeneity due to different genes or environmental risk factors. For many complex diseases, studies have suggested that disease genes influence not only the occurrence of the disease, but also the age at onset. Examples include breast cancer (Claus et al., 1990), prostate cancer (Carter et al., 1992), and Alzheimer disease (Meyer et al., 1998). In fact, affected relatives with different age of onset may be the result of different genetic etiologies and unaffected relatives are censored at the time of the study.

The aim of linkage analysis is to extract inheritance information from pedigrees and to test for co-segregation of genetic markers and disease. There are two broad classes of methods, parametric model-based, and non-parametric model-free methods. The parametric Lod score method, assuming a specific mode of inheritance for a trait-causing gene and also the probability of developing disease conditional on genotype (i.e., penetrance function), is widely used. However, parametric linkage analysis can be highly sensitive to misspecification of the linkage model (Clerget-Darpoux et al., 1986) and can consequently lead to considerable loss of power. The non-parametric model-free methods concentrate on affected relative pairs or affected family members (Weeks and Lange, 1988; Kruglyak et al., 1996). These methods test whether there is an excessive allele sharing among the affected relatives as measured by either sharing the same allele or sharing the same allele from a common ancestor (i.e., allele identical by descent or IBD in short). A number of studies (Morton and Kidd, 1980; Haynes et al., 1986) have evaluated the effects of failure to correct for age of onset on standard Lod score linkage analysis. Their findings emphasized that adjustment for variable age of onset is an important

component of effective linkage analysis. Traditionally, the age of onset data are incorporated into model-based linkage analysis by either specifying different penetrance functions for individuals with different ages or age of onset or assuming a parametric distributional form for the age of onset distribution. These include methods implemented in LIPED (Ott, 1974; Ott, 1976) and GENEHUNTER (Kruglyak et al., 1996) programs. The problem with these approaches is that the penetrance functions or the distribution function can be difficult to specify for complex diseases. For model-free tests, Li and Hsu (2000) demonstrated that the age of onset of the affected siblings can affect the IBD probabilities and therefore the power of the mean IBD test. Thus, naively combining sib pairs with different age of onset can reduce the power of the mean IBD test. One approach is to extend the affected sib pair methods to incorporate age of onset data (Dawson et al., 1990; Flanders and Khoury, 1991; Commenges, 1994). These methods are based on pair-wise allele sharing between two affected sibs or two affected relatives, which can result in lower power of detecting linkage as compared to methods that consider allele sharing among all affected sibs (Kruglyak et al., 1996; Li and Huang, 1998). In addition, a specific distribution of age of onset has to be assumed for the methods of Dawson et al. (1990) and Flanders and Khoury (1991). Another approach is to extend the variance components method to incorporate age of onset data (Daw et al., 1999; Duggirala et al., 1999) with age of onset treated as a quantitative trait. Again for these methods, some parametric distribution such as a multivariate normal distribution or t-distribution for age of onset has to be assumed.

Li and Zhong (2002) extended the model in Li (1999) to include possible effects on the risk of developing disease due to loci not linked to the putative disease locus. Zhong and Li (2002) further extended the one-locus genetic gamma model in Li and Zhong (2002) to a two-locus genetic gamma frailty model in order to simultaneously consider multi-point allele sharing in two genetic regions, and to incorporate age of onset and environmental covariates data in linkage analysis. Li (2002) further extended the additive genetic gamma frailty model of Li (1999) to incorporate inheritance vector information derived from multiple markers and possible variance due to polygenes for nuclear families. The strategy is in contrast to parametric linkage models, where the effect of genotype at a disease locus is characterized by the penetrance function and the likelihood function involves summing over genotypes at the disease locus. We do not differentiate between types of alleles at a disease locus, but instead assign a random effect (frailty) to each founder allele. These frailties each have the same distribution, so that risk alleles are founder alleles that have a high value. Because one can assume that the frailty value of each founder allele is transmitted through the pedigree along with the allele, a mechanism is created for some persons to receive a high risk allele while others receive a low risk allele. However, because Li (2002) never assign an actual risk to each allele (instead the values are integrated over the frailty distribution), the indication that different alleles at the disease locus are as-

sociated with disease is measured by the variability in the hazard distribution at the disease locus. In this way, the model resembles a variance component model, and the correlation structure between individual frailties depends on the inheritance vector at the putative disease locus.

### 13.6.1 Genetic Frailties Defined by Multiple Unlinked Disease Loci

Let us consider only the nuclear family with a father (F), a mother (M), and  $k$  children ( $O_1, \dots, O_k$ ). Assuming that the father and mother are unrelated, there are only four unique alleles that are distinct by descent at a given locus. Arbitrarily label the paternal alleles as 1 and 2 and the maternal alleles as 3 and 4. Consider the setting of Kruglyak et al. (1996), in which we have a series of markers on a chromosomal region that may harbor the disease-causing locus/loci. Suppose  $d$  is a point in the test chromosomal region. We are interested in testing whether the disease susceptibility (DS) locus is linked to locus  $d$ . Let  $l_1, l_2, \dots, l_S$  be the remaining  $S$  disease predisposing loci in addition to a possible DS locus linked to  $d$ . Here, we assume that the loci  $d, l_1, l_2, \dots, l_S$  are not linked. The inheritance pattern at each locus is completely described by an inheritance vector (Kruglyak et al., 1996; Lander and Green, 1987), whose coordinates describe the outcome of the paternal and maternal meioses given rise to the  $k$  sibs. For a sibship with  $k$  sibs, the inheritance vector at the  $d$  locus is the  $2k$ -vector

$$V_d = (v_1, v_2, \dots, v_{2k-1}, v_{2k}),$$

where for  $1 \leq i \leq k$ ,  $v_{2i-1} = 1$  or 2 depending on which paternal allele the  $i$ -th sib receives and  $v_{2i} = 3$  or 4 depending which maternal allele the  $i$ -th sib receives. Similarly, the inheritance vector at the  $l_s$  locus is

$$W_{l_s} = (w_{s1}, w_{s2}, \dots, w_{2k-2}, w_{2k})$$

for  $s = 1, \dots, S$ , where for  $1 \leq i \leq k$ ,  $w_{s2i-1} = 1$  or 2 and  $w_{s2i} = 3$  or 4. The inheritance vector  $V_d$  and  $W_{l_s}$  indicate which parts of the genome at locus  $d$  or  $l_s$  are transmitted to the  $k$  children from the father and the mother during the meioses.

For a given inheritance vector  $v_d$  at the locus  $d$ , and  $w_{l_s}$  at the locus  $l_s$ ,  $s = 1, \dots, S$ , we define the additive genetic gamma frailties for the parents as

$$Z_F = U_{d1} + U_{d2} + \sum_{s=1}^S (U_{l_s 1} + U_{l_s 2}),$$

$$Z_M = U_{d3} + U_{d4} + \sum_{s=1}^S (U_{l_s 3} + U_{l_s 4}),$$

where  $U_{d1}$  and  $U_{d2}$  represent the genetic frailties due to the part of the genome

on the two chromosomes of the father at location  $d$ ;  $U_{d3}$  and  $U_{d4}$  are analogously defined for the mother. Similarly,  $U_{l_s 1}, U_{l_s 2}$  represent the genetic frailties due to the part of the genome on the two chromosomes of the father at the locus  $l_s$ ;  $U_{l_s 3}, U_{l_s 4}$  are analogously defined for the mother. Conditioning on the inheritance vectors  $V_d = v_d$  and  $W_{l_s} = w_{l_s}$  for  $s = 1, 2, \dots, S$ , we define the frailty for the  $j$ -th child as

$$Z_{O_j} = U_{dv_{2j-1}} + U_{dv_{2j}} + \sum_{s=1}^S (U_{l_s w_s 2j-1} + U_{l_s w_s 2j}),$$

for  $j = 1, \dots, k$ . This definition is based on the fact that it is the parts of the genome of the parents that are transmitted to the set of  $k$  offsprings, and the inheritance vectors indicate which parts are transmitted. We assume that the  $U_{d1}, U_{d2}, U_{d3}$ , and  $U_{d4}$  are independently and identically distributed as  $\Gamma(v_d/2, \eta)$ , and that  $U_{l_s 1}, U_{l_s 2}, U_{l_s 3}$ , and  $U_{l_s 4}$  are independently and identically distributed as  $\Gamma(v_s/2, \eta)$ , for  $s = 1, 2, \dots, S$ , where the parameter  $\eta$  is the inverse scale parameter and the parameters  $v_d$  and  $v_s$  are shape parameters. It is easy to verify that both the conditional (on the inheritance vectors) and the marginal means of the frailties are

$$E(Z_F) = E(Z_M) = E(Z_{O_1}) = \dots = E(Z_{O_k}) = \frac{v_d + \sum_{s=1}^S v_s}{\eta},$$

and both the conditional and marginal variance of the frailties are

$$\text{Var}(Z_F) = \text{Var}(Z_M) = \text{Var}(Z_{O_1}) = \dots = \text{Var}(Z_{O_k}) = \frac{v_d + \sum_{s=1}^S v_s}{\eta^2},$$

which is sum of the variance due to DS gene linked to locus  $d$  and DS genes at the remaining loci  $l_s$  for  $s = 1, \dots, S$ .

### 13.6.2 Expected Genetic Frailties over the Inheritance Vectors

In practice, we do not know the number of possible disease-predisposing genes,  $S$ , nor the locations of these genes. We are generally interested in the examination of one locus, say locus  $d$ , at a time. The inheritance vector at locus  $d$  is usually unknown. However, the probability distribution of the inheritance vector at  $d$ , denoted by  $P(v_d)$ , can be calculated by Lander and Green's hidden Markov model algorithm (Kruglyak et al., 1996; Lander and Green, 1987) using multiple markers in the test region. Since the locations of the remaining disease loci are unknown, the inheritance vectors at these remaining loci,  $W_{l_s}$ , for  $s = 1, \dots, S$ , are also unknown. In the absence of any information on these remaining loci, all inheritance vectors are equally likely according to Mendel's first law, and the probability distribution is uniform for each of the remaining loci. Li (2002) defined the expected (over the probability

distributions of the inheritance vectors at the loci  $d$  and  $l_s$  for  $s = 1, \dots, S$ ) genetic frailties for the offsprings as

$$G_{O_j} = E_{V_d, W_{l_s}}(Z_{O_j})$$

$$= E_{V_d}(U_{dv_{2j-1}} + U_{dv_{2j}}) + 0.5 \sum_{s=1}^S (U_{l_s 1} + U_{l_s 2}) + 0.5 \sum_{s=1}^S (U_{l_s 3} + U_{l_s 4}),$$

for  $j = 1, \dots, k$ , where  $E_{V_d, W_{l_s}}$  means the expectation over the distributions of the inheritance vectors  $V_d$  and  $W_{l_s}$ , for  $s = 1, \dots, S$ . Let

$$U_{pF} = \sum_{s=1}^S (U_{l_s 1} + U_{l_s 2}),$$

$$U_{pM} = \sum_{s=1}^S (U_{l_s 3} + U_{l_s 4}).$$

These are the genetic frailties for the parents due to the effects of all the remaining  $S$  loci, which follow a  $\Gamma(v_p, \eta)$  distribution, where  $v_p = \sum_{s=1}^S v_s$ . Then the genetic frailties of the parents can be written as

$$G_F = U_{d1} + U_{d2} + U_{pF}, \quad (13.10)$$

$$G_M = U_{d3} + U_{d4} + U_{pM}, \quad (13.11)$$

and the expected genetic frailty for the  $j$ -th offspring can be written as

$$\begin{aligned} G_{O_j} &= E_{V_d}(U_{dv_{2j-1}} + U_{dv_{2j}}) + 0.5(U_{pF} + U_{pM}) \\ &= \sum_{i=1}^4 b_{ji} U_{di} + 0.5(U_{pF} + U_{pM}) \end{aligned} \quad (13.12)$$

for the  $j$ -th offspring where  $b_{ji}$  depends on the probability distribution of the inheritance vectors  $V_d$ . It is to see that  $b_{ji} = \sum_{v_d} P(v_d) I(dv_{2j-1} = i)$ , for  $i = 1, 2$  and  $b_{ji} = \sum_{v_d} P(v_d) I(dv_{2j} = i)$ , for  $i = 3, 4$  and  $\sum_{i=1}^4 b_{ji} = 2$ , where  $I(\cdot)$  is the indicator function. The expected frailty vector can also be written in matrix notation as

$$G = HU \quad (13.13)$$

where

$$G = \begin{pmatrix} G_F \\ G_M \\ G_{O_1} \\ \vdots \\ G_{O_k} \end{pmatrix}, \quad U = \begin{pmatrix} U_{d1} \\ U_{d2} \\ U_{d3} \\ U_{d4} \\ U_{pF} \\ U_{pM} \end{pmatrix}, \quad H = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \\ b_{11} & b_{12} & b_{13} & b_{14} & 0.5 & 0.5 \\ \vdots & & & & \ddots & \\ b_{k1} & b_{k2} & b_{k3} & b_{k4} & 0.5 & 0.5 \end{pmatrix}.$$

### 13.6.3 Model for Age of Onset Data in a Family

We now introduce a frailty model to relate the age of onset of the disease and the expected genetic frailty defined in the previous subsection. Suppose there are  $n$  nuclear families, indexed by  $i = 1, \dots, n$ , each of which has  $n_i$  individuals. In the following presentation, we drop the subscripts for parents (M, F) and offsprings (O) from the notation. We define some further notation. The subscript  $ik$  indicates the  $k$ -th individual in the  $i$ -th family, where  $k = 1, \dots, n_i$ , and  $i = 1, \dots, n$ . Let  $T_{ik}$  be the age of onset,  $C_{ik}$  be the censoring age,  $t_{ik} = \min(T_{ik}, C_{ik})$ , and  $\delta_{ik} = I(t_{ik} = T_{ik})$ . Thus,  $\delta_{ik} = 1$  if the individual  $ik$  is affected, and 0 otherwise. The observed data are thus  $(t_{ik}, \delta_{ik})$ . Let  $(t, \delta) = \{(t_{ik}, \delta_{ik}), i = 1, \dots, n; k = 1, \dots, n_i\}$  be the vector of the age of onset or age at censoring for all  $n$  families. Let  $Y_{ik}$  be a  $p$ -dimensional vector of covariates which are independent of the genotype. Let  $G_{ik}$  be the expected genetic frailties for each family constructed as in Equations (13.10)-(13.12), where  $U_{id1}, U_{id2}, U_{id3}, U_{id4}$  are independently identically distributed as  $\Gamma(v_d/2, \eta)$  and  $U_{ipF}, U_{ipM}$  are independently and identically distributed as  $\Gamma(v_p, \eta)$  for  $i = 1, \dots, n$ . Finally, let  $U_i = (U_{id1}, U_{id2}, U_{id3}, U_{id4}, U_{ipF}, U_{ipM})$  be the vector of the frailties associated with the parents of the  $i$ -th family.

Conditional on the individual-specific genetic frailty  $G_{ik}$ , we assume that the age of onset data within a family are independent with a hazard function of the  $ik$ -th individual given by

$$h_{ik}(t|G_{ik}) = G_{ik} h_0(t) \exp(Y_{ik}\beta), \quad (13.14)$$

where  $h_0(t)$  is the unspecified baseline hazard function and  $\beta$  is the vector of regression parameters corresponding to the covariate vector  $Y_{ik}$ . The mean frailty of individuals is restricted to one to ensure that  $h_0(t)$  can be identified; i.e., set  $v_p + v_d = \eta$ . This allows an interpretation of  $h_0(t)$  as the baseline hazard function of an ‘average’ individual with  $Y_{ik} = 0$ . Let  $\Theta = \{H_0(t), \beta, \eta, v_p, v_d\} = \{\omega, \eta, v_p, v_d\}$  be the vector of the parameters associated with models (13.13) and (13.14),  $H_0(t)$  denote the cumulative baseline hazard function, and  $\omega = \{H_0(t), \beta\}$ .

### 13.6.4 Conditional Hazards Ratio for Sib Pairs

Since the genetic frailties within a family are dependent due to segregation of a possible disease gene, the hazard function, and therefore the age of onset, for persons within a family are dependent. In order to characterize such dependence structure of the random variable age of onset, we derive the conditional hazard ratio function, a measure widely used in the frailty model literature (e.g., Oakes, 1989; Clayton, 1978). Specifically, consider sib pair  $(l, m)$ . Let  $(T_l, T_m)$  be the random variables of age of onset of  $l$  and  $m$  respectively, and define the conditional hazard ratio

$$\theta(t_l, t_m) = \frac{h(t_l|T_m = t_m)}{h(t_l|T_m > t_m)}$$

where  $h(t_l|T_m = t_m)$  is the instantaneous probability of the  $l$ -th sib having the disease at age  $t_l$  given that the  $m$ -th sib is affected with disease at age  $t_m$ . The function  $h(t_l|T_m > t_m)$  is defined similarly given that the  $m$ -th sib is disease-free at age  $t_m$  (Oakes, 1989; Clayton, 1978). Note that  $\theta(t_l, t_m)$  is a function of both the current age of the  $l$ -th sib and the age of onset of the  $m$ -th sib. This measure of disease aggregation allows the risk of the relatives to depend on the age of onset of the proband. To relate the conditional hazard ratio parameter  $\theta(t_l, t_m)$  to the numbers of IBD shared at locus  $d$  ( $IBD_d$ ) and the parameters associated with the gamma frailty, Li (2002) obtained the conditional hazard ratios for sib pairs  $l$  and  $m$  given different values of  $IBD_d$  for any unspecified baseline hazard function. When there are effects due to both DS locus linked to  $d$  and polygenes, i.e., when  $v_d \neq 0$  and  $v_p \neq 0$ ,  $\theta(t_l, t_m)$  is an increasing function of the numbers  $IBD_d$ . When  $v_d = 0$ , the variance of the frailty due to the gene linked to the locus  $d$  is zero and  $\theta(t_l, t_m)$  does not depend on the IBD number at the locus  $d$ . Therefore, the test of the null hypothesis that the DS gene is not linked to  $d$  can be formulated as a test of  $v_d$  equaling zero. Hence we have,

$$\begin{aligned}
\theta(t_l, t_m) &= 1 + \frac{2v_p H_1^* H_2^*}{(v_d H_{12}^* + 2v_p H_1^*)(v_d H_{12}^* + 2v_p H_2^*)}, \quad V_d \neq 0, IBD_d = 0 \\
&= 1 + \frac{H_1^* H_2^* [v_d/2H_{12}^{*2} + 2v_p H_{12}^2]}{A \times B}, \quad V_d \neq 0, IBD_d = 1 \\
&= 1 + \frac{v_d H_{12}^{*2} + 2v_p H_{12}^2}{(v_d H_{12}^* + 2v_p H_{12})^2}, \quad V_d \neq 0, IBD_d = 2 \\
&= 1 + \frac{1}{2v_p}, \quad V_d = 0, IBD_d = 0 \\
&= 1 + \frac{1}{2v_p}, \quad V_d = 0, IBD_d = 1 \\
&= 1 + \frac{1}{2v_p}, \quad V_d = 0, IBD_d = 2
\end{aligned}$$

where

$$\begin{aligned}
A &= v_d/2H_{12}^*(H_1 + H_{12}^*) + 2v_p H_{12} H_1^* \\
B &= v_d/2H_{12}^*(H_2 + H_{12}^*) + 2v_p H_{12} H_2^* \\
H_1^* &= H_0(t_l) + \eta \\
H_2^* &= H_0(t_m) + \eta \\
H_{12} &= H_0(t_l) + H_0(t_m) + \eta \\
H_{12}^* &= H_0(t_l) + H_0(t_m) + 2\eta \\
\eta &= v_d + v_p.
\end{aligned}$$

### 13.6.5 Estimation Methods and Test of Linkage

Under the assumption that conditional on the genetic frailties  $G = g$ , censoring is independent and non-informative of  $g$  (Arjas and Haara, 1984; Andersen et al., 1992), the likelihood function  $L(t, \delta|\omega, \eta, v_d, v_p)$  based on the model (13.14) is given as

$$L(t, \delta|\omega, \eta, v_d, v_p) = \prod_{i=1}^n L_i(t_i, \delta_i|\omega, \eta, v_d, v_p), \quad (13.15)$$

where  $L_i(t_i, \delta_i|\omega, \eta, v_d, v_p)$  is the likelihood for the  $i$ -th family. If we can specify a parametric model for the underlying hazard function with a small number of unknown parameters, numerical optimization techniques can be used to find maximum likelihood estimates of the parameters. For unspecified baseline hazard function  $h_0(t)$ , an EM algorithm (Dempster et al., 1977), similar to those in Nielsen et al. (1992), and Korsgaard and Andersen (1998) can be developed for parameter estimation. Here, the EM algorithm for unspecified baseline hazard function corresponds to calculating the expected frailties, and then calculating the Nelson-Aalen estimator using the estimated  $\beta$ s and frailties.

Under the additive genetic gamma frailty model, we can test the null hypothesis that the variance of age of onset due to the putative disease gene linked to the locus  $d$  equals zero, i.e.,  $v_d = 0$ , by comparing the likelihood of this restricted model with the likelihood (13.15) under the full model in which the variance due to locus  $d$  is estimated. Under the restricted model,  $U_{idj} = 0$  for all  $i = 1, \dots, n$  and  $j = 1, 2, 3, 4$ . We denote the likelihood function under this restricted model as  $L_0(t, \delta|\omega, \eta, v_p, v_d = 0)$ . We define a Lod score statistic as

$$Lod = \ln_{10} \frac{\max L(t, \delta|\omega, \eta, v_p, v_d)}{\max L_0(t, \delta|\omega, \eta, v_p, v_d = 0)}$$

where under the full model,  $\eta = v_p + v_d$ , and under the restricted model  $\eta = v_p$ . Since we restrict  $v_d \geq 0$ , the asymptotic distribution of  $2 \ln_{10} Lod$  under the null hypothesis is a 50-50 mixture of 0 and  $\chi_1^2$  (Self and Liang, 1987). The significance value for the Lod score is therefore 0.588 for a significance level of  $\alpha = 0.05$ , and 2.074 for  $\alpha = 0.001$ .

### 13.6.6 Breast Cancer Data Example

Hall et al. (1990) provided strong evidence of linkage of early-onset familial breast cancer to chromosome 17q21, marker D17S74 by using a parametric linkage analysis for a subset of breast cancer families with an average age at breast cancer diagnosis of less than or equal to 45. Close examination of the late-onset families as defined by Hall et al. (1990) reveals that almost all the late-onset families in fact include a mixture of early- and late-onset breast cancer cases. Therefore, in general, the correlation of age of onset between

relatives in late-onset families is expected to be smaller than such correlation between relatives in early onset families. Since the BRCA1 gene segregates within these early-onset family, one would expect more allele sharing IBD between relatives in these early-onset families. Therefore, sib pairs who share more allele IBD at the disease locus are expected to have higher correlation in their age of onset. Li (2002) applied his method to a breast cancer data set which includes 17 nuclear families. From the collection of 23 families reported by Hall et al. (1990), he identified 17 nuclear families which are fully informative at locus D17S74. The inheritance pattern of these 17 families are therefore uniquely determined. Using his methods, he obtained a Lod score of 1.76 with corresponding  $p = 0.002$ , indicating moderate linkage between breast cancer and locus D17S74. Li (2002) showed the estimated conditional hazard ratio curves  $\theta(t_l, t_m)$  for sib pairs who share 0 and 1 IBD at the marker locus by assuming a Weibull baseline hazard function. It can be seen that the conditional hazard ratio is a function of both age of onset of one sib ( $t_m$ ) and the current age of the other sib ( $t_l$ ), indicating that risk of breast cancer to the relative of a breast cancer proband is dependent upon the age at which the proband is affected. In addition, for given  $t_l$  and  $t_m$ , the conditional hazard is higher for sib pairs who share 1 allele IBD than for sib pairs who share 0 allele IBD. For example, the increased relative risk ratio for an individual at age 50 is 4.17, given that her sister gets breast cancer at age 40 compared to her sister being breast cancer free at age 40 if they share 1 allele IBD at the locus D17S74. However, the increased relative risk ratio is 3.07 if they share 0 allele IBD, and is 5.68 if they share 2 alleles IBD.



# Chapter 14

---

## *Identifiability of Bivariate Frailty Models*

14.1 Introduction .....	273
14.2 Identifiability of Bivariate Frailty Models .....	275
14.3 Identifiability of Correlated Frailty Models .....	276
14.4 Non-Identifiability of Frailty Models without Observed Covariates .....	277
14.4.1 Bivariate Frailty Models with Infinite Mean .....	277
14.4.2 Bivariate Frailty Models with Finite Mean .....	278
14.5 Discussion .....	279

---

### 14.1 Introduction

One of the goals of frailty modeling is to study the properties of the frailty distribution and the baseline hazard function in models describing real survival data. Such analysis is possible if the distribution of frailty as well as the age trajectory of the underlying hazard can be reconstructed (identified) from survival data.

In a univariate frailty model the observed survival time  $T$  and unobserved frailty variable  $Z$  are related by the proportional hazards assumption:

$$h(t|Z) = Z h_0(t). \quad (14.1)$$

In applications, data on observed covariates  $\mathbf{Y}$  are often available together with survival information. In this case the frailty model (14.1) may be extended to include the effects of the observed covariates:

$$h(t|Z, \mathbf{Y}) = Z r(\mathbf{Y}) h_0(t), \quad (14.2)$$

where  $r(\mathbf{Y})$  is an unknown risk function. The bivariate model was derived by extending (14.1) to the 2-dimensional case under the assumption of conditional independence of  $T_1, T_2$  given the shared frailty  $Z$ , resulting in the following bivariate survival function:

$$S(t_1, t_2|Z) = e^{-ZH(t_1)} e^{-ZH(t_2)}. \quad (14.3)$$

In many applications it is not reasonable to assume that the frailties corresponding to the two subjects are identical. Bivariate frailty models allow us to

overcome this limitation. These models extend the idea of individual frailty to the bivariate case and include shared frailty models (14.3) as particular cases. In bivariate frailty models related individuals have different but dependent frailties  $Z_1$  and  $Z_2$  yielding the conditional survival function

$$S(t_1, t_2 | Z_1, Z_2) = e^{-Z_1 H(t_1)} e^{-Z_2 H(t_2)} \quad (14.4)$$

under the assumption of conditional independence of  $T_1, T_2$  given  $Z_1, Z_2$ .

Including the effects of observed covariates  $\mathbf{Y}_1, \mathbf{Y}_2$  measured for the two subjects in the bivariate frailty model (14.4) yields

$$S(t_1, t_2 | Z_1, Z_2, \mathbf{Y}_1, \mathbf{Y}_2) = e^{-Z_1 r(\mathbf{Y}_1) H(t_1)} e^{-Z_2 r(\mathbf{Y}_2) H(t_2)}. \quad (14.5)$$

Note that (14.5) is a straightforward extension of (14.2) to the bivariate case under the conditional independence assumption. Identifiability of model (14.5) was proved by Honoraé (1993) under the assumption of finite mean of  $Z_i, i = 1, 2$ .

The bivariate frailty distribution in models (14.4) and (14.5) may be constructed using independent additive components with one common component for both frailties (i.e.  $Z_i = X_0 + X_i, i = 1, 2$ ), introducing an additional parameter characterizing the correlation between the frailties (Yashin and Iachine 1994), hence the name “correlated frailty models”. To allow for different variances of  $Z_1$  and  $Z_2$ , Yashin and Iachine (1994) introduced an additional scalar parameter  $\alpha > 0$  resulting in the conditional survival function of the form

$$S(t_1, t_2 | X_0, X_1, X_2) = e^{-(X_0 + X_1) H(t_1)} e^{-(\alpha X_0 + X_2) H(t_2)}. \quad (14.6)$$

Under the assumption that  $X_i, i = 0, 1, 2$  are gamma-distributed, Yashin and Iachine (1999) proved the identifiability of the correlated frailty model (14.6) without observed covariates. Their proof was based on the relation between the moments of the bivariate frailty distribution and the partial derivatives of the respective marginal bivariate survival function in  $t_1 = t_2 = 0$ .

This chapter shows that under weak regularity conditions correlated frailty models of the form (14.6) are identifiable (for a discussion of the conditions please refer to Iachine and Yashin (1998)). The result actually holds for a slightly broader class of models, the so-called generalized shared frailty models.

However, it turns out that general multivariate frailty models (14.4) with finite mean do not have the identifiability property when no covariates are observed. To illustrate the non-identifiability problem we present an example of two finite mean bivariate frailty models which produce the same bivariate survival time distribution: a shared frailty model and a model with non-shared, but dependent frailties.

## 14.2 Identifiability of Bivariate Frailty Models

We consider identifiability of bivariate frailty models in the sense of McLachlan and Basford (1988) where the bivariate Laplace transform  $L$  and the cumulative hazard function  $H$  are treated as parameters of the survival model.

**Definition:** A class  $\mathcal{C}$  of bivariate frailty models  $(L, H)$ , where  $L$  is the bivariate frailty Laplace transform and  $H$  is the cumulative hazard function  $H$ , is called *identifiable* if for all  $(L_1, H_1), (L_{II}, H_{II}) \in \mathcal{C}$  the condition

$$L_1(H_I(t_1), H_I(t_2)) = L_{II}(H_{II}(t_1), H_{II}(t_2)), \quad \forall t_1, t_2 \geq 0$$

implies that

$$L_I(s_1, s_2) = L_{II}(s_1, s_2), \quad H_I(t) = H_{II}(t)$$

holds for all  $s_1, s_2 \geq 0, t \geq 0$ .

There is a class of bivariate frailty models for which the identifiability property holds - it is the class of *models based on additive independent components* (Iachine and Yashin, 1998), i.e. let

$$Z_i = X_0 + X_i, \quad i = 1, 2$$

where  $X_0, X_1, X_2$  are independent. Then

$$S(t_1, t_2) = L_0(H(t_1) + H(t_2))L_1(H(t_1))L_2(H(t_2))$$

with  $L_i(s)$  being the Laplace transform of  $X_i$ . The proof is based on the results presented below.

Let  $Z$  be a non-negative random variable and let  $H(t)$  be a cumulative hazard function (i.e. it is non-decreasing and  $H(0) = 0$ ). The following assumptions will be used in the formulation and in the proof of the main identifiability result:

- (i)  $P(Z = 0) = 0$
- (ii)  $H(1) = 1$
- (iii)  $H(t)$  is continuous and  $\lim_{t \rightarrow \infty} H(t) = \infty$
- (iv)  $\inf\{z | P(Z \leq z) > 0\} = 0$ .

The main identifiability result is formulated in the following theorem:

**Theorem 14.1** *Let  $Z$  be some non-negative random variables with Laplace transform  $L(s)$  for which conditions (i) and (ii) hold. Let conditions (ii) and (iii) hold for cumulative hazards  $H_i(t), i = 1, 2$ . Let  $\alpha > 0$  be a real number and let  $\tilde{H}_i, i = 1, 2$  be some cumulative hazards. Define*

$$S(t_1, t_2) = L(H_1(t_1) + \alpha H_2(t_2))e^{-\tilde{H}_1(t_1) - \tilde{H}_2(t_2)}.$$

*If  $S(t_1, t_2)$  is known then the functions  $L(s), H_i(t), \tilde{H}_i(t), i = 1, 2$  and the parameter  $\alpha$  can be identified.*

As it can be seen from the formulation of the above theorem, the identifiability property holds for a slightly broader class of frailty models which may be called *generalized shared frailty models*. The family of models based on the additive independent components is included in this class. Note that the condition of finite mean required for identifiability of a frailty model with observed covariates (Elbers and Ridder, 1982) is not needed in this case. As independent proof of identifiability of the correlated gamma-frailty model was given by Yashin and Iachine (1999b).

---

### 14.3 Identifiability of Correlated Frailty Models

We consider the identifiability property of the correlated frailty model (14.6) without observed covariates. However, as noted by Honore (1993), the proof may be also applied to correlated frailty models with observed covariates, where the dependence between the conditional hazard and the observed covariates given frailty is arbitrary (i.e. non-proportional, e.g., Iachine, 2002).

While searching for a way to prove identifiability for model (14.6), it turned out to be possible to show this property for a slightly more general class of bivariate random effect models, which may be called generalized shared frailty models (GSF). These models are based on the conditional hazard function of the form

$$h(t|Z) = Z h_0(t) + \bar{h}(t), \quad (14.7)$$

with shared frailty  $Z$  and two hazard functions  $h_0(t)$  and  $\bar{h}(t)$ . As shown in the proof of the Corollary below, the correlated frailty model (14.6) is a particular case of the GSF model. The main identifiability result for model (14.7) is contained in the theorem below.

Before formulating this main result we turn to the assumptions which are used in the identifiability proof (for a discussion of the assumptions please refer to Iachine and Yashin, 1998). Let  $Z$  be a non-negative random variable. The function  $H : [0, +\infty) \rightarrow [0, +\infty)$  is a cumulative hazard function if it is not-decreasing and  $H(0) = 0$ . The following assumptions will be used in the formulation and in the proof of the main identifiability result.

- (i)  $P(Z = 0) = 0$
- (ii)  $H(1) = 1$
- (iii)  $H(t)$  is continuous and  $\lim_{t \rightarrow \infty} H(t) = \infty$
- (iv)  $\inf\{z | P(Z \leq z) > 0\} = 0$ .

The main result of this chapter is formulated in the following statement:

**Theorem 14.2** *Let  $Z$  be a non-negative random variable with Laplace transform  $L(s)$  for which conditions (i) and (iv) hold. Let conditions (ii) and (iii)*

hold for cumulative hazards  $H_i(t)$ ,  $i = 1, 2$ . Let  $\alpha > 0$  be a real number and let  $\bar{H}_i(t)$ ,  $i = 1, 2$  be some cumulative hazards. Define:

$$S(t_1, t_2) = L(H_1(t_1) + \alpha H_2(t_2))e^{-\bar{H}_1(t_1)}e^{-\bar{H}_2(t_2)}.$$

If  $S(t_1, t_2)$  is known then the functions  $L(s)$ ,  $H_i(t)$ ,  $\bar{H}_i(t)$ ,  $i = 1, 2$  and the parameter  $\alpha$  can be identified.

**Corollary 14.1** Let  $Z_1 = X_0 + X_1$  and  $Z_2 = \alpha X_0 + X_2$  where  $X_0, X_1, X_2$  are independent non-negative random variables with Laplace transforms  $L_0(s)$ ,  $L_1(s)$ ,  $L_2(s)$ , respectively and where  $\alpha > 0$  be a real number. Assume that  $X_0$  satisfies condition (iv). Let  $H_i(t) = \int_0^t h_{0i}(u)du$ ,  $i = 1, 2$  be cumulative hazards satisfying conditions (ii) and (iii). Define:

$$\begin{aligned} S(t_1, t_2) &= E[e^{-Z_1 H_1(t_1) - Z_2 H_2(t_2)}] \\ &= L_0(H_1(t_1) + \alpha H_2(t_2))L_1(H_1(t_1))L_2(H_2(t_2)). \end{aligned} \quad (14.8)$$

If  $S(t_1, t_2)$  is known then the functions  $L_0(s)$ ,  $H_i(t)$ ,  $L_i(s)$ ,  $i = 1, 2$ , and the parameter  $\alpha$  can be identified.

## 14.4 Non-Identifiability of Frailty Models without Observed Covariates

### 14.4.1 Bivariate Frailty Models with Infinite Mean

It has been known for quite some time, that bivariate frailty models (14.4) with  $EZ = \infty$  are not identifiable, as illustrated by the following extension of the non-identifiability example of Hougaard (2000, p. 357), which is based on the positive stable distribution family,  $PS(\alpha)$ ,  $\alpha \in (0, 1)$ . The idea of this example is to construct two different bivariate frailty distributions (and respective cumulative hazard functions) that yield the same marginal survival distribution.

Consider a model with frailties  $Z_1, Z_2$  and cumulative baseline hazard  $H(t)$ :

$$S(t_1, t_2) = E[e^{-Z_1 H(t_1) - Z_2 H(t_2)}].$$

Define new frailty variables  $\bar{Z}_i = Z_i^{1/\alpha} X_i$ ,  $i = 1, 2$ , and cumulative hazard functions  $\bar{H}(t) = H^{1/\alpha}(t)$  where  $X_1, X_2$  are iid  $PS(\alpha)$ ,  $\alpha \in (0, 1)$ . The marginal bivariate survival function for this frailty model is

$$\begin{aligned} \bar{S}(t_1, t_2) &= E[e^{-\bar{Z}_1 \bar{H}(t_1) - \bar{Z}_2 \bar{H}(t_2)}] \\ &= E[E(e^{-X_1 Z_1^{1/\alpha} H^{1/\alpha}(t_1)} | Z_1, Z_1) E(e^{-X_2 Z_2^{1/\alpha} H^{1/\alpha}(t_2)} | Z_1, Z_2)] \\ &= E[e^{-Z_1 H(t_1) - Z_2 H(t_2)}] = S(t_1, t_2). \end{aligned}$$

This implies non-identifiability, since the frailty model with  $\bar{Z}_1, \bar{Z}_2, \bar{H}(t)$  yields the same marginal bivariate survival function as the frailty model based on  $Z_1, Z_2, H(t)$ . Note that  $E\bar{Z}_i = \infty$  and  $L_{\bar{Z}_1, \bar{Z}_2}(s_1, s_2) = L(s_1^\alpha, s_2^\alpha)$ . To make it clear, that the distributions  $(Z_1, Z_2)$  and  $(\bar{Z}_1, \bar{Z}_2)$  are substantially different (i.e., not rescaled versions of each other) assume that  $(Z_1, Z_2)$  corresponds to a shared frailty model, i.e.  $Z_1 = Z_2 = Z$ . However, the frailties  $\bar{Z}_1, \bar{Z}_2$  are clearly not identical in this case, which completes the proof.

This non-identifiability is a reflection of the observation made by Ridder (1990), that without assumptions on  $Z_i, i = 1, 2$  (such as  $EZ_i < \infty$ ) the frailty distribution is only identifiable up to a power transformation (see also Honoré, 1993).

#### 14.4.2 Bivariate Frailty Models with Finite Mean

The “intuitive” reasons for non-identifiability of the bivariate frailty model may be illustrated by rewriting the frailty model in a sort of an GAFT-form (Ridder, 1990):

$$T_i = H^{-1} \left( \frac{V_i}{Z_i} \right), \quad i = 1, 2 \quad (14.9)$$

where  $(V_1, V_2)$  are iid  $\text{Exp}(1)$  random variable which are independent of the frailties  $Z_1, Z_2$ , contributing to the independent variation of  $T_1, T_2$  when the frailties  $Z_1, Z_2$  are fixed. That is, these variables play a role of “random noise” or “error terms”.

On the other hand, when the frailties  $Z_1, Z_2$  are not identical, they contain independent noise terms of their own, best illustrated by the correlated frailty model (14.6), where the terms  $X_1, X_2$  are exactly of this nature. The same observation was made by Hougaard (2000), p. 364, who noted that “having two random effects for the same source of variation, of course, implies that it will be difficult or impossible to separate the random effects”.

But why is the correlated frailty model identifiable? It seems that in order to create non-identifiability one must be able to “transfer” some of the independent variation between  $V_i$  and  $X_i$ . However, this is impossible to do within the class of correlated frailty models with finite mean because of the non-linearity of the relationship  $V_i/(X_0 + X_i)$ , unless  $X_0 = 0$ , which corresponds to the case of independence between  $T_1$  and  $T_2$ , in which case the model is, of course, non-identifiable (see also similar arguments given by Honoré (1993) for the shared frailty model). In the non-identifiability example above this independent variation is actually transferred by means of the introduced positive stable components, destroying the finite frailty mean property.

It turns out that it is still possible to construct a counter example for the identifiability property of bivariate frailty model with finite mean. However, in this case the independent variation must be introduced in an alternative way to ensure the finite mean property, namely through double conditioning or compound frailty construction, as shown below.

Let  $X$  be an absolutely divisible random variable with Laplace transform  $L(s)$  and let  $V$  be a positive random variable. Define the frailty variables  $Z_1, Z_2$  by assuming  $Z_1 \perp Z_2 | V$  and that conditional on  $V$ ,  $Z_i$  has Laplace transform  $L_{Z|V}(s|v) = L^v(s)$ . If the distribution of  $X$  is not degenerated, this yields a non-shared frailty model with survival function:

$$\begin{aligned} S(t_1, t_2) &= E[e^{-Z_1 H(t_1) - Z_2 H(t_2)}] \\ &= E[E(e^{-Z_1 H(t_1)} | V) E(e^{-Z_2 H(t_2)} | V)] \\ &= E[L^V(H(t_1)) L^V(H(t_2))] = E[e^{-V(\bar{H}(t_1) + \bar{H}(t_2))}] \end{aligned}$$

where  $\bar{H}(t) = -\ln L(H(t))$ . This implies non-identifiability, since the last expression also corresponds to a shared frailty model with frailty  $V$  and cumulative baseline hazard  $\bar{H}(t)$ . Note that  $Z_i$  and  $V$  may both have finite means (e.g., when  $V$  is a two-point discrete and  $X$ -gamma distributed).

In this example the residual random variation is represented by the conditional distribution of the frailties  $Z_1, Z_2$  given  $V$ , which are assumed to be independent under this condition.

---

## 14.5 Discussion

Identifiability of shared and correlated frailty models without observed covariates has been used to gain useful insight into the observed dependence patterns in bivariate survival data e.g., Vaupel et al. (1992), Yashin and Iachine (1995a), Iachine et al. (1998), Giard et al. (2002), Ripatti et al. (2003), and Wienke et al. (2003). The property of identifiability possessed by the correlated frailty models under fairly general conditions provides evidence for the fact that not all survival distributions may be satisfactorily described by the shared frailty models, despite their persisting popularity. The correlated frailty construction provides not only a more flexible way of modeling the bivariate dependence structure, but also a tool for evaluating the shared frailty assumption, because the two models are nested.

On the other hand, the non-identifiability property illustrates that there are limits for what can be learned about the dependence of frailty from the bivariate survival data alone. It is clear that each frailty derived bivariate survival model gives rise to an entire equivalence class of bivariate frailty distributions with finite mean (and respective baseline hazard functions) exactly reproducing the original survival distribution.

In genetic studies of twins and relatives (Neale and Cardon, 1992) the focus on the analysis of the second moments (i.e., variances and covariances) of the quantity of interest (e.g., frailty, Yashin and Iachine, 1995a). In this regard, it is interesting to note that the results of Yashin and Iachine (1995b) show that each of these “equivalent” frailty models, however different, still produce the

same covariance between the corresponding random hazards ( $Z_1 h_0(t), Z_2 h(t)$ ), because this covariance may be directly obtained from the bivariate survival function (as the mixed second partial derivative of its logarithm evaluated at 0).

The non-identifiability example shows, however, that frailty variance is not constant within the equivalence class and may be inflated by the compounding procedure. Further analysis is needed to obtain the bounds for the frailty variance within an equivalence class and to quantify its implications for genetic studies of latent traits (such as frailty).

Another conclusion deals with bivariate frailty models with observed covariates entering the model in the proportional hazard (PH) form (14.5), which were shown to be identifiable when  $EZ < \infty$  (Honoré 1993). In the view of the non-identifiability property when no PH-covariates are observed, it should be noted that information about the marginal frailty distribution (e.g., the frailty variance) is obtained from by identifiability of the univariate survival model (Elbers and Ridder, 1982). But the magnitude of unobserved heterogeneity in this model is determined by the degree of deviation of the univariate survival model from the PH-assumption (i.e., the Cox regression model) and not by the presence or absence of dependence between the related survival times.

This limitation is closely related to the so-called confounding problem (Hougaard, 1985; Iachine, 2002) present in finite mean frailty models with PH-covariates (14.5). In these models, the marginal distribution of frailty plays a dual role: on one hand it describes the deviations of the marginal univariate survival model from the PH-assumption (e.g., zero frailty variance implies a PH-model), on the other hand it describes the strength of dependence between the related survival times (e.g., zero frailty variance implies no dependence).

Consequently, the estimates of marginal frailty parameters (e.g., variance) must be interpreted with care, especially in situations where it might be suspected that the chosen frailty model does not provide a satisfactory explanation for the data. Fortunately, the confounding situation above is relatively easy to diagnose by comparing the results of univariate frailty model analysis with those obtained in the analysis of multivariate data. In this case, deviations between the marginal frailty parameters obtained in the two analyses may be interpreted as evidence of confounding.

---

## Appendix

The Following Table A.1 shows the data on 137 bone marrow transplant patients. The variables represented in the data set are as follows:

g--Disease Group  
    1-ALL  
    2-AML Low Risk  
    3-AML High Risk  
T1 -- Time To Death Or On Study Time  
T2 -- Disease Free Survival Time (Time To Relapse, Death Or End Of Study)  
I1 -- Death Indicator  
    1-Dead 0-Alive  
I2 -- Relapse Indicator  
    1-Relapsed, 0-Disease Free  
I3--Disease Free Survival Indicator  
    (1-Dead Or Relapsed, 0-Alive Disease Free)  
TA--Time To Acute Graft-Versus-Host Disease  
IA--Acute GVHD Indicator  
    (1-Developed Acute GVHD 0-Never Developed Acute GVHD)  
TC--Time To Chronic Graft-Versus-Host Disease  
IC--Chronic GVHD Indicator  
    1-Developed Chronic GVHD 0-Never Developed Chronic GVHD  
TP--Time To Return of Platelets to Normal Levels  
IP--Platelet Recovery Indicator  
    1-Platelets Returned To Normal, 0-Platelets Never Returned to Normal  
Z1--Patient Age In Years  
Z2--Donor Age In Years  
Z3--Patient Sex  
    1-Male, 0-Female  
Z4--Donor Sex  
    1-Male, 0-Female  
Z5--Patient CMV Status  
    1-CMV Positive, 0-CMV Negative  
Z6--Donor CMV Status  
    1-CMV Positive, 0-CMV Negative  
Z7--Waiting Time to Transplant In Days  
Z8--FAB  
    1-FAB Grade 4 Or 5 and AML, 0-Otherwise  
Z9--Hospital  
    1-The Ohio State University, 2-Alferd, 3-St. Vincent, 4-Hahnemann  
Z10--MTX Used as a Graft-Versus-Host- Prophylactic 1-Yes 0-No

g	T1	T2	I1	I2	I3	TA	IA	TC	IC	TP	IP	Z1	Z2	Z3	Z4	Z5	Z6	Z7	Z8	Z9	Z10
1	2081	2081	0	0	0	67	1	121	1	13	1	26	33	1	0	1	1	98	0	1	0
1	1602	1602	0	0	0	1602	0	139	1	18	1	21	37	1	1	0	0	1720	0	1	0
1	1496	1496	0	0	0	1496	0	307	1	12	1	26	35	1	1	1	0	127	0	1	0
1	1462	1462	0	0	0	70	1	95	1	13	1	17	21	0	1	0	0	168	0	1	0
1	1433	1433	0	0	0	1433	0	236	1	12	1	32	36	1	1	1	1	93	0	1	0
1	1377	1377	0	0	0	1377	0	123	1	12	1	22	31	1	1	1	1	2187	0	1	0
1	1330	1330	0	0	0	1330	0	96	1	17	1	20	17	1	0	1	1	1006	0	1	0
1	996	996	0	0	0	72	1	121	1	12	1	22	24	1	0	0	0	1319	0	1	0
1	226	226	0	0	0	226	0	226	0	10	1	18	21	0	1	0	0	208	0	1	0
1	1199	1199	0	0	0	1199	0	91	1	29	1	24	40	1	1	0	1	174	0	3	1
1	1111	1111	0	0	0	1111	0	1111	0	22	1	19	28	1	1	0	1	236	0	3	1
1	530	530	0	0	0	38	1	84	1	34	1	17	28	1	1	0	0	151	0	3	1
1	1182	1182	0	0	0	1182	0	112	1	22	1	24	23	0	0	0	1	203	0	2	1
1	1167	1167	0	0	0	39	1	487	1	1167	0	27	22	0	1	1	1	191	0	2	1
1	418	418	1	0	1	418	0	220	1	21	1	18	14	1	1	0	0	110	0	1	0
1	417	383	1	1	1	417	0	417	0	16	1	15	20	1	1	0	0	824	0	1	0
1	276	276	1	0	1	276	0	81	1	21	1	18	5	0	0	0	0	146	0	1	0
1	156	104	1	1	1	28	1	156	0	20	1	20	33	1	1	0	1	85	0	1	0
1	781	609	1	1	1	781	0	781	0	26	1	27	27	1	0	1	1	187	0	1	0
1	172	172	1	0	1	22	1	172	0	37	1	40	37	0	0	0	1	129	0	1	0
1	487	487	1	0	1	487	0	76	1	22	1	22	20	1	1	0	0	128	0	1	0
1	716	662	1	1	1	716	0	716	0	17	1	28	32	1	1	0	0	84	0	1	0
1	194	194	1	0	1	194	0	94	1	25	1	26	32	0	1	0	0	329	0	1	0
1	371	230	1	1	1	371	0	184	1	9	1	39	31	0	1	0	1	147	0	1	0
1	526	526	1	0	1	526	0	121	1	11	1	15	20	1	1	0	0	943	0	1	0
1	122	122	1	0	1	88	1	122	0	13	1	20	26	1	0	0	1	2616	0	1	0
1	1279	129	1	1	1	1279	0	1279	0	22	1	17	20	0	0	0	0	937	0	3	1
1	110	74	1	1	1	110	0	110	0	49	1	28	25	1	0	1	0	303	0	3	1
1	243	122	1	1	1	243	0	243	0	23	1	37	38	0	1	1	1	170	0	3	1
1	86	86	1	0	1	86	0	86	0	86	0	17	26	1	0	1	0	239	0	3	1
1	466	466	1	0	1	466	0	119	1	100	1	15	18	1	1	0	0	508	0	3	1
1	262	192	1	1	1	10	1	84	1	59	1	29	32	1	1	1	0	74	0	3	1
1	162	109	1	1	1	162	0	162	0	40	1	36	43	1	1	1	0	393	0	2	1
1	262	55	1	1	1	262	0	262	0	24	1	23	16	0	1	1	1	331	0	2	1
1	1	1	1	0	1	1	0	1	0	1	0	42	48	1	1	0	0	196	0	2	1
1	107	107	1	0	1	107	0	107	0	107	0	30	19	1	1	1	1	178	0	2	1
1	269	110	1	1	1	269	0	120	1	27	1	29	20	0	1	1	1	361	0	2	1
1	350	332	1	0	1	350	0	350	0	33	1	22	20	1	0	0	0	834	0	2	1
2	2569	2569	0	0	0	2569	0	2569	0	21	1	19	13	1	1	1	0	270	1	1	0
2	2506	2506	0	0	0	2506	0	2506	0	17	1	31	34	1	1	0	0	60	0	1	0
2	2409	2409	0	0	0	2409	0	2409	0	16	1	35	31	1	1	1	1	120	0	1	0
2	2218	2218	0	0	0	2218	0	2218	0	11	1	16	16	1	1	1	0	60	1	1	0
2	1857	1857	0	0	0	1857	0	260	1	15	1	29	35	0	0	1	0	90	0	1	0
2	1829	1829	0	0	0	1829	0	1829	0	19	1	19	18	1	1	1	0	210	0	1	0
2	1562	1562	0	0	0	1562	0	1562	0	18	1	26	30	1	1	1	1	90	0	1	0
2	1470	1470	0	0	0	1470	0	180	1	14	1	27	34	1	1	0	1	240	0	1	0
2	1363	1363	0	0	0	1363	0	200	1	12	1	13	24	1	1	1	0	90	0	1	0
2	1030	1030	0	0	0	1030	0	210	1	14	1	25	29	0	0	0	0	210	0	1	0
2	860	860	0	0	0	860	0	860	0	15	1	25	31	0	1	0	1	180	0	1	0
2	1258	1258	0	0	0	1258	0	120	1	66	1	30	16	0	1	1	0	180	0	2	1

2	2246	2246	0	0	0	52	1	380	1	15	1	45	39	0	0	0	0	105	0	4	0
2	1870	1870	0	0	0	1870	0	230	1	16	1	33	30	0	0	1	1	225	0	4	0
2	1799	1799	0	0	0	1799	0	140	1	12	1	32	23	1	0	0	0	120	0	4	0
2	1709	1709	0	0	0	20	1	348	1	19	1	23	28	0	1	1	0	90	1	4	0
2	1674	1674	0	0	0	1674	0	1674	0	24	1	37	34	1	1	0	0	60	1	4	0
2	1568	1568	0	0	0	1568	0	1568	0	14	1	15	19	1	0	0	0	90	0	4	0
2	1527	1527	0	0	0	1527	0	1527	0	13	1	22	12	0	1	0	1	450	1	4	0
2	1324	1324	0	0	0	25	1	1324	0	15	1	46	31	1	1	1	1	75	0	4	0
2	957	957	0	0	0	957	0	957	0	69	1	18	17	1	1	0	0	90	0	4	0
2	932	932	0	0	0	29	1	932	0	7	1	27	30	0	0	0	0	60	1	4	0
2	847	847	0	0	0	847	0	847	0	16	1	28	29	1	1	0	0	75	0	4	0
2	848	848	0	0	0	848	0	155	1	16	1	23	26	1	1	0	0	180	0	4	0
2	1850	1850	0	0	0	1850	0	1850	0	9	1	37	36	0	0	0	1	180	0	3	1
2	1843	1843	0	0	0	1843	0	1843	0	19	1	34	32	0	0	1	1	270	0	3	1
2	1535	1535	0	0	0	1535	0	1535	0	21	1	35	32	0	1	0	0	180	1	3	1
2	1447	1447	0	0	0	1447	0	220	1	24	1	33	28	0	1	1	1	150	0	3	1
2	1384	1384	0	0	0	1384	0	200	1	19	1	21	18	0	0	0	0	120	0	3	1
2	414	414	1	0	1	414	0	414	0	27	1	21	15	1	1	0	1	120	1	1	0
2	2204	2204	1	0	1	2204	0	2204	0	12	1	25	19	0	0	0	1	60	0	1	0
2	1063	1063	1	0	1	1063	0	240	1	16	1	50	38	1	0	1	0	270	1	1	0
2	481	481	1	0	1	30	1	120	1	24	1	35	36	1	0	1	1	90	1	1	0
2	105	105	1	0	1	21	1	105	0	15	1	37	34	1	0	1	1	120	0	1	0
2	641	641	1	0	1	641	0	641	0	11	1	26	24	1	1	0	0	90	0	1	0
2	390	390	1	0	1	390	0	390	0	11	1	50	48	1	1	0	0	120	0	1	0
2	288	288	1	0	1	18	1	100	1	288	0	45	43	1	1	1	1	90	0	1	0
2	522	421	1	1	1	25	1	140	1	20	1	28	30	1	1	0	1	90	1	1	0
2	79	79	1	0	1	16	1	79	0	79	0	43	43	0	0	0	0	90	0	1	0
2	1156	748	1	1	1	1156	0	180	1	18	1	14	19	1	0	0	0	60	0	1	0
2	583	486	1	1	1	583	0	583	0	11	1	17	14	0	1	0	0	120	0	1	0
2	48	48	1	0	1	48	0	48	0	14	1	32	33	0	1	1	0	150	1	1	0
2	431	272	1	1	1	431	0	431	0	12	1	30	23	0	1	1	0	120	1	1	0
2	1074	1074	1	0	1	1074	0	120	1	19	1	30	32	1	1	1	0	150	1	1	0
2	393	381	1	1	1	393	0	100	1	16	1	33	28	0	0	0	0	120	1	1	0
2	10	10	1	0	1	10	0	10	0	10	0	34	54	1	0	1	1	240	0	2	1
2	53	53	1	0	1	53	0	53	0	53	0	33	41	0	1	1	1	180	0	2	1
2	80	80	1	0	1	10	1	80	0	80	0	30	35	0	0	1	1	150	0	2	1
2	35	35	1	0	1	35	0	35	0	35	0	23	25	0	1	1	1	150	0	2	1
2	1499	248	0	1	1	1499	0	1499	0	9	1	35	18	1	1	0	1	30	0	4	0
2	704	704	1	0	1	36	1	155	1	18	1	29	21	0	1	1	0	105	0	4	0
2	653	211	1	1	1	653	0	653	0	23	1	23	16	1	0	0	0	90	1	4	0
2	222	219	1	1	1	222	0	123	1	52	1	28	30	1	1	1	1	120	1	3	1
2	1356	606	0	1	1	1356	0	1356	0	14	1	33	22	1	1	1	0	210	1	3	1
3	2640	2640	0	0	0	2640	0	2640	0	22	1	18	23	1	1	0	0	750	0	1	0
3	2430	2430	0	0	0	2430	0	2430	0	14	1	29	26	1	1	0	1	24	0	1	0
3	2252	2252	0	0	0	2252	0	150	1	17	1	35	31	1	0	0	0	120	0	1	0
3	2140	2140	0	0	0	2140	0	220	1	18	1	27	17	1	1	1	1	210	0	1	0
3	2133	2133	0	0	0	2133	0	250	1	17	1	36	39	0	1	0	0	240	0	1	0
3	1238	1238	0	0	0	1238	0	250	1	18	1	24	28	1	0	1	1	240	0	1	0
3	1631	1631	0	0	0	1631	0	150	1	40	1	27	21	1	0	1	0	690	1	2	1
3	2024	2024	0	0	0	2024	0	180	1	16	1	35	41	0	1	0	0	105	1	4	0
3	1345	1345	0	0	0	32	1	360	1	14	1	50	36	1	1	1	1	120	0	4	0
3	1136	1136	0	0	0	1136	0	140	1	15	1	47	27	1	0	1	0	900	0	3	1
3	845	845	0	0	0	845	0	845	0	20	1	40	39	0	0	1	1	210	1	3	1
3	491	422	1	1	1	491	0	180	1	491	0	22	21	0	0	0	0	210	1	1	0

3	162	162	1	0	1	162	0	162	0	13	1	22	23	1	0	0	1	300	0	1	0
3	1298	84	1	1	1	1298	0	1298	0	1298	0	8	2	0	0	1	0	105	1	1	0
3	121	100	1	1	1	28	1	121	0	65	1	39	48	1	1	1	1	210	1	1	0
3	2	2	1	0	1	2	0	2	0	2	0	20	19	1	1	0	0	75	1	1	0
3	62	47	1	1	1	62	0	62	0	11	1	27	25	1	1	0	0	90	1	1	0
3	265	242	1	1	1	265	0	210	1	14	1	32	32	1	0	0	0	180	1	1	0
3	547	456	1	1	1	547	0	130	1	24	1	31	28	1	0	1	1	630	1	1	0
3	341	268	1	1	1	21	1	100	1	17	1	20	23	0	1	1	1	180	1	1	0
3	318	318	1	0	1	318	0	140	1	12	1	35	40	0	1	1	1	300	0	1	0
3	195	32	1	1	1	195	0	195	0	16	1	36	39	1	1	0	0	90	1	1	0
3	469	467	1	1	1	469	0	90	1	20	1	35	33	0	0	1	0	120	0	1	0
3	93	47	1	1	1	93	0	93	0	28	1	7	2	1	1	0	0	135	1	1	0
3	515	390	1	1	1	515	0	515	0	31	1	23	25	1	1	1	0	210	1	1	0
3	183	183	1	0	1	183	0	130	1	21	1	11	7	0	1	0	0	120	1	1	0
3	105	105	1	0	1	105	0	105	0	105	0	14	18	1	0	0	0	150	1	1	0
3	128	115	1	1	1	128	0	128	0	12	1	37	35	0	0	1	1	270	0	1	0
3	164	164	1	0	1	164	0	164	0	164	0	19	32	0	0	0	1	285	1	1	0
3	129	93	1	1	1	129	0	129	0	51	1	37	34	0	1	1	0	240	1	1	0
3	122	120	1	1	1	122	0	122	0	12	1	25	29	0	1	1	1	510	1	1	0
3	80	80	1	0	1	21	1	80	0	0	1	35	28	1	0	0	0	780	1	1	0
3	677	677	1	0	1	677	0	150	1	8	1	15	14	1	1	1	0	150	1	1	0
3	73	64	1	1	1	73	0	73	0	38	1	45	42	0	1	1	0	180	1	2	1
3	168	168	1	0	1	168	0	200	1	48	1	32	43	0	1	1	1	150	1	2	1
3	74	74	1	0	1	29	1	74	0	24	1	41	29	0	1	1	1	750	0	2	1
3	16	16	1	0	1	16	0	16	0	16	0	27	36	0	0	1	0	180	0	4	0
3	248	157	1	1	1	248	0	100	1	52	1	33	39	0	0	1	1	180	1	4	0
3	732	625	1	1	1	732	0	732	0	18	1	39	43	0	1	1	1	150	1	4	0
3	105	48	1	1	1	105	0	105	0	30	1	17	14	0	1	0	0	210	1	4	0
3	392	273	1	1	1	392	0	122	1	24	1	43	50	1	1	1	0	240	0	3	1
3	63	63	1	0	1	38	1	63	0	16	1	44	37	1	1	0	0	360	1	3	1
3	97	76	1	1	1	97	0	97	0	97	0	48	56	1	1	1	1	330	0	3	1
3	153	113	1	1	1	153	0	153	0	59	1	31	25	0	1	1	1	240	0	3	1
3	363	363	1	0	1	363	0	363	0	19	1	52	48	1	1	1	0	180	0	3	1

Table A.1: Bone marrow transplantation data

The following Table A.2 shows the following variable in columns 1:11 respectively for diabetic retinopathy data.

1:Subject id(SI)

2:laser type(LT): 1=xenon, 2=argon

3:treated eye(TE): 1=right 2=left

4:age at diagnosis of diabetes (Age):

5:type of diabetes(TD): 1=juvenile (age at dx \$<\$ 20), 2=adult

6:Outcome for the treated eye (OT): risk group: 6-12

7:status(ST): 0=censored, 1=blindness

8:follow-up time for treated eye (TT):

9:Outcome for the untreated eye (OU): risk group: 6-12

10:status(SU): 0=censored, 1=blindness

11:follow-up time for untreated eye (TU):

The risk group variable was used to define the 'high risk' samples.

SI	LT	TE	Age	TD	OT	ST	TT	OU	SU	TU
5	2	2	28	2	9	0	46.23	9	0	46.23
14	2	1	12	1	8	0	42.5	6	1	31.3
16	1	1	9	1	11	0	42.27	11	0	42.27
25	2	2	9	1	11	0	20.6	11	0	20.6
29	1	2	13	1	9	0	38.77	10	1	0.3
46	1	1	12	1	9	0	65.23	9	1	54.27
49	2	1	8	1	8	0	63.5	6	1	10.8
56	1	1	12	1	8	0	23.17	9	0	23.17
61	2	1	16	1	9	0	1.47	10	0	1.47
71	2	1	21	2	9	0	58.07	9	1	13.83
100	2	2	23	2	9	1	46.43	9	0	48.53
112	2	1	44	2	11	0	44.4	12	1	7.9
120	1	2	47	2	11	0	39.57	6	0	39.57
127	1	1	48	2	6	1	30.83	10	1	38.57
133	2	1	26	2	10	0	66.27	9	1	14.1
150	2	2	10	1	9	1	20.17	10	1	6.9
167	2	2	23	2	12	0	58.43	9	1	41.4
176	2	1	5	1	9	0	58.2	9	0	58.2
185	2	1	46	2	6	0	57.43	8	0	57.43
190	2	1	5	1	11	0	56.03	12	0	56.03
202	2	2	13	1	9	0	67.53	9	0	67.53
214	1	1	45	2	9	0	61.4	12	1	0.6
220	2	1	11	1	10	1	10.27	10	1	1.63
243	1	2	1	1	12	0	66.2	11	0	66.2
255	1	2	10	1	12	1	5.67	9	1	13.83
264	2	1	12	1	9	0	58.83	11	1	29.97
266	2	1	36	2	9	0	60.27	11	1	26.37
284	2	2	53	2	10	1	5.77	12	1	1.33
295	1	2	10	1	11	1	5.9	11	1	35.53
300	1	2	25	2	10	1	25.63	9	1	21.9
302	2	2	14	1	9	1	33.9	9	1	14.8
315	2	2	16	1	10	1	1.73	8	1	6.2
324	1	2	38	2	9	0	46.9	8	1	22
328	2	2	14	1	9	0	31.13	9	0	31.13
335	2	2	10	1	9	1	30.2	11	1	22
342	2	2	17	1	11	0	70.9	12	0	70.9
349	1	1	44	2	11	1	25.8	8	1	13.87
357	2	2	21	2	9	1	5.73	9	1	48.3
368	1	2	19	1	9	0	53.43	9	0	53.43
385	2	1	13	1	10	1	1.9	12	0	51.1
396	2	2	40	2	10	1	9.9	9	1	9.9

405	2	1	9	1	8	0	34.2	12	0	34.2
409	2	2	48	2	10	0	46.73	12	1	2.67
419	1	1	42	2	10	0	18.73	10	1	13.83
429	1	1	24	2	11	0	32.03	10	1	4.27
433	2	2	55	2	10	0	69.87	10	1	13.9
445	2	2	17	1	11	0	66.8	9	0	66.8
454	2	1	5	1	9	0	64.73	8	0	64.73
468	2	1	6	1	10	1	1.7	6	1	1.7
480	1	1	19	1	9	1	1.77	12	1	43.03
485	1	1	12	1	8	0	29.03	8	0	29.03
491	1	2	45	2	9	0	56.57	9	0	56.57
503	1	2	27	2	9	1	8.3	10	1	8.3
515	2	1	43	2	8	0	21.57	9	1	18.43
522	1	2	4	1	9	0	31.57	8	0	31.57
538	2	2	45	2	10	0	31.63	9	1	31.63
547	1	1	32	2	11	0	39.77	10	0	39.77
550	1	2	3	1	10	1	18.7	11	1	6.53
554	2	1	14	1	8	0	18.9	9	0	18.9
557	2	1	13	1	10	0	56.8	8	1	22.23
561	1	1	15	1	10	0	55.6	9	1	14
568	2	2	10	1	9	1	42.17	10	1	42.17
572	1	2	6	1	9	0	10.7	9	1	5.33
576	1	2	17	1	9	0	66.33	10	1	59.8
581	2	2	37	2	12	0	52.33	11	1	5.83
606	1	1	18	1	9	0	58.17	12	1	2.17
610	1	2	13	1	9	1	14.3	8	1	48.43
615	1	2	14	1	12	0	25.83	10	0	25.83
618	1	1	12	1	10	0	45.4	9	0	45.4
624	2	1	9	1	6	0	47.6	9	0	47.6
631	2	2	11	1	12	1	13.33	10	1	9.6
636	2	2	10	1	11	0	42.1	11	0	42.1
645	1	1	5	1	9	0	39.93	6	0	39.93
653	1	2	15	1	9	1	14.27	12	1	7.6
662	2	2	7	1	12	1	34.57	12	1	1.8
664	2	2	2	1	9	0	65.8	12	1	4.3
683	1	2	22	2	8	1	4.1	6	1	12.2
687	2	2	5	1	11	0	60.93	12	0	60.93
701	1	2	4	1	10	0	57.2	9	0	57.2
706	1	1	27	2	8	0	38.07	10	1	12.73
717	1	1	53	2	11	0	54.1	11	1	54.1
722	1	1	10	1	9	0	59.27	12	1	9.4
731	1	2	13	1	12	1	21.57	10	1	9.9
740	2	2	12	1	8	0	54.1	8	0	54.1
749	1	2	24	2	9	0	50.47	11	0	50.47
757	2	2	17	1	9	0	46.17	12	0	46.17
760	1	2	8	1	11	0	46.3	11	0	46.3
766	1	1	58	2	12	0	38.83	9	0	38.83
769	1	2	17	1	9	0	44.6	9	0	44.6
772	2	2	12	1	12	0	43.07	9	0	43.07
778	1	2	25	2	8	1	26.23	8	0	40.03
780	2	1	15	1	9	0	41.6	11	1	18.03
793	1	2	21	2	11	0	38.07	11	0	38.07
800	1	1	20	2	9	0	65.23	9	0	65.23
804	1	1	23	2	10	1	7.07	12	0	66.77
810	1	2	13	1	10	1	13.77	12	1	13.77
815	1	2	45	2	10	0	9.63	12	1	9.63
832	2	1	5	1	12	0	46.23	12	0	46.23
834	2	2	8	1	9	0	45.73	12	1	1.5
838	1	2	30	2	9	1	33.63	9	1	33.63
857	2	2	7	1	8	0	40.17	6	0	40.17

866	2	2	39	2	11	1	63.33	10	1	27.6
887	2	1	26	2	10	1	38.47	10	1	1.63
903	1	1	50	2	6	0	55.23	8	0	55.23
910	1	2	34	2	6	0	52.77	10	1	25.3
920	1	2	10	1	6	0	57.17	8	1	46.2
925	2	1	40	2	10	0	9.87	10	1	1.7
931	1	2	13	1	12	0	57.9	11	0	57.9
936	2	2	7	1	10	0	5.9	12	0	5.9
945	1	2	11	1	12	0	32.2	12	0	32.2
949	1	1	13	1	9	1	10.33	10	1	0.83
952	2	2	9	1	12	1	6.13	9	0	50.9
962	2	2	5	1	9	0	43.67	9	1	25.93
964	2	2	10	1	9	0	38.3	9	0	38.3
971	2	1	23	2	9	0	38.77	10	1	19.4
978	2	2	2	1	12	0	38.07	10	1	21.97
983	1	1	12	1	11	0	38.3	11	0	38.3
987	1	2	7	1	10	1	26.2	9	0	70.03
1002	2	2	13	1	9	0	62.57	11	1	18.03
1017	1	2	50	2	8	1	13.83	9	1	1.57
1029	1	1	20	2	12	0	46.5	12	1	13.37
1034	2	2	15	1	11	1	11.07	9	1	1.97
1037	1	1	30	2	9	0	42.47	9	1	22.2
1042	1	2	32	2	9	0	38.73	9	0	38.73
1069	2	2	39	2	11	0	51.13	11	0	51.13
1074	1	2	4	1	10	1	6.1	11	0	46.5
1098	2	2	3	1	10	1	2.1	11	1	11.3
1102	2	1	10	1	10	1	17.73	9	0	42.3
1112	2	2	6	1	9	0	26.47	10	0	26.47
1117	2	1	15	1	11	0	10.77	11	0	10.77
1126	2	1	33	2	12	0	55.33	10	0	55.33
1135	1	1	15	1	9	0	58.67	9	0	58.67
1145	2	2	44	2	10	1	12.93	9	1	4.97
1148	2	2	48	2	9	0	54.2	12	1	26.47
1167	2	2	4	1	10	0	49.57	9	0	49.57
1184	2	2	46	2	12	1	24.43	11	1	9.87
1191	1	2	25	2	9	0	50.23	9	0	50.23
1205	1	2	12	1	11	1	13.97	12	1	30.4
1213	1	1	12	1	6	0	43.33	10	1	43.33
1228	2	1	26	2	8	0	42.23	8	0	42.23
1247	1	2	11	1	9	0	74.93	9	0	74.93
1250	2	1	36	2	9	0	66.93	9	0	66.93
1253	1	1	12	1	9	0	73.43	9	0	73.43
1267	1	1	50	2	12	0	67.47	11	1	38.57
1281	1	1	44	2	9	0	3.67	10	1	3.67
1287	2	2	8	1	10	1	48.87	9	0	67.03
1293	1	1	14	1	9	0	65.6	9	0	65.6
1296	1	2	18	1	10	0	15.83	10	1	15.83
1309	1	1	56	2	11	0	20.07	8	1	8.83
1312	1	2	9	1	10	0	67.43	9	0	67.43
1317	2	1	15	1	9	0	1.47	6	0	1.47
1321	1	1	5	1	11	0	62.93	9	1	22.13
1333	2	1	1	1	9	1	6.3	11	0	56.97
1347	1	1	1	1	10	0	59.7	10	1	18.93
1361	1	2	14	1	10	1	13.8	9	1	19
1366	1	1	57	2	6	0	55.13	6	0	55.13
1373	2	2	8	1	11	1	13.57	11	1	5.43
1397	1	1	33	2	11	0	42.2	11	0	42.2
1410	1	2	46	2	10	0	38.27	9	0	38.27
1413	2	1	3	1	8	0	7.1	12	1	7.1
1425	2	2	35	2	11	0	63.63	11	1	26.17

1447	2	2	8	1	11	0	59	11	1	24.73
1461	1	1	30	2	9	0	54.37	10	0	54.37
1469	1	1	51	2	8	0	54.6	12	1	10.97
1480	2	2	42	2	9	0	63.87	9	1	21.1
1487	2	1	20	2	9	0	62.37	8	1	43.7
1491	1	2	23	2	11	0	62.8	9	0	62.8
1499	1	1	22	2	11	0	63.33	9	1	14.37
1503	1	2	25	2	9	0	58.53	9	0	58.53
1513	2	2	45	2	12	0	58.07	11	0	58.07
1524	2	2	20	2	10	0	58.5	9	0	58.5
1533	2	2	41	2	9	1	1.5	9	0	14.37
1537	1	1	19	1	12	0	54.73	10	1	38.4
1552	1	1	4	1	12	0	50.63	11	1	2.83
1554	1	2	36	2	10	0	51.1	9	0	51.1
1562	1	1	20	2	9	0	49.93	9	1	6.57
1572	1	1	24	2	9	0	46.27	9	1	46.27
1581	2	1	28	2	9	0	10.6	10	0	10.6
1585	1	1	51	2	8	0	42.77	12	0	42.77
1596	2	1	16	1	9	1	34.37	10	0	42.27
1600	1	1	16	1	10	0	42.07	10	0	42.07
1603	1	2	10	1	9	0	38.77	9	0	38.77
1619	1	2	20	2	9	0	74.97	12	1	61.83
1627	1	2	10	1	10	1	6.57	12	0	66.97
1636	2	1	16	1	6	1	38.87	6	0	68.3
1640	1	2	10	1	11	1	42.43	9	1	46.63
1643	1	1	11	1	9	0	67.07	9	0	67.07
1649	2	1	1	1	10	1	2.7	12	0	2.7
1666	2	2	17	1	6	0	63.8	8	0	63.8
1672	2	2	7	1	9	0	32.63	9	0	32.63
1683	1	1	29	2	10	0	62	8	0	62
1688	1	2	5	1	11	1	13.1	10	0	54.8
1705	1	2	1	1	8	0	8	8	0	8
1717	2	2	22	2	12	0	51.6	11	1	42.33
1727	2	1	33	2	9	0	49.97	10	1	2.9
1746	2	1	3	1	10	0	45.9	10	1	1.43
1749	2	1	32	2	9	0	41.93	9	0	41.93

Table A.2: Diabetic retinopathy data

The following Table A.3 presents the data on Myeloma having the following variables.

Time(T)  
 Status(S)  
 LogBUN(LBUN)  
 HGB  
 Platelet(P)  
 Age  
 LogWBC(LW)  
 Frac(F)  
 LogPBM(LPBM)  
 Protein(Pr)  
 SCalc(SC)

T	S	LBUN	HGB	P	Age	LW	F	LPBM	Pr	SC
1.25	1	2.2175	9.4	1	67	3.6628	1	1.9542	12	10
1.25	1	1.9395	12.0	1	38	3.9868	1	1.9542	20	18
2.00	1	1.5185	9.8	1	81	3.8751	1	2.0000	2	15
2.00	1	1.7482	11.3	0	75	3.8062	1	1.2553	0	12
2.00	1	1.3010	5.1	0	57	3.7243	1	2.0000	3	9
3.00	1	1.5441	6.7	1	46	4.4757	0	1.9345	12	10
5.00	1	2.2355	10.1	1	50	4.9542	1	1.6628	4	9
5.00	1	1.6812	6.5	1	74	3.7324	0	1.7324	5	9
6.00	1	1.3617	9.0	1	77	3.5441	0	1.4624	1	8
6.00	1	2.1139	10.2	0	70	3.5441	1	1.3617	1	8
6.00	1	1.1139	9.7	1	60	3.5185	1	1.3979	0	10
6.00	1	1.4150	10.4	1	67	3.9294	1	1.6902	0	8
7.00	1	1.9777	9.5	1	48	3.3617	1	1.5682	5	10
7.00	1	1.0414	5.1	0	61	3.7324	1	2.0000	1	10
7.00	1	1.1761	11.4	1	53	3.7243	1	1.5185	1	13
9.00	1	1.7243	8.2	1	55	3.7993	1	1.7404	0	12
11.00	1	1.1139	14.0	1	61	3.8808	1	1.2788	0	10
11.00	1	1.2304	12.0	1	43	3.7709	1	1.1761	1	9
11.00	1	1.3010	13.2	1	65	3.7993	1	1.8195	1	10
11.00	1	1.5682	7.5	1	70	3.8865	0	1.6721	0	12
11.00	1	1.0792	9.6	1	51	3.5051	1	1.9031	0	9
13.00	1	0.7782	5.5	0	60	3.5798	1	1.3979	2	10
14.00	1	1.3979	14.6	1	66	3.7243	1	1.2553	2	10
15.00	1	1.6021	10.6	1	70	3.6902	1	1.4314	0	11
16.00	1	1.3424	9.0	1	48	3.9345	1	2.0000	0	10
16.00	1	1.3222	8.8	1	62	3.6990	1	0.6990	17	10
17.00	1	1.2304	10.0	1	53	3.8808	1	1.4472	4	9
17.00	1	1.5911	11.2	1	68	3.4314	0	1.6128	1	10
18.00	1	1.4472	7.5	1	65	3.5682	0	0.9031	7	8
19.00	1	1.0792	14.4	1	51	3.9191	1	2.0000	6	15
19.00	1	1.2553	7.5	0	60	3.7924	1	1.9294	5	9
24.00	1	1.3010	14.6	1	56	4.0899	1	0.4771	0	9
25.00	1	1.0000	12.4	1	67	3.8195	1	1.6435	0	10
26.00	1	1.2304	11.2	1	49	3.6021	1	2.0000	27	11
32.00	1	1.3222	10.6	1	46	3.6990	1	1.6335	1	9

35.00	1	1.1139	7.0	0	48	3.6532	1	1.1761	4	10
37.00	1	1.6021	11.0	1	63	3.9542	0	1.2041	7	9
41.00	1	1.0000	10.2	1	69	3.4771	1	1.4771	6	10
41.00	1	1.1461	5.0	1	70	3.5185	1	1.3424	0	9
51.00	1	1.5682	7.7	0	74	3.4150	1	1.0414	4	13
52.00	1	1.0000	10.1	1	60	3.8573	1	1.6532	4	10
54.00	1	1.2553	9.0	1	49	3.7243	1	1.6990	2	10
58.00	1	1.2041	12.1	1	42	3.6990	1	1.5798	22	10
66.00	1	1.4472	6.6	1	59	3.7853	1	1.8195	0	9
67.00	1	1.3222	12.8	1	52	3.6435	1	1.0414	1	10
88.00	1	1.1761	10.6	1	47	3.5563	0	1.7559	21	9
89.00	1	1.3222	14.0	1	63	3.6532	1	1.6232	1	9
92.00	1	1.4314	11.0	1	58	4.0755	1	1.4150	4	11
4.00	0	1.9542	10.2	1	59	4.0453	0	0.7782	12	10
4.00	0	1.9243	10.0	1	49	3.9590	0	1.6232	0	13
7.00	0	1.1139	12.4	1	48	3.7993	1	1.8573	0	10
7.00	0	1.5315	10.2	1	81	3.5911	0	1.8808	0	11
8.00	0	1.0792	9.9	1	57	3.8325	1	1.6532	0	8
12.00	0	1.1461	11.6	1	46	3.6435	0	1.1461	0	7
11.00	0	1.6128	14.0	1	60	3.7324	1	1.8451	3	9
12.00	0	1.3979	8.8	1	66	3.8388	1	1.3617	0	9
13.00	0	1.6628	4.9	0	71	3.6435	0	1.7924	0	9
16.00	0	1.1461	13.0	1	55	3.8573	0	0.9031	0	9
19.00	0	1.3222	13.0	1	59	3.7709	1	2.0000	1	10
19.00	0	1.3222	10.8	1	69	3.8808	1	1.5185	0	10
28.00	0	1.2304	7.3	1	82	3.7482	1	1.6721	0	9
41.00	0	1.7559	12.8	1	72	3.7243	1	1.4472	1	9
53.00	0	1.1139	12.0	1	66	3.6128	1	2.0000	1	11
57.00	0	1.2553	12.5	1	66	3.9685	0	1.9542	0	11
77.00	0	1.0792	14.0	1	60	3.6812	0	0.9542	0	12

Table A.3: Myeloma data

---

## Bibliography

- [1] Aaberge, R., Kravdel, O. and Wennemo, T. (1989). Unobserved heterogeneity in models of marriage dissolution. Discussion Paper 42, Central Bureau of Statistics, Norway.
- [2] Aalen, O.O. (1987). Two examples of modelling heterogeneity in survival analysis. *Scandinavian Journal of Statistics*, 14, 19-25.
- [3] Aalen, O.O. (1988). Heterogeneity in survival analysis. *Statistics in Medicine*, 7, 1121-37.
- [4] Aalen, O.O. (1992). Modelling heterogeneity in survival analysis by the compound Poisson distribution. *Annals of Applied Probability*, 2, 951-72.
- [5] Aalen, O.O. (1994). Effects of frailty in survival analysis. *Statistical Methods in Medical Research*, 3, 227-43.
- [6] Aalen, O.O. and Gjessing, H.K. (2001). Understanding the shape of the hazard rate: A process point of view. *Statistical Science*, 16, 1-22.
- [7] Aalen, O.O. and Hjort, N.L. (2002). Frailty models that yield proportional hazards. *Statistics and Probability Letters*, 58, 335-42.
- [8] Aalen, O.O. and Tretli, S. (1999). Analyzing incidence of tests cancer by means of a frailty model. *Cancer Causes Control*, 10, 285-92.
- [9] Abramowitz, M. and Stegun, I.A. (1965). *Handbook of Mathematical Functions*. Dover: New York.
- [10] Ahmadi, K.A. and Hanagal, D.D. (2008). Bayesian hypotheses tests for some bivariate exponential distributions. *Model Assisted Statistics and Applications*, 3(3), 269-278.
- [11] Ahmadi, K.A. and Hanagal, D.D. (2009). Bayesian estimation of parameters in bivariate exponential regression models. *Journal of Statistical Theory and Application*, 8(4), 428-47.
- [12] Andersen, E.W. (2000). Two-stage estimation in copula models used in frailty studies. Research Report 00/12, Department of Biostatistics, University of Copenhagen.

- [13] Andersen, P.K., Borgan, O., Gill, R.D. and Keiding, N. (1994). *Statistical Models Based on Counting Processes*. Springer: New York.
- [14] Arjas, E. and Haara, P. (1984). A marked point process approach to censored failure data with complicated covariates. *Scandinavian Journal of Statistics*, 11, 193-209.
- [15] Avalos, B.R., Klein, J.L., Kapoor, N., Tutschka, P.J., Klein, J.P. and Copelan, E.A. (1993). Preparation of bone marrow transplantation in Hodgkin's and non-Hodgkin's lymphoma using Bu/Cy. *Bone Marrow Transplantation*, 13, 133-138.
- [16] Babikar, A. and Cuzick, J. (1994). A simple frailty model for family studies with covariates. *Statistics in Medicine*, 13, 1679-92.
- [17] Batchelor, J.R. and Hackett, M. (1970). HLA Matching in the treatment of burned patients with skin allografts. *Lancet*, 2, 581-83.
- [18] Begun, A.Z., Iachine, I.A. and Yashin, A.I. (2000). Genetic nature of individual frailty: Comparison of two approaches. *Twin Research*, 3(1), 51-57.
- [19] Bertoin, J. (1996). *Levy Processes*. Cambridge University Press, Cambridge.
- [20] Billingsley, P. (1968). *Covergence of Probability Measures*. Wiley: New York.
- [21] Bjarnason, H. and Hougaard, P. (2000). Fisher information for two gamma frailty bivariate Weibull models. *Lifetime Data Analysis*, 6, 59-71.
- [22] Block, H.W. and Basu, A.P. (1974). A continuous bivariate exponential extension. *Journal of the American Statistical Association*, 69, 1031-37.
- [23] Bolstad, W.M. and Manda, S.O. (2001). Investigating child mortality in Malawi using family and community random effects: A Bayesian analysis. *Journal of the American Statistical Association*, 96, 12.
- [24] Boyles, R.A. (1983). On the convergence of the EM algorithm. *Journal of Royal Statistical Society, B*, 45, 47-50.
- [25] Breslow, N.E. (1972). Discussion of Professor Cox's paper. *Journal Royal Statistical Society, B*, 34, 216-17.
- [26] Breslow, N.E. (1974). Covariate analysis of censored survival data. *Biometrics*, 30, 89-100.
- [27] Breslow, N.E. and Clayton, D.G. (1993). Approximate inference in generalized linear models. *Journal of the American Statistical Association*, 88, 9-23.

- [28] Carter, B.S., Beaty, H.B., Steinberg, G.D., Childs, B. and Walsh, P.C. (1992). Mendelian inheritance of familial prostate cancer. *Proceedings of National Academy of Sciences, USA*, 89, 3367-71.
- [29] Chernoff, H. (1954). On the distribution of likelihood ratio. *Annals of Mathematical Statistics*, 25, 573-78.
- [30] Claeskens, G., Nguti, R. and Janssen, P. (2006). One-sided tests in shared frailty models. (Preprint).
- [31] Clarke, C., Collins, R.A., Leavitt, B.R., Andrews, D.F., Hayden, M.R., Lumsden, C.J. and McInnes, R.R. (2000). A one-hit model of cell death in inherited neuronal degenerations. *Nature*, 406, 195-99.
- [32] Clarke, C., Lumsden, C.J. and McInnes, R.R. (2001). Inherited neurodegenerative diseases: The on-hit model of neurodegeneration. *Human Molecular Genetics*, 10, 2269-75.
- [33] Claus, E.B., Risch, N.J. and Thompson, W.D. (1990). Using age of onset to distinguish between subforms of breast cancer. *Annals of Human Genetics*, 54, 169-77.
- [34] Clayton, D.G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65, 141-51.
- [35] Clayton, D.G. (1988). The analysis of event history data: A review of progress and outstanding problems. *Statistics in Medicine*, 7, 819-41.
- [36] Clayton, D.G. and Cuzick, J. (1985). Multivariate generalizations of the proportional hazard model (with discussion). *Journal of Royal Statistical Society, A*, 148, 82-117.
- [37] Clerget-Darpoux, F., Bonaiti-Pellie, C. and Hochez, J. (1986). Effects of misspecifying genetic parameters in Lod score analysis. *Biometrics*, 42, 393-99.
- [38] Commenges, D. (1994). Robust genetic linkage analysis based on a score test of homogeneity: The weighted pairwise correlation statistic. *Genetic Epidemiology*, 11, 189-200.
- [39] Commenges, H. and Gadda, H.J. (1997). Generalized score test of homogeneity based on correlated random effects models. *Journal of the Royal Statistical Society, B*, 59(1), 157-71.
- [40] Congdon, P. (2008). A bivariate frailty model for events with a permanent survivor fraction and non-monotonic hazards; with an application to age at first maternity. *Computational Statistics and Data Analysis*, 52, 4346-56.

- [41] Corbeil, R.R. and Searle, S.R. (1978). Restricted maximum likelihood estimation of variance components in the mixed model. *Technometrics*, 18, 31-38.
- [42] Cox, D.R. (1959). The analysis of exponentially distributed life-times with two types of failures. *Journal of the Royal Statistical Society, B*, 21, 411-21.
- [43] Cox, D.R. (1972). Regression models and life tables (with discussions). *Journal of the Royal Statistical Society, B*, 34, 187-220.
- [44] Cox, D.R. (1975). Partial likelihood. *Biometrika*, 62, 269-76.
- [45] Cox, D.R. and Oakes, D. (1984). *Analysis of Survival Data*. Chapman & Hall: London.
- [46] Daw, E.W., Heath, S.C. and Wijsman, E.M. (1999). Multipoint oligogenetic analysis of age-at-onset data with applications to Alzheimer disease pedigrees. *American Journal of Human Genetics*, 64, 839-51.
- [47] Dawson, D.V., Kaplan, E.B. and Elston, R.C. (1990). Extensions to sib-pair linkage tests applicable to disorders characterized by delayed onset. *Genetic Epidemiology*, 7, 453-66.
- [48] Dempster, A.P., Laird, N.M. and Rubin, D.R. (1977). Maximum likelihood estimation from incomplete data via the EM-algorithm (with discussion). *Journal of the Royal Statistical Society, B*, 39, 1-38.
- [49] Duchateau, L. and Janssen, P. (2008). *The Frailty Model*. Springer: New York.
- [50] Duchateau, L., Janssen, P., Lindsey, P., Legrand, C., Nguti, R. and Sylvester, R. (2002). The shared frailty model and the power for heterogeneity tests in multicenter trials. *Computational Statistics and Data Analysis*, 40, 603-20.
- [51] Duggirala, R., Blangero, J., Almasy, L., Dyer, T.D., Williams, K.L., Leach, R.J., O'Connell, P. and Stern, M.P. (1999). Linkage of type 2 diabetes mellitus and age at onset to a genetic location on chromosome 10q in Mexican Americans. *American Journal of Human Genetics*, 64, 1127-40.
- [52] Efron, B. (1977). The efficiency of Cox's likelihood function for censored data. *Journal of the American Statistical Association*, 72, 557-65.
- [53] Elbers, C and Rider, G. (1982). True and spurious duration dependence: The identifiability of the proportional hazard model. *Review of Economic Studies*, 49, 403-09.

- [54] Falconer, D.S. (1965). The inheritance of liability to certain diseases estimated from the incidence among relatives. *Annals of Human Genetics*, 29, 51-76.
- [55] Fan, J.J., Hsu, L. and Prentice, R.L. (2000). Dependence estimation over a finite bivariate failure region. *Lifetime Data Analysis*, 6, 343-55.
- [56] Feller, W. (1996). *An Introduction to Probability Theory and its Applications*. John Wiley & Sons: New York.
- [57] Finkelstein, M. and Esaulova, V. (2006). *On Asymptotic Failure Rates in Bivariate Frailty Competing Risk Models*. Maxplanck Institute, Rostock, Germany. (Preprint).
- [58] Flanders, W.D. and Khoury, M.J. (1991). Extensions to methods of sib-pair linkage analyses. *Genetic Epidemiology*, 8, 399-408.
- [59] Fleming, T.R. and Harrington, D.P. (1984). Nonparametric estimation of survival distribution in censored data. *Communications in Statistics, Theory and Methods*, 13, 2469-86.
- [60] Flinn, C.J. and Heckman, J.J. (1982). New methods for analyzing individual event histories. *Sociological Methodology*, 99-140.
- [61] Freireich, E.J., Gehan, E., Frei, E., Schroeder, L.R., Wolman, I.J., Anbari, R., Burgert, E.O., Mills, S.D., Pinkel, D., Selawry, O.S., Moon, J.H., Gendel, B.R., Spurr, C.L., Storrs, R., Haurani, F., Hoogstraten, B., and Lee, S. (1963). The effect of 6-Mercaptopurine on the duration of steroid-induced remissions in acute Leukemia: A model for evaluation of other potential useful therapy. *Blood*, 21, 699-716.
- [62] Freund, J.E. (1961). A bivariate extension of exponential distribution. *Journal of the American Statistical Association*, 56, 971-77.
- [63] Garibotti, G., Smith, K.R., Kerber, R.A. and Boucher, K.M. (2006). Longevity and correlated frailty in multigenerational families. *Journal of Gerontology*, 61A(12), 1253-61.
- [64] Genest, C. and Mackay, R.J. (1986). The joy of copulas: Bivariate distributions with given marginals. *American Statistician*, 40, 280-83.
- [65] Genest, C., Ghoudi, K. and Rivest, L.P. (1993). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82, 543-52.
- [66] Giard, N., Lichtenstein, P. and Yashin, A.I. (2002). A multistate model for the genetic analysis of the aging process. *Statistics in Medicine*, 21, 2511-26.

- [67] Gilks, W., Richardson, S. and Spiegelhalter, D. (1996). *Markov Chain Monte Carlo in Practice*. Chapman & Hall: New York.
- [68] Gill, R.D. (1980). Censoring and stochastic integrals. *Mathematical Centre Tracks* 124: Amsterdam, Mathematische Centre.
- [69] Gill, R.D. (1985). Discussion of the paper by D.G. Clayton and J. Cuzick. *Journal of Royal Statistical Society, A*, 148, 108-09.
- [70] Gill, R.D. and Johansen, S. (1990). A survey of product-integration with a view towards application in survival analysis. *Annals of Statistics*, 18, 1501-55.
- [71] Gjessing, H.K., Aalen, O.O. and Hjort, N.L. (2003). Frailty models based on Levy processes. *Advances in Applied Probability*, 35, 532-50.
- [72] Glidden, D.V. (1999). Checking the adequacy of the gamma frailty model for multivariate failure times. *Biometrika*, 86, 381-94.
- [73] Glidden, D.V. (2000). A two-stage estimator of the dependence parameter for the Clayton-Oakes model. *Lifetime Data Analysis*, 6(2), 141-56.
- [74] Good, I.G. and Gaskin, R.A. (1971). Non-parametric roughness penalties for probability densities. *Biometrika*, 58, 255-77.
- [75] Green, P.J. and Silverman, B.W. (1994). *Nonparametric Regression and Generalized Linear Models. A Roughness Penalty Approach*. Chapman & Hall: London.
- [76] Greenwood, M. and Yule, G.U. (1920). An enquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease as repeated accidents. *Journal of Royal Statistical Society*, 83, 255-79.
- [77] Gumbel, E.J. (1958). *Statistics of Extremes*. Columbia University Press: New York.
- [78] Gumbel, E.J. (1960). Bivariate exponential distribution. *Journal of the American Statistical Association*, 55, 698-707.
- [79] Hall, J.M., Lee, M.K., Newman, B., Morrow, J.E., Anderson, L.A., Huey, B. and King, M.C. (1990). Linkage of early-onset familial breast cancer to chromosome 17q21. *Science*, 250, 1684-89.
- [80] Hanagal, D.D. (1992a). Some inference results in bivariate exponential distributions based on censored samples. *Communications in Statistics, Theory and Methods*, 21(5), 1273-95.
- [81] Hanagal, D.D. (1992b). Some inference results in modified Freund's bivariate exponential distribution. *Biometrical Journal*, 34(6), 745-56.

- [82] Hanagal, D.D. (2004). Parametric bivariate regression analysis based on censored samples: A Weibull model. *Economic Quality Control*, 19, 83-90.
- [83] Hanagal, D.D. (2005a). A bivariate Weibull regression model. *Economic Quality Control*, 20, 143-50.
- [84] Hanagal, D.D. (2005b). Weibull extension of a bivariate exponential regression model. *Economic Quality Control*, 20, 247-53.
- [85] Hanagal, D.D. (2005c). A positive stable frailty regression model in bivariate survival data. *Journal of Indian Society for Probability and Statistics*, 9, 35-44.
- [86] Hanagal, D.D. (2005d). Bivariate Weibull regression with gamma frailty for censored data. Proceedings of 1st International Conference on Reliability and Safety Engineering, December 21-23, 2005, held at Indian Institute of Technology, Bhuvaneshwar, India. pp. 13-20.
- [87] Hanagal, D.D. (2006a). Bivariate Weibull regression model based on censored samples. *Statistical Papers*, 47, 137-148.
- [88] Hanagal, D.D. (2006b). A gamma frailty regression model in bivariate survival data. *IAPQR Transactions*, 31, 73-83.
- [89] Hanagal, D.D. (2006c). Weibull extensions of bivariate exponential regression model for a survival data. *Statistical Methods*, 8, 110-122.
- [90] Hanagal, D.D. (2006d). Bivariate Weibull regression model with gamma frailty for the survival data. *Statistical Methods*, (Special issue on National Seminar on Modelling and Analysis of Life Time Data), 37-51.
- [91] Hanagal, D.D. (2006e). Bivariate Weibull regression with positive stable frailty for censored data. Proceedings of 2nd International Conference on Reliability and Safety Engineering, December 18-20, 2006, held at Indian Institute of Technology, Chennai, India. pp. 184-192.
- [92] Hanagal, D.D. (2006f). Weibull extension of bivariate exponential regression model with gamma frailty for survival data. *Economic Quality Control*, 21, 165-174.
- [93] Hanagal, D.D. (2007a). Positive stable frailty regression models in mixture distributions. Proceedings of 3rd International Conference on Reliability and Safety Engineering, December 17-19, 2007, held at Indian Institute of Technology, Udaipur, India. pp. 350-356.
- [94] Hanagal, D.D. (2007b). Gamma frailty regression models in mixture distributions. *Economic Quality Control*, 22(2), 295-302.
- [95] Hanagal, D.D. (2007c). A power variance function frailty regression model in bivariate survival data. *IAPQR Transactions*, 32(2), 117-129.

- [96] Hanagal, D.D. (2008a). Frailty regression models in mixture distributions. *Journal of Statistical Planning and Inference*, 138(8), 2462-68.
- [97] Hanagal, D.D. (2008b). Modelling heterogeneity for bivariate survival data by the lognormal distribution. *Statistics and Probability Letters*, 78(9), 1101-09.
- [98] Hanagal, D.D. (2008c). *Multivariate Reliability Models and Methods*. Wiley's Encyclopedia of Quantitative Risk Assessment, Wiley & Sons: New York, Vol 3, 1144-47.
- [99] Hanagal, D.D. (2009a). Weibull extension of bivariate exponential regression model with different frailty distributions. *Statistical Papers*, 50, 29-49.
- [100] Hanagal, D.D. (2009b). Modeling heterogeneity for bivariate survival data by power variance function distribution. *Journal of Reliability and Statistical Studies*, 2, 14-27.
- [101] Hanagal, D.D. (2010a). Correlated compound Poisoon frailty model for the bivariate survival data. *International Journal of Statistics and Management Systems*, 5, 127-40.
- [102] Hanagal, D.D. (2010b). Modeling heterogeneity for bivariate survival data by compound Poisson distribution. *Model Assisted Statistics and Applications*, 5(1), 01-09.
- [103] Hanagal, D.D. (2010c). Modeling heterogeneity for bivariate survival data by Weibull distribution. *Statistical Papers*, 51, (to appear).
- [104] Hanagal, D.D. (2010d). Modeling heterogeneity for bivariate survival data by the compound Poisson distribution with random scale. *Statistics and Probability Letters*, 80, 1781-90.
- [105] Hanagal, D.D. and Kale, B.K. (1991a). Large sample tests of  $\lambda_3$  in the bivariate exponential distribution. *Statistics and Probability Letters*, 12(4), 311-13.
- [106] Hanagal, D.D. and Kale, B.K. (1991b). Large sample tests of independence in an absolutely continuous bivariate exponential distribution. *Communications in Statistics, Theory and Methods*, 20(4), 1301-13.
- [107] Hanagal, D.D. and Kale, B.K. (1992). Large sample tests for testing symmetry and independence in some bivariate exponential models. *Communications in Statistics, Theory and Methods*, 21(9), 2625-43.
- [108] Hanagal, D.D. and Ahmadi, K.A. (2008a). Bayesian estimation of parameters in some bivariate exponential models. *Advances and Applications in Statistics*, 10(2), 179-93.

- [109] Hanagal, D.D. and Ahmadi, K.A. (2008b). Estimation of parameters by EM algorithm in bivariate exponential distribution based on censored samples. *Economic Quality Control*, 23(2), 257-66.
- [110] Hanagal, D.D. and Ahmadi, K.A. (2009a). Bayesian estimation of the parameters of bivariate exponential distributions. *Communications in Statistics, Simulation and Computation*, 38(7), 1391-1413.
- [111] Hanagal, D.D. and Ahmadi, K.A. (2009b). Bayesian estimation using EM algorithm in bivariate exponential distributions based on censored samples. *Journal of Applied Statistical Science*, 17(2), 181-198.
- [112] Harrington, D. and Fleming, T.R. (1982c). A class of rank test procedures for censored survival data. *Biometrika*, 69, 553-66.
- [113] Harville, D.A. (1977). Maximum likelihood approaches to variance component estimation and related problems. *Journal of the American Statistical Association*, 72, 320-40.
- [114] Hastie, T. and Tibshirani, R. (1986). Generalized additive models. *Statistical Science*, 1, 297-318.
- [115] Hauge, M. (1981). *The Danish twin register: Prospective longitudinal research, an empirical basis for the primary prevention of psychosocial disorders*. Oxford University Press, London, 218-21.
- [116] Haynes, C., Pericak-Vance, M.A. and Dawson, D. (1986). Genetic analysis workshop IV: Analysis of Huntington's disease linkage and age of onset curves. *Genetic Epidemiology*, 1, 109-22.
- [117] Heckman, J. and Singer, B. (1984). The identifiability of the proportional hazard model. *Review of Econometric Studies*, 51(2), 231-41.
- [118] Hedeker, D., Siddiqui, O. and Hu, F.B. (1996). *Random Effects Regression Analysis of Correlated Group Time Survival Data*. School of Public Health and Prevention Research Cancer. University of Illinois: Chicago.
- [119] Henderson, C.R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics*, 31, 423-47.
- [120] Herskind, A.M., McGue, M.M., Iachine, I.A., Holm, N. Sorensen, Y.I.A., Harvald, B. and Vaupel, J.W. (1996). Untagling genetic influences on smoking body mass index and longevity: A multivariate study of 2464 Danish twins followed for 28 years. *Human Genetics*, 98, 467-75.
- [121] Hjort, N.L. (2003). *Topic in Non-Parametric Bayesian Statistics (with discussion)*. Highly Structured Stochastic Systems. Eds. E.J. Green, N.L. Hjort and S. Richardson, Oxford University Press: London, 455-78.

- [122] Holt, J.D. and Prentice, R.L. (1974). Survival analysis in twin studies and matched pair experiments. *Biometrika*, 61(1), 17-30.
- [123] Honore, B.E. (1993). Identification results for duration models with multiple spells. *Review of Economic Studies*, 60, 241-46.
- [124] Hougaard, P. (1984). Life table methods for heterogeneous populations: Distributions describing heterogeneity. *Biometrika*, 71, 75-83.
- [125] Hougaard, P. (1985). Discussion of the paper by D.G. Clayton and J. Cuzick. *Journal of the Royal Statistical Society, A*, 148, 113-14.
- [126] Hougaard, P. (1986a). A class of multivariate failure time distributions. *Biometrika*, 73, 671-78.
- [127] Hougaard, P. (1986b). Survival models for heterogeneous populations derived from stable distributions. *Biometrika*, 73, 387-96.
- [128] Hougaard, P. (1987). Modeling multivariate survival. *Scandinavian Journal of Statistics*, 14, 291-304.
- [129] Hougaard, P. (1991). Modeling heterogeneity in survival data. *Journal of Applied Probability*, 28, 695-701.
- [130] Hougaard, P. (1999). Fundamentals of survival data. *Biometrics*, 55(1), 13-22.
- [131] Hougaard, P. (2000). *Analysis of Multivariate Survival Data*. Springer: New York.
- [132] Hougaard, P., Muglogaard and Borch-Johnsen, K. (1994). Heterogeneity models of disease susceptibility with application to diabetic nephropathy.
- [133] Biometrics, 50(4), 1178-88.
- [134] Huster, W.J., Brookmeyer, R. and Self, S.G. (1989). Modelling paired survival data with covariates. *Biometrics*, 45, 145-56.
- [135] Iachine, I.A. (1995a). Correlated frailty concept in the analysis of bivariate survival data. Bachelor project, Department of Mathematics and Computer Science, Odense University, Denmark.
- [136] Iachine, I.A. (1995b). Parameter estimation in the bivariate correlated frailty model with observed covariates via the EM-algorithm. Working Paper Series: Population Studies of Aging 16, CHS, Odense University, Denmark.
- [137] Iachine, I.A. (2001). The use of twin and family survival data in the population studies of aging: Statistical methods based on multivariate survival data. Ph.D. thesis, Department of Statistics and Demography, University of Southern Denmark, Denmark.

- [138] Iachine, I.A. (2006). Identifiability of bivariate frailty models. Working Paper, Department of Statistics and Demography, University of Southern Denmark, Denmark.
- [139] Iachine, I.A. and Yashin, A.I. (1998). Identifiability of bivariate frailty models based on additive independent components. Research Report 8, Department of Statistics and Demography, Odense University, Denmark.
- [140] Iachine, I.A., Holm, N.V., Harris, J.R., Begun, A.Z., Iachina, M.K., Laitinen, M., Kaprio, J. and Yashin, A.I. (1998). How heritable is individual susceptibility to death? The results of an analysis of survival data on Danish, Swedish and Finnish twins. *Twin Research*, 1(4), 196-205.
- [141] Iachine, I.A. and Yashin, A. I. (1999). Bivariate quadratic hazard model: Results of analysis of Danish twins survival data. Research Report 21, Department of Statistics and Demography, Odense University, Denmark.
- [142] Ibrahim, J.G., Chen, M.H. and Sinha, D. (2001). *Bayesian Survival Analysis*. Springer-Verlag, New York. Inc.
- [143] Inclan, C. and Tiao, G.C. (1994). Use of cumulative sums of squares for retrospective detection of changes of variances. *Journal of the American Statistical Association*, 89, 913-23.
- [144] Kalbfleisch, J.D. (1974). Some efficiency calculations for survival distributions. *Biometrika*, 61, 31-38.
- [145] Kalbfleisch, J.D. and Prentice, R. (2002). *The Statistical Analysis of Failure Time Data*. John Wiley & Sons: New York.
- [146] Kaplan, E.L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of American Statistical Association*, 53, 457-81.
- [147] Kebir, Y. (1991). On hazard rate processes. *Naval Research Logistics*, 38, 865-76.
- [148] Keiding, N., Andersen, P.K. and Klein, J.P (1997). The role of frailty models and accelerated failure time models in describing heterogeneity due to omitted covariates. *Statistics in Medicine*, 16(1-3), 215-24.
- [149] Keyfitz, N. and Littman, G. (1979). Mortality in a heterogeneous population. *Population Studies*, 33, 333-42.
- [150] Kim, H., Sun, D. and Tsutakawa, R. (2002). Lognormal vs. gamma: Extra variations. *Biometrical Journal*, 44, 305-323.
- [151] Klein, J.P. (1992). Semiparametric estimation of random effects using the Cox model based on EM algorithm. *Biometrics*, 48, 795-806.

- [152] Klein, J.P. and Moeschberger, M.L. (2003). *Survival Analysis*. Second edition. Springer: New York.
- [153] Korsgaard, I.R. and Andersen, A.H. (1998). The additive genetic gamma frailty model. *Scandinavian Journal of Statistics*, 25(2), 255-69.
- [154] Kotz, S., Balakrishnan, N. and Johnson, N.L. (2000). *Continuous Multivariate Distributions*. John Wiley & Sons: New York.
- [155] Koziol, J.A. and Green, S.B. (1976). A cramer-von Mises statistic for randomly censored data. *Biometrics*, 63, 465-74.
- [156] Krall, J.M., Uthoff, V.A. and Harley, J.B. (1975). A step-up procedure for selecting variables associated with survival, *Biometrics*, 31, 4957.
- [157] Kruglyak, L., Daly, M.J., Reeve-Daly, M.P. and Lander, E.S. (1996). Parametric and nonparametric linkage analysis: A unified multipoint approach. *American Journal of Human Genetics*, 58, 1347-63.
- [158] Lagakos, S.W. and Schoenfeld, D. (1984). Properties of proportional hazards score tests under misspecified regression models. *Biometrics*, 40, 1037-48.
- [159] Lancaster, T. (1990). *The Economic Analysis of Transition Data*. University Press Cambridge: London.
- [160] Lander, E. and Green, P. (1987). Construction of multilocus genetic maps in humans. *Proceedings of National Academy of Sciences*, USA, 84, 2363-67.
- [161] Lawless, J.F. (1982). *Statistical Models and Methods for Lifetime Data*. John Wiley & Sons: New York.
- [162] Lee, E.W., Wei, L.J. and Amato, D.A. (1992). Cox-type regression analysis for large numbers of small groups of correlated failure time observations. Eds: Klein, J.P and Goel, P.K.; *Survival Analysis: State of Art*, Kluwer Academic Press, Boston: 237-48.
- [163] Lee, Y. and Nelder, J.A. (1996). Hierarchical generalized linear models. *Journal of Royal Statistical Society*, B, 58, 619-78.
- [164] Lee, E.T. and Wang, J.W. (2003). *Statistical Methods for Survival Data Analysis*. John Wiley & Sons: New York.
- [165] Li, H. (1999). The additive genetic gamma frailty model for linkage analysis. *Annals of Human Genetics*, 63, 455-68.
- [166] Li, H. (2002). An additive genetic gamma frailty model for linkage analysis of diseases with variable age of onset using nuclear families. *Lifetime Data Analysis*, 8, 315-34.

- [167] Li, H. and Hsu, L. (2000). Effects of age at onset on the power of the affected sib pair and transmission/disequilibrium tests. *Annals of Human Genetics*, 64, 239-254.
- [168] Li, H. and Huang, J. (1998). Semiparametric linkage analysis using pseudod likelihoods on neighbouring sets. *Annals of Human Genetics*, 62, 323-336.
- [169] Li, H. and Zhong, X. (2002). Multivariate survival models induced by genetic frailties with application to linkage analysis. *Biostatistics*, 3, 57-75.
- [170] Liang, K.Y. (1991). Estimating effects of probands characteristic in familial risk: 1. Adjustments for censoring and correlated ages at onset. *Genetic Epidemiology*, 8, 329-38.
- [171] Lillard, L.A. (1993). Simultaneous equations for hazards: Marriage duration and fertility timing. *Journal of Econometrics*, 56, 189-217.
- [172] Lillard, L.A., Brien, M.J. and Waite, L.J. (1995). Premarital cohabitation and subsequent marital dissolution: A matter of self-selection? *Demography*, 32, 437-57.
- [173] Lillard, L.A. and Panis, C.W.A. (2000). *AML User's Guide and Reference Manual*. EconWare: Los Angeles.
- [174] Lin, D.Y. and Wei, L.J. (1989). The robust inference of the Cox proportional hazard model. *Journal of the American Statistical Association*, 84, 1074-78.
- [175] Lindstrom, D.P. (1996). Economic opportunities in Mexico and return to the United States. *Demography*, 33, 357-74.
- [176] Little, R.J.A. and Rubin, D.B. (1987). *Statistical Analysis with Missing Data*. Wiley: New York.
- [177] Lo, C.C., Amaral, L., Halvin, S., Ivanov, P.C., Penzel, T., Peter, J.H. and Stanley, H.E. (2002). Dynamics of sleep-wake transitions during sleep. *Europhysics Letters*, 57, 625-31.
- [178] Louis, T.A. (1982). Finding the observed informationmatrix when using the EM algorithm. *Journal of Royal Statistical Society, B*, 44, 226-33.
- [179] Lunn, D., Thomas, A., Best, N. and Spiegelhalter, D. (2000). WinBUGS A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*, 10, 325-337.
- [180] Manda, S. and Meyer, R. (2005). Bayesian inference for recurrent events data using time dependent frailty. *Statistics in Medicine*, 24, 1263-74.

- [181] Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemotherapy Reports*, 50, 163-70.
- [182] Mantel, N., Bohidar, N.R. and Ciminera, J.L. (1977). Mantel-Haenzel analysis of litter-matched time-to-response data, with modifications for recovery of interlitter information. *Cancer Research*, 37, 3863-68.
- [183] Marshall, A.W. and Olkin, I. (1967). A multivariate exponential distribution. *Journal of the American Statistical Association*, 62, 30-44.
- [184] McCall, B.P. (1994). Testing the proportional hazard assumptions in the presence of unmeasured heterogeneity: An application to the unemployment durations of displaced workers. *Journal of Applied Econometrics*, B, 44, 226-33.
- [185] McGilchrist, C.A. (1993). REML estimation for survival models with frailty. *Biometrics*, 49, 221-25.
- [186] McGilchrist, C.A. (1994). Estimation in generalized mixed models. *Journal of the Royal Statistical Society*, B, 56, 61-69.
- [187] McGilchrist, C.A. and Aisbett, C.W. (1991). Regression with frailty in survival analysis. *Biometrics*, 47, 461-66.
- [188] McGue, M., Vaupel, J.W., Holm, N. and Harvald, B. (1993). Longevity is moderately heritable in a sample of Danish twins born 1870-1880. *Journal of Gerontology Biological Science*, 48(6), 237-44.
- [189] McLachlan, G.J. and Basford, K.E. (1998). *Mixture Models: Inference and Applications to Clustering*, Vol 84 of Statistics; Eds: Owen, Cornell, Kennedy, Kshirsagar, Schilling. Marcel Dekker: New York.
- [190] Melino, A. and Sueyoshi, G.T. (1990). A simple approach to the identifiability of the proportional hazard model. *Economics Letters*, 33, 63-68.
- [191] Meyer, M.R., Tschanz, J.T., Norton, M.C., Welsh-Bohmer, K.A., Steffens, D.C., Wyse, B.W. and Breitner, J.C.S. (1998). APOE genotype predicts when-not-whether- on is predisposed to develop Alzheimer disease. *Nature Genetics*, 19, 321-322.
- [192] Michael, J.R. and Schucany, W.R. (1986). *Analysis of Data From Censored Samples: Goodness of Fit Techniques*. Edited by D'Agostino, R.B. and Stephens, M.A. Marcel Dekker: New York.
- [193] Moger, T.A., Aalen, O.O., Heimdal, K. and Gjessing, H.K. (2004). Analysis of testicular cancer data by means of a frailty model with familial dependence. *Statistics in Medicine*, 23, 617-32.

- [194] Moger, T.A. and Aalen, O.O. (2005). A distribution for multivariate frailty based on the compound Poisson distribution with random scale. *Lifetime Data Analysis*, 11, 41-59.
- [195] Moger, T.A and Aalen, O.O. (2006). Hierarchical Levy frailty models and a frailty analysis of data on infant mortality in Norwegian siblings. Working paper Series 290, University of Washington.
- [196] Morgan, B.J.T. (1992). *Analysis of Quantal Response Data*. Chapman & Hall: London.
- [197] Morgan, T.M. (1986). Omitting covariates from the proportional hazards model. *Biometrics*, 42, 993-95.
- [198] Morton, L.A. and Kidd, K.K. (1980). The effects of variable age-of-onset and diagnostic criteria on the estimates of linkage: An example using manic-depressive illness and color blindness. *Social Biology*, 27, 1 10.
- [199] Murphy, S.A. (1994). Consistency in a proportional hazards model incorporating a random effect. *Annals of Statistics*, 22, 712-31.
- [200] Murphy, S.A. (1995). Asymptotic theory for the frailty model. *Annals of Statistics*, 23(1), 182-98.
- [201] Murphy, S.A. and Vaart, A.V. (2000). On profile likelihood. *Journal of the American Statistical Association*, 95, 449-85.
- [202] Nahman, N.S., Middendorf, D.F., Bay, W.H., McElligot, R., Powell, S. and Anderson, J. (1992). Modification of percutaneous approach to peritoneal dialysis catheter placement under peritoneoscopic visualization: Clinical results in 78 patients. *Journal of the American Society of Nephrology*, 3, 103-107.
- [203] Naylor, J.C. and Smith, A.F.M. (1982). Applications of a method for the efficient computation of posterior distribution. *Journal of Applied Statistics*, 31, 214-25.
- [204] Neale, M.C. and Cardon, L.R. (1992). *Methodology for Genetic Studies of Twins and Families*. NATO ASCI Series D: Behavioral and Social Sciences, Kluwer Academic Publishers: Netherlands.
- [205] Nguti, R., Claeskens, G. and Janssen, P. (2004). Likelihood ratio tests for a shared frailty model: A non-standard problem. *Technical Report*, 309, Limburgs Central University, Belgium.
- [206] Nelson, W. (1969). Hazard plotting for incomplete failure data. *Journal Quality Technology*, 1, 27-52.

- [207] Nielson, G.G, Gill, R.D., Andersen, P.K. and Sorensen, T.I.A. (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scandinavian Journal of Statistics*, 19(1), 25-43.
- [208] Noh, M., Ha, I.D. and Lee, Y. (2006). Dispersion frailty models and HGLMs. *Statistics in Medicine*, 25, 1341-54.
- [209] Oakes, D. (1977). The asymptotic information in censored survival data. *Biometrika*, 64, 441-48.
- [210] Oakes, D. (1982). A model for association in bivariate survival data. *Journal of the Royal Statistical Society, B*, 44, 414-93.
- [211] Oakes, D. (1986). Semiparametric inference in bivariate survival data. *Biometrika*, 73, 353-61.
- [212] Oakes, D. (1989). Bivariate survival models induced by frailties. *Journal of the American Statistical Association*, 84, 487-93.
- [213] Oakes, D. (1994). Multivariate survival distributions. *Journal Nonparametric Statistics*, 3, 343-354.
- [214] Ohman, P. and Eberly, L. (2001). A shifted frailty distribution for correlated time to event data. Research report. Department of Statistics, University of Florida, Gainesville.
- [215] Orchard, T. and Woodbury, M.A. (1972). A missing information principle: Theory and applications. *Proceedings of 6th Berkley Symposium on Mathematical Statistics and Probability*, Vol 1, 697-715.
- [216] Ott, J. (1974). Estimation of the recombination fraction in human pedigrees: Efficient computation of the likelihood for human linkage studies. *American Journal of Human Genetics*, 26, 588-597.
- [217] Ott, J. (1976). A computer program for linkage analysis of general human pedigrees. *American Journal of Human Genetics*, 28, 528-529.
- [218] Paik, M., Tsai, W. and Ottman, R. (1994). Multivariate survival analysis using piecewise gamma frailty. *Biometrics*, 50, 975-88.
- [219] Panis, C.W.A and Lillard, L.A. (1995). Child mortality in Malaysia: Explaining ethnic differences and the recent decline. *Population Studies*, 49, 463-79.
- [220] Parner, E. (1997). Inference in semi-parametric frailty models. Ph.D. Thesis. University of Aarhus, Denmark.
- [221] Parner, E. (1998). Asymptotic theory for the correlated gamma-frailty model. *The Annals of Statistics*, 26(1), 183-214.

- [222] Petersen, J.H. (1998). An additive frailty model for correlated life times. *Biometrics*, 54, 646-61.
- [223] Peterson, J.H., Anderson, P.K. and Gill, R.D. (1996). Variance components models for survival data. *Statistica Neerlandica*, 50(1), 193-211.
- [224] Peto, R. (1972). Contribution to the discussion of paper by D. R. Cox. *Journal of the Royal Statistical Society, B*, 34, 205-207.
- [225] Peto, R. and Peto, J. (1972). Asymptotically efficient rank invariant test procedures (with discussion). *Journal of the Royal Statistical Society, A*, 135, 185-206.
- [226] Pickles, A. and Crouchley, R. (1994). A simple frailty model for family studies with covariates. *Statistics in Medicine*, 3, 263-78.
- [227] Pickles, A., Crouchley, R., Simonoff, E., Eaves, L., Meyer, J., Rutter, M., Hewit, J. and Silberg, J. (1994). Survival models for developmental genetic data: Age at onset of puberty and antisocial behaviour in twins. *Genetic Epidemiology*, 11, 155-70.
- [228] Pickles, A. and Crouchley, R. (1995). A comparison of frailty models for multivariate survival data. *Statistics in Medicine*, 14, 1447-61.
- [229] Prentice, R.L., Williams, B.J. and Peterson, A.V. (1981). On the regression analysis of multivariate failure time data. *Biometrika*, 68, 373-79.
- [230] Proschan, F. and Sullo, P. (1974). Estimating the Parameters of Bivariate Exponential Distributions in Several Sampling Situations. Reliability and Biometry. Edited by F. Proschan and R.J. Serfling, Philadelphia: SIAM: 423-40.
- [231] Quaas, R.L. (1976). Computing the diagonal elements and inverse of a large numerator relationship matrix. *Biometrics*, 32, 949-53.
- [232] Ridder, G. (1990). The non-parametric identification of generalized accelerated failure time models. *Review of Economic Studies*, 57(2), 167-81.
- [233] Ripatti, S. and Palmgren, J. (2000). Estimation of multivariate frailty models using penalized likelihood. *Biometrics*, 56, 1016-22.
- [234] Ripatti, S., Larsen, K. and Palmgren, J. (2002). Maximum likelihood inference for multivariate frailty models using an automated Monte Carlo EM algorithm. *Lifetime Data Analysis*, 8, 349-60.
- [235] Ripatti, S., Gatz, M., Pedersen, N.L. and Palmgren, J. (2003). Three state frailty model for age at onset of dementia and death in Swedish twins. *Genetic Epidemiology*, 24(2), 139-49.
- [236] Robertson, T., Wright, F.T. and Dykstra, R.L. (1988). *Order-Restricted Statistical Inference*. Wiley: New York.

- [237] Robinson, G.K. (1991). That BLUP is a good thing: The estimation of random effects. *Statistical Science*, 6, 15-51.
- [238] Sastry, N. (1997). A nest frailty model for survival data with an application to the study of child survival in Northeast Brazil. *Journal of the American Statistical Association*, 92, 426-35.
- [239] Savage, I.R. (1956). Contributions to the theory of rank order statistics-the two sample case. *Annals of Mathematical Statistics*, 27, 590-615.
- [240] Schall, R. (1991). Estimation in generalized linear models with random effects. *Biometrika*, 78, 719-27.
- [241] Scheike, T.H., Peterson, J.H. and Martinussen, T. (1999). Retrospective ascertainment of recurrent events: An application to time to pregnancy. *Journal of the American Statistical Association*, 94(447), 713-25.
- [242] Self, S.G. and Liang, K.Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, 82, 605-10.
- [243] Sen, P.K. and Silvapulle, M.J. (2002). An appraisal of some aspects of statistical inference under inequality constraints. *Journal of Statistical Planning and Inference*, 107, 3-43.
- [244] Shih, J.H. (1998). A goodness-of-fit test for association in a bivariate survival model. *Biometrika*, 85, 189-200.
- [245] Shih, J.H. and Louis, T.A. (1995). Inferences on the association parameter in copula models for bivariate data. *Biometrics*, 51(4), 1384-99.
- [246] Shinnar, S., Berg, A.T., O'Dell, C., Newstein, D., Moshe, S.L. and Hauser, W.A. (2000). Predictors of multiple seizures in a cohort of children prospectively followed from the time of their first unprovoked seizure. *Annals of Neurology*, 48, 140-47.
- [247] Shu, Y. and Klein, J.P. (1999). A SAS macro for the positive stable frailty model. *American Statistical Association Proceedings of Statistical Computing Section*, 47-52.
- [248] Silvapulle, M.J. and Silvapulle, P. (1995). A score test against one-sided alternatives. *Journal of the American Statistical Association*, 90, 342-49.
- [249] Singpurwalla, N.D. (1995). Survival in dynamic environments. *Statistical Science*, 10, 86-103.
- [250] Slate, E.H. and Turnbull, B.W. (2000). Statistical models for longitudinal biomarkers of disease onset. *Statistics in Medicine*, 19, 617-37.

- [251] Smith, A.F.M., Skene, A.M., Shaw, J.E.H. and Naylor, J.C. (1987). Progress with numerical and graphical methods for practical Bayesian statistics. *Statistician*, 36, 75-82.
- [252] Sorensen, T.I.A., Nielsen, G.G., Andersen, P.K. and Jeasdale, T.W. (1988). Genetic and environmental influences on premature death in adult adoptees. *New England Journal of Medicine*, 318, 727-32.
- [253] Stram, D.O. and Lee, J.W. (1994). Variance components testing in the longitudinal mixed effects models. *Biometrics*, 50, 1171-77.
- [254] Struthers, C. and Kalbfleisch, J.D. (1986). Misspecified proportional hazards models. *Biometrika*, 73, 363-69.
- [255] Tapia, R.A. and Thompson, J.R. (1978). *Non-Parametric Probability Density Estimation*. John Hopkins University Press: Baltimore, MD.
- [256] Therneau, T.M. and Grambsch, P.M. (2000). *Modeling Survival Data: Extending the Cox Model*. Springer: New York.
- [257] Therneau, T.M., Grambsch, P.M. and Pankratz, V.S. (2000). Penalized survival models and frailty. *Technical Report*, 66, Mayo Foundation.
- [258] Triarhou, L.C. (1998). Rate of neuronal fallout in a transsynaptic cerebellar model. *Brain Research Bulletin*, 47, 219-22.
- [259] Trussel, J. and Richards, T. (1985). Correcting for unobserved heterogeneity in hazard models using Heckman-Singer procedure. *Sociological Methodology*, 242-76.
- [260] Tutz, G. and Kauermann, G. (2003). Generalized linear random effects models with varying coefficients. *Computational Statistics and Data Analysis*, 43, 1328.
- [261] Vaida, F. and Xu, R. (2000). Proportional hazards model with random effects. *Statistics in Medicine*, 19, 3309-24.
- [262] Vaupel, J.W. (1988). Inherited frailty and longevity. *Demography*, 25(2), 139-49.
- [263] Vaupel, J.W., Manton, K.G. and Stallard, E. (1979). The impact of heterogeneity on individual frailty on the dynamic of mortality. *Demography*, 16(3), 439-54.
- [264] Vaupel, J.W. and Yashin, A.I. (1985). Heterogeneity's rules: Some surprising effects of selection on population dynamics. *The American Statistician*, 39, 176-85.

- [265] Vaupel, J.W., Yashin, A.I., Hauge, M., Harvald, B., Holm, N.Y. and Xue, L. (1992). Survival analysis in genetics: Danish twins data applied to gerontological question. Eds: Klein, J.P. and Goel, P.K., *Survival Analysis: State of Art*, Kluwer: Dordrecht.
- [266] Verbeke, G. and Molenberghs, G. (2003). The use of score tests for inference on variance components. *Biometrics*, 59, 254-62.
- [267] Viswanathan, B. and Manatunga, A.K. (2001). Diagnostic plots for assessing the frailty distribution in multivariate survival data. *Lifetime Data Analysis*, 7, 143-55.
- [268] Vu, H.T.V. and Zhou, S. (1997). Generalization of likelihood ratio tests under nonstandard conditions. *The Annals of Statistics*, 25, 897-916.
- [269] Wang, S.T., Klein, J.P. and Moeschberger, M.L. (1995). Semi-parametric estimation of covariate effects using the positive stable frailty model. *Applied Stochastic Models Data Analysis*, 11, 121-33.
- [270] Weeks, D.E. and Lange, K. (1988). The affected-pedigree-member method of linkage analysis. *American Journal of Human Genetics*, 42, 315-326.
- [271] Wienke, A., Arbeev, K.G.A., Locantelli, I. and Yashin, A.I. (2005). A comparison of different bivariate correlated frailty models and estimation strategies. *Mathematical Sciences*, 198(1), 1-13.
- [272] Wienke, A., Christensen, K., Skythe, A. and Yashin, A.I. (2002). Genetic analysis of cause of death in a mixture model with bivariate lifetime data. *Statistical Modelling*, 2, 89-102.
- [273] Wienke, A., Holm N., Skythe, A. and Yashin, A.I. (2001). The heritability of mortality due to heart diseases: A correlated frailty model applied to Dansih twins. *Twin Research*, 4, 266-74.
- [274] Wienke, A., Lichtenstein, P. and Yashin, A.I. (2003). A bivariate frailty model with a cure fraction for modeling familial correlations in diseases. *Biometrics*, 59, 1178-83.
- [275] Wienke, A., Locantelli, I. and Yashin, A.I. (2006). The modeling of a cure fraction in bivariate time-to-event data. *Austrian Journal of Statistics*, 35(1), 67-76.
- [276] Wintrebert, C.M., Putter, H., Zwinderman, A. and Van Houwelingen, H.C. (2004). Centre-effect on survival after bone marrow transplantation: Application of time-dependent frailty models. *Biometrical Journal*, 46, 512-25.
- [277] Wright, S. (1922). Coeficients of in breeding and relationship. *American Naturalist*, 56, 330-38.

- [278] Wu, C.F.J. (1983). On the convergence properties of EM algorithm. *The Annals of Statistics*, 1(1), 95-103.
- [279] Xue, S. and Brookmeyer, R. (1996). Bivariate frailty model for the analysis of multivariate survival time. *Lifetime Data Analysis*, 2, 277-90.
- [280] Xue, X. and Ding, Y. (1999). Assessing heterogeneity and correlation of paired failure times with the bivariate frailty model. *Statistics in Medicine*, 18, 907-18.
- [281] Yashin, A.I. and Iachine, I.A. (1993a). Survival of related individuals: An extension of some fundamental results of heterogeneity analysis. Research Report, Population Studies of Aging, 8, Odense University, Denmark.
- [282] Yashin, A.I. and Iachine, I.A. (1993b). How to reduce standard errors analyzing bivariate survival data. An approach based on correlated frailty model. Research Report, Population Studies of Aging, 9, Odense University, Denmark.
- [283] Yashin, A.I., Vaupel, J.W. and Iachine, I.A. (1993a). Correlated individual frailty: An advantageous approach to survival analysis of bivariate data. Working paper Series: Population Studies of Aging 7, CHS, Odense University, Denmark.
- [284] Yashin, A.I., Vaupel, J.W. and Iachine, I.A. (1993b). Duality of aging: The equivalence of mortality models based on radically different concepts. *Mechanisms of Aging and Development*, 74, 1-14.
- [285] Yashin, A.I. and Iachine, I.A. (1994). Environment determines 50% of variability in individual frailty: Results from Dansih twin study, Research Report, Population Studies of Aging, 10, Odense University, Denmark.
- [286] Yashin, A.I. and Iachine, I.A. (1995a). Genetic analysis of durations: Correlated frailty model applied to survival of Dansih twins. *Genetic Epidemiology*, 12, 529-38.
- [287] Yashin, A.I. and Iachine, I.A. (1995b). How long can humans live? Lower bound for biological limit of human longevity calculated from Danish twin data using correlated frailty model. *Mechanisms of Aging and Development*, 80, 147-69.
- [288] Yashin, A.I. and Iachin, I.A. (1995c). Survival of related individuals: An extension of some fundamental results of heterogeneity analysis. *Mathematical Population Studies*, 5(4), 321-39.
- [289] Yashin, A.I., Vaupel, J.W. and Iachine, I.A. (1995). Correlated individual frailty: An advantages approach to survival analysis of bivariate data. *Mathematical Population Studies*, 5, 145-59.

- [290] Yashin, A.I. and Iachine, I.A. (1996). Random effect models of bivariate survival: Quadratic hazard as a new alternative. Eds: G. Kristensen. Transactions of Symposium in Avendt Statistik, 22-27. January 1996, Okonomisk Institute, Odense University.
- [291] Yashin, A.I., Manton, K. and Iachine, I.A. (1996). Genetic and environmental factors in duration studies. *Journal of Epidemiology and Biostatistics*, 1(2), 115-20.
- [292] Yashin, A.I. and Iachine, I.A. (1997). How frailty models can be used in the analysis of mortality and longevity limits. *Demography*, 34, 31-48.
- [293] Yashin, A.I. and Manton, K.G. (1997). Effects of unobserved and partially observed covariates processes on system failure: A review of models and estimation strategies. *Statistical Science*, 12, 20-34.
- [294] Yashin, A.I. and Iachine, I.A. (1999). What difference does the dependence between duration make? Insights for population studies of aging. *Lifetime Data Analysis*, 5, 5-22.
- [295] Yashin, A.I., Iachine, I.A. and Harris, J.R. (1999). Half of variation in susceptibility is genetic: Findings from Swedish twin survival data. *Behavior Genetics*, 29(1), 11-19.
- [296] Yashin, A.I., Iachine, I.A. and Begun, A.S. (2000). Mortality modeling: A review. *Mathematical Population Studies*, 8(4), 305-32.
- [297] Yashin, A.I., Iachine, I.A., Begun, A.Z. and Vaupel, J.W. (2001). Hidden frailty: Myths and reality. Research Report 34, Department of Statistics and Demography, Odense University, Denmark.
- [298] Yates, F. (1993). The analysis of replicated experiments when the field results are incomplete. *Empire Journal of Experimental Agriculture*, 1, 129-42.
- [299] Yau, K.K.W. and McGilchrist, C.A. (1998). ML and REML estimation in survival analysis with time dependent correlated frailty. *Statistics in Medicine*, 17, 1201-13.
- [300] Zahl, P.H. (1997). Frailty modelling for the excess hazard. *Statistics in Medicine*, 16, 1573-85.
- [301] Zelterman, D. (1992). A statistical distribution with an unbounded hazard function and its application to a theory from demography. *Biometrics*, 48, 807-18.
- [302] Zhong, X. and Li, H. (2002). An additive genetic gamma frailty model for two-locus linkage analysis using subship age of onset data. *Statistical Applications in Genetics and Molecular Biology*, 1, 1-24.