

Texts in Statistical Science

Multivariate Survival Analysis and Competing Risks

Martin Crowder



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

Contents

Preface	xxi
---------------	-----

Part I Univariate Survival Analysis

1. Scenario	3
1.1 Survival Data	3
1.1.1 Waiting Times	3
1.1.2 Discrete Time	4
1.1.3 Censoring	4
1.2 Some Small Data Sets	5
1.2.1 Strengths of Cords	5
1.2.2 Cancer Survival	6
1.2.3 Catheter Infection	6
1.3 Inspecting the Data with R	7
1.4 Fitting Models with R	9
1.5 Simulating Data with R	9
2. Survival Distributions	11
2.1 Continuous Lifetimes	11
2.2 Some Continuous Survival Distributions	12
2.2.1 The Exponential Distribution	12
2.2.2 The Weibull Distribution	13
2.2.3 The Pareto Distribution	13
2.2.4 Other Distributions	14
2.2.5 The Shape of Hazard	14
2.3 Discrete Lifetimes	15
2.4 Some Discrete Survival Distributions	16
2.4.1 The Geometric Distribution	16
2.4.2 The Negative Binomial Distribution	17
2.5 Mixed Discrete-Continuous Survival Distributions	17
2.5.1 From Discrete to Continuous	18
2.5.2 Riemann–Stieltjes Integrals	19
2.6 Reliability Topics	20
2.6.1 Reliability of Systems	20
2.6.2 k-out-of-n Systems	21
2.6.3 Survival Aspect	21

2.6.4	Degradation Processes	21
2.6.5	Stress and Strength	22
2.7	Exercises	22
2.7.1	Survival Distributions	22
2.7.2	Reliability of Systems	23
2.7.3	Degradation Processes	24
2.7.4	Stress and Strength	24
2.8	Hints and Solutions	24
2.8.1	Survival Distributions	24
2.8.2	Reliability of Systems	26
2.8.3	Degradation Processes	26
2.8.4	Stress and Strength	26
3.	Continuous Time-Parametric Inference	29
3.1	Parametric Inference: Frequentist and Bayesian	29
3.1.1	Frequentist Approach	29
3.1.2	Bayesian Approach	30
3.1.3	Proceed with Caution	31
3.2	Random Samples	32
3.2.1	Type-I Censoring	33
3.2.2	Type-II Censoring	33
3.2.3	Left Truncation	34
3.2.4	Probabilities of Observation versus Censoring	34
3.2.5	Weibull Lifetimes	35
3.2.6	Strengths of Cords	36
3.2.7	Survival of Breast Cancer Patients	38
3.3	Regression Models	39
3.3.1	Business Start-Ups	40
3.3.2	Proportional Hazards (PH)	41
3.3.3	Accelerated Life (AL)	42
3.3.4	Proportional Odds (PO)	42
3.3.5	Mean Residual Life (MRL)	42
3.3.6	Catheter Infection	43
3.4	Goodness of Fit	44
3.4.1	Enhanced Models	45
3.4.2	Uniform Residuals	45
3.4.3	Cox-Snell Residuals	46
3.4.4	Right-Censored Times	46
3.4.5	Other Residuals	46
3.4.6	Tests on Residuals	47
3.5	Frailty and Random Effects	47
3.5.1	Frailty	47
3.5.2	Recovering the Frailty Distribution	48
3.5.3	Discrete Random Effects and Frailty	49
3.5.4	Accommodating Zero Frailty	51

3.6	Time-Dependent Covariates	51
3.7	Exercises	53
3.7.1	Regression Models	53
3.7.2	Residuals	53
3.7.3	Discrete Frailty	54
3.8	Hints and Solutions	55
3.8.1	Regression Models	55
3.8.2	Residuals	55
4.	Continuous Time: Non- and Semi-Parametric Methods	57
4.1	Random Samples	57
4.1.1	The Kaplan–Meier Estimator	57
4.1.2	Strengths of Cords	59
4.1.3	The Integrated and Cumulative Hazard Functions	60
4.1.4	Interval-Censored Data	60
4.2	Explanatory Variables	62
4.2.1	Cox’s Proportional Hazards Model	62
4.2.2	Cox’s Partial Likelihood	62
4.2.3	Inference	63
4.2.4	Computation	64
4.2.5	Catheter Infection	64
4.3	Some Further Aspects	65
4.3.1	Stratification	65
4.3.2	Tied Lifetimes	66
4.3.3	The Baseline Survivor Function	66
4.3.4	The Log-Rank Test	67
4.3.5	Schoenfeld Residuals	68
4.3.6	Time-Dependent Covariates	68
4.3.7	Interval-Censored Data	69
4.4	Task Completion Times	69
4.5	Accelerated Life Models	72
4.6	Exercises	75
4.6.1	Random Samples	75
4.6.2	Partial Likelihood	75
4.6.3	Applications	75
4.6.4	Accelerated Life Models	76
4.7	Hints and Solutions	76
4.7.1	Random Samples	76
4.7.2	Accelerated Life Models	76
5.	Discrete Time	77
5.1	Random Samples: Parametric Methods	77
5.1.1	Geometric Lifetimes	77
5.1.2	Career Promotions	78
5.1.3	Probabilities of Observation versus Censoring	80

5.2	Random Samples: Non- and Semi-Parametric Estimation	81
5.2.1	Career Promotions	82
5.2.2	Large-Sample Theory	83
5.3	Explanatory Variables	84
5.3.1	Likelihood Function	84
5.3.2	Geometric Waiting Times	85
5.3.3	The Driving Test	85
5.3.4	Proportional Hazards	87
5.3.5	Proportional Odds	87
5.3.6	The Driving Test	89
5.3.7	The Baseline Odds	89
5.4	Interval-Censored Data	90
5.4.1	Cancer Survival Data	90
5.5	Frailty and Random Effects	92
5.5.1	Geometric Distribution	92
5.5.2	Random Effects	93
5.5.3	Beta-Geometric Distribution	93
5.5.4	Cycles to Pregnancy	94
5.5.5	The Driving Test	95
5.6	Exercises	95
5.6.1	Random Samples	95
5.6.2	Explanatory Variables	96
5.6.3	Gamma and Beta Distributions	96
5.7	Hints and Solutions	97
5.7.1	Random Samples	97

Part II Multivariate Survival Analysis

6.	Multivariate Data and Distributions	101
6.1	Some Small Data Sets	101
6.1.1	Repeated Response Times	101
6.1.2	Paired Response Times	102
6.1.3	Lengths and Strengths of Fibres	102
6.1.4	Household Energy Usage	102
6.2	Multivariate Survival Distributions	105
6.2.1	Joint and Marginal Distributions	105
6.2.2	Conditional Distributions	105
6.2.3	Dependence and Association	105
6.2.4	Hazard Functions and Failure Rates	106
6.2.5	Gumbel's Bivariate Exponential	107
6.3	Exercises	108
6.3.1	Joint and Marginal Distributions	108
6.3.2	Dependence and Association	108
6.4	Hints and Solutions	109

6.4.1	Joint and Marginal Distributions	109
6.4.2	Dependence and Association	109
7.	Some Models and Methods	111
7.1	The Multivariate Log-Normal Distribution	111
7.2	Applications	112
7.2.1	Household Energy Usage	112
7.2.2	Repeated Response Times	113
7.3	Bivariate Exponential	114
7.3.1	Discrete-Time Version	114
7.4	Bivariate Exponential	115
7.4.1	Discrete-Time Version	116
7.5	Some Other Bivariate Distributions	116
7.5.1	Block and Basu (1974)	116
7.5.2	Lawrance and Lewis (1983)	116
7.5.3	Arnold and Brockett (1983)	117
7.5.4	Cowan (1987)	117
7.5.5	Yet More Distributions	117
7.6	Non- and Semi-Parametric Methods	118
7.7	Exercises	120
8.	Frailty, Random Effects, and Copulas	121
8.1	Frailty: Construction	121
8.2	Some Frailty-Generated Distributions	122
8.2.1	Multivariate Burr	122
8.2.2	Multivariate Weibull	123
8.2.3	Distribution 3	124
8.2.4	Multivariate Beta-Geometric	125
8.2.5	Multivariate Gamma-Poisson	126
8.2.6	Marshall-Olkin Families	126
8.3	Applications	127
8.3.1	Paired Response Times	127
8.3.2	Household Energy Usage	129
8.3.3	Cycles to Pregnancy	130
8.4	Copulas: Structure	131
8.5	Further Details	133
8.6	Applications	135
8.6.1	Clayton Copula	135
8.7	Exercises	136
8.7.1	Frailty-Generated Distributions	136
8.7.2	Applications	137
8.7.3	Copulas	137
8.8	Hints and Solutions	138
8.8.1	Copulas	138

9. Repeated Measures	139
9.1 Pure Frailty Models: Applications	139
9.1.1 Lengths and Strengths of Fibres	140
9.1.2 Visual Acuity	141
9.2 Models with Serial Correlation: Application	143
9.2.1 Repeated Response Times	144
9.3 Matched Pairs	145
9.4 Discrete Time: Applications	146
9.4.1 Proportional Odds	146
9.4.2 Beta-Geometric Model	147
9.4.3 Bird Recapture	147
9.4.4 Antenatal Knowledge	149
9.5 Milestones: Applications	152
9.5.1 Educational Development	152
9.5.2 Pill Dissolution Rates	153
9.5.3 Timber Slip	153
9.5.4 Loan Default	154
9.6 Exercises	156
9.6.1 Bird Recapture Data	156
9.6.2 Some Background for the Bivariate Beta Distribution	156
9.6.3 Antenatal Data	157
9.6.4 Binomial Waiting Times	158
9.6.5 Milestones Data	158
10. Recurrent Events	161
10.1 Some Recurrence Data	161
10.2 The Event Rate	163
10.3 Basic Recurrence Processes	165
10.3.1 Poisson Processes	165
10.3.2 Renewal Processes	166
10.3.3 Recurrence of Medical Condition	166
10.3.4 Simulation	167
10.4 More Elaborate Models	168
10.4.1 Poisson Process	168
10.4.2 Intensity Functions	168
10.5 Other Fields of Application	169
10.5.1 Repair and Warranty Data	169
10.5.2 Sports Data	169
10.5.3 Institution Data	170
10.5.4 Alternating Periods	170
10.6 Event Counts	171
10.6.1 Continuous Time: Poisson Process	171
10.6.2 Discrete Time	172
10.6.3 Multivariate Negative Binomial	172

10.7	Quasi-Life Tables.....	174
10.7.1	Estimation.....	174
10.8	Exercises.....	176
10.9	Hints and Solutions.....	177
11.	Multi-State Processes.....	179
11.1	Markov Chain Models.....	179
11.1.1	Discrete Time.....	179
11.1.2	Continuous Time.....	180
11.1.3	Hidden Markov Chains.....	181
11.1.4	Estimation.....	182
11.2	The Wiener Process.....	183
11.3	Wear and Tear and Lack of Care.....	184
11.4	Cumulative Models.....	185
11.4.1	Compound Poisson Process.....	186
11.4.2	Compound Birth Process.....	188
11.4.3	Gamma Process.....	188
11.4.4	Customer Lifetime Value.....	190
11.5	Some Other Models and Applications.....	190
11.5.1	Empirical Equation Models.....	190
11.5.2	Models for the Stress Process.....	192
11.5.3	Stress-Strength Models.....	193
11.5.4	Other Models and Applications.....	193
11.6	Exercises.....	194
11.6.1	Markov Chains.....	194
11.6.2	Wiener Process.....	195
11.6.3	Cumulative Damage Models.....	197
11.6.4	Compound Poisson Process.....	197
11.6.5	Compound Birth Process.....	197
11.6.6	Gamma Process.....	198
11.7	Hints and Solutions.....	198
11.7.1	Markov Chains.....	198
11.7.2	Wiener Process.....	198
11.7.3	Cumulative Damage Models.....	199
11.7.4	Compound Poisson Process.....	199
11.7.5	Compound Birth Process.....	199

Part III Competing Risks

12.	Continuous Failure Times and Their Causes.....	203
12.1	Some Small Data Sets.....	203
12.1.1	Gubbins.....	203
12.1.2	Catheter Infection.....	203
12.1.3	Superalloy Testing.....	204
12.2	Basic Probability Functions: Continuous Time.....	204
12.2.1	Exponential Mixture.....	206

12.3	Hazard Functions	207
12.3.1	Exponential Mixture	207
12.3.2	Lemma	208
12.3.3	Weibull Sub-Hazards	208
12.4	Proportional Hazards	209
12.4.1	Weibull Sub-Hazards	209
12.4.2	Theorem	210
12.5	Regression Models	211
12.5.1	Proportional Hazards (PH)	211
12.5.2	Accelerated Life (AL)	211
12.5.3	Proportional Odds (PO)	211
12.5.4	Mean Residual Life (MRL)	212
12.6	Examples	212
12.6.1	Exponential Mixture	212
12.7	Exercises	214
12.8	Hints and Solutions	214
13.	Continuous Time: Parametric Inference	215
13.1	The Likelihood for Competing Risks	215
13.1.1	Forms of the Likelihood Function	215
13.1.2	Uncertainty about C	216
13.1.3	Uncertainty about T	216
13.1.4	Maximum Likelihood Estimates	217
13.1.4.1	Weibull Sub-Hazards	217
13.1.4.2	Exponential Mixture	217
13.2	Model Checking	218
13.2.1	Goodness of Fit	218
13.2.2	Uniform Residuals	218
13.3	Inference	219
13.4	Some Applications	221
13.4.1	Gubbins	221
13.4.2	Survival Times of Mice	222
13.4.3	Fracture Toughness	224
13.4.4	Length of Hospital Stay	225
13.5	Some Examples of Hazard Modelling	226
13.5.1	Exponential Mixture	226
13.5.2	Gumbel's Bivariate Exponential	227
13.5.3	Bivariate Makeham Distribution	227
13.5.4	Kimber and Grace: The Dream Team	227
13.5.5	A Clinical Trial	229
13.6	Masked Systems	231
13.7	Exercises	233
13.7.1	Applications	233
14.	Latent Lifetimes	235
14.1	Basic Probability Functions	235

14.1.1	Tsiatis's Lemma	235
14.1.2	Gumbel's Bivariate Exponential	236
14.2	Some Examples	238
14.2.1	Freund's Bivariate Exponential	238
14.2.2	Frailty Models	238
14.2.3	Multivariate Burr (MB)	239
14.2.4	Multivariate Weibull (MW)	239
14.2.5	A Stochastic Process Model	241
14.2.6	Other Applications	241
14.3	Further Aspects	242
14.3.1	Latent Failure Times versus Hazard Functions	242
14.3.2	Marginals versus Sub-Distributions	242
14.3.3	Peterson's Bounds	243
14.4	Independent Risks	244
14.4.1	Gail's Theorem	245
14.4.2	Other Applications	247
14.5	The Makeham Assumption	247
14.5.1	Proportional Hazards	248
14.6	A Risk-Removal Model	249
14.7	A Degradation Process	250
14.7.1	Wiener Process	251
14.7.2	Compound Poisson and Compound Birth Processes	251
14.7.3	Gamma Process	251
14.8	Exercises	251
14.9	Hints and Solutions	252
15.	Continuous Time: Non- and Semi-Parametric Methods	253
15.1	The Kaplan-Meier Estimator	253
15.1.1	Survival Times of Mice	255
15.2	Actuarial Approach	257
15.3	Proportional Hazards and Partial Likelihood	259
15.3.1	The Proportional Hazards (PH) Model	259
15.3.2	The Partial Likelihood	259
15.3.3	A Clinical Trial	260
15.4	The Baseline Survivor Functions	261
15.5	Other Methods and Applications	262
16.	Discrete Lifetimes	265
16.1	Basic Probability Functions	265
16.1.1	Geometric Mixture	266
16.2	Latent Lifetimes and Sub-Odds Functions	267
16.2.1	Sub-Odds Theorem	268
16.3	Some Examples	271
16.3.1	Discrete Version of Gumbel	271
16.3.2	Discrete Version of Freund	272

16.3.3	Discrete Version of Marshall–Olkin	273
16.3.4	Mixture Models	273
16.4	Parametric Estimation	274
16.4.1	Likelihood Function	275
16.4.2	Discrete Marshall–Olkin	275
16.4.3	Psychiatric Wards	276
16.5	Non-Parametric Estimation from Random Samples	277
16.5.1	Gubbins	279
16.5.2	Interval-Censored Data	281
16.5.3	Superalloy Testing	281
16.6	Asymptotic Distribution of Non-Parametric Estimators	283
16.7	Proportional Odds and Partial Likelihood	284
16.7.1	Psychiatric Wards	285
16.8	Exercises	286
16.9	Hints and Solutions	286
17.	Latent Lifetimes: Identifiability Crises	287
17.1	The Cox–Tsiatis Impasse	287
17.1.1	Tsiatis’s Theorem	287
17.1.2	Gumbel’s Bivariate Exponential	289
17.2	More General Identifiability Results	290
17.2.1	Miller’s Theorem	290
17.2.2	The LPQ Theorem	291
17.2.3	The Marshall–Olkin Distribution	295
17.3	Specified Marginals	296
17.4	Discrete Lifetimes	299
17.4.1	Discrete Freund	301
17.4.2	Discrete Marshall–Olkin	301
17.4.3	A Test for Independence of Risks	301
17.5	Regression Case	302
17.5.1	Heckman and Honore’s Theorem	302
17.5.2	Gumbel’s Bivariate Exponential	303
17.6	Censoring of Survival Data	304
17.7	Parametric Identifiability	306
 Part IV Counting Processes in Survival Analysis		
18.	Some Basic Concepts	311
18.1	Probability Spaces	311
18.2	Conditional Expectation	312
18.3	Filtrations	314
18.4	Martingales in Discrete Time	315
18.4.1	Likelihood Ratios	316
18.5	Martingales in Continuous Time	317
18.6	Counting Processes	319
18.7	Product Integrals	320

19. Survival Analysis	323
19.1 A Single Lifetime	323
19.1.1 The Intensity Process	323
19.1.2 Parametric Likelihood Function	324
19.2 Independent Lifetimes	325
19.3 Competing Risks	326
19.4 Right-Censoring	328
20. Non- and Semi-Parametric Methods	331
20.1 Survival Times	331
20.2 Competing Risks	333
20.3 Large-Sample Results	334
20.3.1 Consistency	334
20.3.2 Asymptotic Normality	334
20.3.3 Confidence Intervals	334
20.4 Hypothesis Testing	335
20.4.1 Single-Sample Case	335
20.4.2 Several Samples	336
20.5 Regression Models	338
20.5.1 Intensity Models and Time-Dependent Covariates	338
20.5.2 Proportional Hazards Model	339
20.5.3 Martingale Residuals	339
Appendix A: Terms, Notations, and Abbreviations	341
Appendix B: Basic Likelihood Methods	343
Appendix C: Some Theory for Partial Likelihood	347
Appendix D: Numerical Optimisation of Functions	351
References	355
Epilogue to First Edition	375
Index	377