
Contents

Preface	xix
Acknowledgments	xxi
List of Figures	xxiii
List of Tables	xxvii
Co-authors	xxxix
Notations	xxxiii
1 Introduction	1
1.1 General presentation of the book	1
1.2 Organization of the book	2
1.3 Notation	3
1.4 Presentation of examples	4
1.4.1 Dementia: the PAQUID study	4
1.4.2 AIDS: the ALBI trial	5
1.4.3 Colorectal cancer	6
I Classical Biostatistical Models	7
2 Inference	9
2.1 Generalities on inference: the concept of model	9
2.2 Likelihood and applications	9
2.2.1 The likelihood	9
2.2.2 Asymptotic properties of the MLE	12
2.2.3 Asymptotic tests	17
2.3 Other types of likelihoods and estimation methods	18
2.3.1 Other types of likelihood	18
2.3.2 Other methods of inference	19
2.3.2.1 Other estimation methods	19
2.3.2.2 Methods for variance estimation and tests . .	20
2.4 Model choice	21
2.4.1 Akaike criterion: AIC	21
2.4.2 Cross-validation	22

2.5	Optimization algorithms	23
2.5.1	Generalities	23
2.5.2	Newton-Raphson and Newton-like algorithms	24
2.5.2.1	Newton-Raphson	24
2.5.2.2	Marquardt algorithm	25
2.5.2.3	BHHH and BFGS algorithms	25
2.5.3	EM algorithm	26
3	Survival analysis	29
3.1	Introduction	29
3.2	Event, origin and functions of interest	30
3.2.1	Choice of the event of interest	30
3.2.2	Choice of the time-origin	30
3.2.3	Functions in survival analysis	31
3.3	Observation patterns: censoring and truncation	32
3.3.1	Censoring	32
3.3.1.1	Right-censoring	32
3.3.1.2	Left- and interval-censoring	32
3.3.2	Truncation	33
3.3.3	Risk set	33
3.3.4	The likelihood for censored and truncated data	34
3.3.4.1	Censoring	34
3.3.4.2	Truncation	35
3.4	Estimation of the survival function	35
3.4.1	The Kaplan-Meier estimator	35
3.4.2	Parametric estimators	37
3.4.2.1	The exponential model	37
3.4.2.2	The Weibull model	37
3.4.2.3	The Gamma model	37
3.4.3	Penalized likelihood approach	38
3.5	The proportional hazards model	40
3.5.1	The proportional hazards model	40
3.5.2	The partial likelihood	40
3.5.3	Parameters interpretation	42
3.5.4	Confidence interval and tests	42
3.5.5	Example: survival after entry in institution	43
3.5.6	The stratified Cox model	44
3.5.7	Time-dependent explanatory variables	44
3.5.7.1	Endogenous time-dependent explanatory variables and causality	45
3.5.8	Model checking: proportional hazards assumption	46
3.5.8.1	Graphical check	46
3.5.8.2	Residuals	47
3.5.8.3	Test for interaction between time and the co-variate	48

3.5.8.4	Solutions	49
3.5.8.5	Example	49
3.5.9	Parametric and penalized likelihood approaches	50
3.6	Accelerated failure time model	51
3.7	Counting processes approach	52
3.7.1	Counting processes	52
3.7.2	Filtration	52
3.7.3	Martingales and Doob-Meyer decomposition	53
3.7.4	Nelson-Aalen and Breslow estimators	54
3.8	Additive hazards models	55
3.8.1	The additive risks models: specification and inference	55
3.8.2	Application to mortality in institution	58
3.9	Degradation models	59
4	Models for longitudinal data	63
4.1	Linear mixed models	64
4.1.1	Genesis	64
4.1.2	Definition of the linear mixed model	65
4.1.3	Estimation	69
4.1.3.1	Maximum likelihood estimation	69
4.1.3.2	Restricted or Residual Maximum Likelihood (REML)	69
4.1.4	Inference for fixed effects	70
4.1.5	Choice of the covariance structure	71
4.1.6	Estimation of random effects	73
4.1.7	Predictions	73
4.1.8	Residuals	74
4.1.9	Goodness of fit analyses	75
4.1.10	Application	76
4.2	Generalized mixed linear models	79
4.2.1	Model specification	80
4.2.2	The mixed-effect logistic model	81
4.2.3	The mixed proportional-odds and cumulative probit models	81
4.2.4	The mixed Poisson model	82
4.2.5	Estimation	83
4.2.6	Confidence intervals and tests	85
4.2.7	Estimation of random effects	85
4.2.8	Application	85
4.2.8.1	Interpretation of the fixed effects	87
4.2.8.2	Covariance structure for the random effects	87
4.3	Non-linear mixed models	88
4.4	Marginal models and generalized estimating equations (GEE)	89
4.4.1	The GEE approach	90

4.4.2	Mixed models and maximum likelihood <i>vs.</i> marginal models and GEE	92
4.4.2.1	Required assumptions	92
4.4.2.2	Parameter interpretation	93
4.4.3	Application	94
4.4.3.1	Estimated parameter interpretation	94
4.5	Incomplete longitudinal data	96
4.5.1	Terminology	96
4.5.2	Analysis under the MAR assumption	97
4.5.3	Analysis in the MNAR case	98
4.5.4	Censored longitudinal data	99
4.6	Modeling strategies	101
4.6.1	Description of the data	101
4.6.2	Choice of the fixed effects	101
4.6.3	Choice of the covariance matrix	102
4.6.4	Goodness of fit	103

II Advanced Biostatistical Models 105

5	Extensions of mixed models	107
5.1	Mixed models for curvilinear outcomes	107
5.1.1	Model specification	108
5.1.2	Identifiability	109
5.1.3	Maximum likelihood estimation	109
5.1.4	Difference with the generalized linear mixed model	110
5.1.5	Concept of underlying latent process	110
5.1.6	Application	111
5.2	Mixed models for multivariate longitudinal data	115
5.2.1	Multivariate linear mixed model	116
5.2.1.1	Model specification	116
5.2.1.2	Application	117
5.2.2	Multivariate mixed model with latent process	119
5.2.2.1	Model specification	122
5.2.2.2	Identifiability	123
5.2.2.3	Estimation	123
5.2.2.4	Application	124
5.3	Latent class mixed models	128
5.3.1	Model specification	129
5.3.2	Maximum likelihood estimation	130
5.3.3	Posterior classification	132
5.3.4	Application	133

6	Advanced survival models	139
6.1	Relative survival	139
6.1.1	Principle	139
6.1.2	Specification	140
6.1.2.1	Relative survival	140
6.1.2.2	Excess death rate	140
6.1.3	Inference	141
6.1.4	Illustration	142
6.2	Competing risks models	144
6.2.1	Examples	144
6.2.2	Specification	145
6.2.3	Inference	146
6.2.4	Regression models with competing risks	147
6.2.4.1	Regression on cause-specific hazards	147
6.2.4.2	Regression for different types of events	148
6.2.4.3	Regression on cumulative incidence functions: Fine and Gray model	148
6.2.5	Illustration with R	149
6.2.5.1	Estimation of cumulative incidence functions	149
6.2.5.2	Log-rank test for investigating the equality of the cumulative incidence functions	149
6.2.5.3	Regression on cause-specific hazards	150
6.2.5.4	Joint model	151
6.2.5.5	Subdistribution hazards model	152
6.3	Frailty models	153
6.3.1	Type of data	153
6.3.1.1	Clustered data	153
6.3.1.2	Recurrent events	153
6.3.2	Principle	154
6.3.3	Specification	156
6.3.4	Inference	157
6.3.4.1	Semi-parametric approach using penalized likelihood	158
6.3.4.2	Approach using partial penalized likelihood	159
6.3.5	Estimation of the random effects	160
6.3.6	Application with the R package <code>frailtypack</code>	161
6.3.7	Models with interaction between frailties and observed explanatory variables	162
6.4	Extension of frailty models	164
6.4.1	Joint models for recurrent events and a terminal event	164
6.4.1.1	Specification	165
6.4.1.2	Inference	165
6.4.1.3	Application with the R package <code>frailtypack</code>	166
6.4.2	Models with nested frailties	167
6.4.2.1	Specification and inference	167

6.4.2.2	Application with the R package <i>frailtypack</i>	169
6.5	Cure models	169
7	Multistate models	175
7.1	Introduction	175
7.2	Multistate processes	176
7.2.1	Definition	176
7.2.2	Markov property and Chapman-Kolmogorov equation	177
7.2.3	Discrete-time processes	178
7.2.4	Continuous-time processes	178
7.2.5	Graphical representation	180
7.2.6	The illness-death model	181
7.2.7	Semi-Markov processes	182
7.3	Multistate models: generalities	183
7.3.1	General assumptions	183
7.3.2	Models for transition intensities	184
7.3.3	Models for variability: fixed effects and frailties	185
7.3.4	The illness-death model	185
7.4	Observation schemes	186
7.4.1	Truncation	186
7.4.2	Right-censoring	186
7.4.3	Discrete-time observations: interval-censoring	186
7.4.4	Ignorability	187
7.5	Statistical inference for multistate models observed in continuous time	188
7.5.1	Likelihood	188
7.5.2	Non-parametric inference: Aalen-Johansen estimator	189
7.5.3	Inference for semi-parametric regression models	190
7.5.4	Parametric inference	190
7.5.5	Inference by penalized likelihood	190
7.5.6	Illness-death model applied to bone-marrow graft	191
7.6	Inference for multistate models from interval-censored data	195
7.6.1	Likelihood	195
7.6.2	Non-parametric inference	196
7.6.3	Parametric inference	196
7.6.4	Inference by penalized likelihood	197
7.6.5	Inference for the illness-death model in case of interval censoring	197
7.6.5.1	The illness-death models with a mixed observation scheme	197
7.6.6	Application of the illness-death model to dementia	198
7.7	Complex functions of parameters: individualized hazards, sojourn times	206
7.8	Approach by counting processes	208
7.9	Other approaches	209

7.9.1	Prediction by "landmarking"	209
7.9.2	Pseudo-values approach	209
8	Joint models for longitudinal and time-to-event data	211
8.1	Introduction	211
8.2	Models with shared random effects	214
8.2.1	Specification	214
8.2.2	Estimation	218
8.2.2.1	Estimation in two stages	218
8.2.2.2	Estimation by joint likelihood	218
8.2.3	Assessment of the fit	220
8.2.4	Application	222
8.3	Latent class joint model	225
8.3.1	Specification	226
8.3.2	Estimation	228
8.3.3	Assessment of the fit	229
8.3.3.1	Posterior classification	229
8.3.3.2	Comparison of predictions and observations	230
8.3.3.3	Conditional independence assumption	231
8.3.4	Application	232
8.4	Latent classes vs. shared random effects	236
8.5	The joint model as prognostic model	237
8.6	Extension of joint models	239
8.6.1	Complex longitudinal data	239
8.6.2	Multiple events	241
9	The dynamic approach to causality	243
9.1	Introduction	243
9.1.1	The causality issue	243
9.1.1.1	Philosophical aspects	243
9.1.1.2	Causality in the life sciences	244
9.1.1.3	Causality in epidemiology	245
9.1.1.4	Causality in biostatistics	246
9.1.2	The counterfactual approach	246
9.1.2.1	Principle	246
9.1.2.2	Criticism	247
9.1.3	The dynamical approach to causality	247
9.1.3.1	Representation by stochastic processes	248
9.1.3.2	Local influence in stochastic processes	249
9.1.3.3	Systems and causal influences	250
9.1.3.4	Applications	251
9.2	Local independence, direct and indirect influence	252
9.2.1	Notations	252
9.2.2	Local independence, direct and indirect influence	252
9.2.3	Extension to random horizon and the problem of death	254

9.2.4	Differential equation representation	255
9.2.5	The deterministic case	255
9.2.6	Graph representation	256
9.3	Causal influences	256
9.3.1	Systems and causal influence	256
9.3.2	Conditional and marginal causal effects	260
9.3.2.1	Conditional causal effects	260
9.3.2.2	Expected causal effect	263
9.3.2.3	Marginal causal effect	263
9.3.3	Systems with no unmeasured confounders	264
9.3.4	Assumptions for causal interpretation	264
9.3.5	Observations	265
9.4	The dynamic approach to causal reasoning in ageing studies	266
9.4.1	The particular significance of death	266
9.4.2	Examples of causal reasoning in ageing studies	266
9.4.3	Observation and inference	269
9.5	Mechanistic models	270
9.5.1	Generalities	270
9.5.2	The general structure of a mechanistic model	271
9.5.3	Inference in mechanistic models	273
9.5.4	Computational challenge in mechanistic models	275
9.5.5	Modeling the HIV-immune system interaction	276
9.5.5.1	The ODE system	276
9.5.5.2	Intra- and inter-individual variability	278
9.5.5.3	Observation model	278
9.5.5.4	Application on the ALBI trial	278
9.6	The issue of dynamic treatment regimes	282
9.6.1	Effect of HAART on CD4 counts: notation and MSM	283
9.6.2	Assumptions and inference for the dynamic approach	284
9.6.2.1	Assumptions	284
9.6.2.2	Inference	286
9.6.2.3	Dynamic treatment regimes and mechanistic models	287
9.6.3	Effect of HAART on CD4 counts: results from the dynamic models	288
10	Appendix: software	291
10.1	Appendix for Chapter 3: survival analysis	291
10.1.1	SAS and R softwares	291
10.1.2	PROC LIFETEST	291
10.1.3	PROC PHREG	292
10.1.3.1	Testing the proportionality assumption with residuals	293
10.1.3.2	Testing the proportionality assumption with interaction with time	294

10.1.3.3	Time-dependent covariates	294
10.1.4	The R package <code>survival</code>	295
10.1.5	The package <code>SmoothHazard</code>	296
10.2	Appendix for Chapter 4: models for longitudinal data	296
10.2.1	Linear mixed models	296
10.2.2	Generalized linear mixed models	300
10.2.3	Non-linear mixed models	303
10.2.4	Marginal models	303
10.3	Appendix for Chapter 5: extensions of mixed models	304
10.3.1	Mixed models for curvilinear data	304
10.3.2	Multivariate linear mixed models	306
10.3.3	Multivariate mixed models with latent process	308
10.3.4	Latent class mixed models	310
10.4	Appendix for Chapter 6: advanced survival	311
10.4.1	Relative survival models with the R package <code>relsurv</code>	311
10.4.2	Competing risks models with the R packages <code>cmprsk</code> and <code>mstate</code>	312
10.4.3	Shared frailty models with R packages <code>coxph</code> and <code>frailtypack</code>	315
10.4.4	Joint frailty models with the R package <code>frailtypack</code> and the SAS procedure <code>NLMIXED</code>	317
10.4.5	Nested frailty models with the R package <code>frailtypack</code>	320
10.5	Appendix for Chapter 7: multistate models	322
10.5.1	The package <code>mstate</code>	322
10.5.2	The package <code>SmoothHazard</code>	325
10.6	Appendix for Chapter 8: joint models for longitudinal data and time to events	327
10.6.1	Joint models with shared random effects	327
10.6.2	Latent class joint models	329
	Bibliography	331
	Index	355