

Odd O. Aalen, Ørnulf Borgan and Håkon K. Gjessing Survival and Event History Analysis

Time-to-event data are ubiquitous in fields such as medicine, biology, demography, sociology, economics and reliability theory. Recently, a need to analyze more complex event histories has emerged. Examples are individuals that move among several states, frailty that makes some units fail before others, internal time-dependent covariates, and the estimation of causal effects from observational data.

The aim of this book is to bridge the gap between standard textbook models and a range of models where the dynamic structure of the data manifests itself fully. The common denominator of such models is stochastic processes. The authors show how counting processes, martingales, and stochastic integrals fit very nicely with censored data. Beginning with standard analyses such as Kaplan–Meier plots and Cox regression, the presentation progresses to the additive hazard model and recurrent event data. Stochastic processes are also used as natural models for individual frailty; they allow sensible interpretations of a number of surprising artifacts seen in population data.

The stochastic process framework is naturally connected to causality. The authors show how dynamic path analyses can incorporate many modern causality ideas in a framework that takes the time aspect seriously.

To make the material accessible to the reader, a large number of practical examples, mainly from medicine, are developed in detail. Stochastic processes are introduced in an intuitive and non-technical manner. The book is aimed at investigators who use event history methods and want a better understanding of the statistical concepts. It is suitable as a textbook for graduate courses in statistics and biostatistics.

Odd O. Aalen is professor of medical statistics at the University of Oslo, Norway. His Ph.D. from the University of California, Berkeley in 1975 introduced counting processes and martingales in event history analysis. He has also contributed to numerous other areas of event history analysis, such as additive hazards regression, frailty, and causality through dynamic modeling.

Ørnulf Borgan is professor of statistics at the University of Oslo, Norway. Since his Ph.D. in 1984 he has contributed extensively to event history analysis. He is co-author of the monograph *Statistical Models Based on Counting Processes* and is editor of *Scandinavian Journal of Statistics*.

Håkon K. Gjessing is professor of medical statistics at the Norwegian Institute of Public Health and the University of Bergen, Norway. Since his Ph.D. in probability in 1995, he has worked on a broad range of theoretical and applied problems in biostatistics.

STATISTICS |
LIFE SCIENCES, MEDICINE,
HEALTH SCIENCES

ISBN 978-0-387-20287-7



Contents

Preface	vii
1 An introduction to survival and event history analysis	1
1.1 Survival analysis: basic concepts and examples	2
1.1.1 What makes survival special: censoring and truncation	3
1.1.2 Survival function and hazard rate	5
1.1.3 Regression and frailty models	7
1.1.4 The past	9
1.1.5 Some illustrative examples	9
1.2 Event history analysis: models and examples	16
1.2.1 Recurrent event data	17
1.2.2 Multistate models	18
1.3 Data that do not involve time	24
1.4 Counting processes	25
1.4.1 What is a counting process?	25
1.4.2 Survival times and counting processes	28
1.4.3 Event histories and counting processes	32
1.5 Modeling event history data	33
1.5.1 The multiplicative intensity model	34
1.5.2 Regression models	34
1.5.3 Frailty models and first passage time models	35
1.5.4 Independent or dependent data?	36
1.6 Exercises	37
2 Stochastic processes in event history analysis	41
2.1 Stochastic processes in discrete time	43
2.1.1 Martingales in discrete time	43
2.1.2 Variation processes	44
2.1.3 Stopping times and transformations	45
2.1.4 The Doob decomposition	47
2.2 Processes in continuous time	48

2.2.1	Martingales in continuous time	48
2.2.2	Stochastic integrals	50
2.2.3	The Doob-Meyer decomposition	52
2.2.4	The Poisson process	52
2.2.5	Counting processes	53
2.2.6	Stochastic integrals for counting process martingales	55
2.2.7	The innovation theorem	56
2.2.8	Independent censoring	57
2.3	Processes with continuous sample paths	61
2.3.1	The Wiener process and Gaussian martingales	61
2.3.2	Asymptotic theory for martingales: intuitive discussion	62
2.3.3	Asymptotic theory for martingales: mathematical formulation	63
2.4	Exercises	66
3	Nonparametric analysis of survival and event history data	69
3.1	The Nelson-Aalen estimator	70
3.1.1	The survival data situation	71
3.1.2	The multiplicative intensity model	76
3.1.3	Handling of ties	83
3.1.4	Smoothing the Nelson-Aalen estimator	85
3.1.5	The estimator and its small sample properties	87
3.1.6	Large sample properties	89
3.2	The Kaplan-Meier estimator	90
3.2.1	The estimator and confidence intervals	90
3.2.2	Handling tied survival times	94
3.2.3	Median and mean survival times	95
3.2.4	Product-integral representation	97
3.2.5	Excess mortality and relative survival	99
3.2.6	Martingale representation and statistical properties	103
3.3	Nonparametric tests	104
3.3.1	The two-sample case	105
3.3.2	Extension to more than two samples	109
3.3.3	Stratified tests	110
3.3.4	Handling of tied observations	111
3.3.5	Asymptotics	112
3.4	The empirical transition matrix	114
3.4.1	Competing risks and cumulative incidence functions	114
3.4.2	An illness-death model	117
3.4.3	The general case	120
3.4.4	Martingale representation and large sample properties	123
3.4.5	Estimation of (co)variances	124
3.5	Exercises	126

4 Regression models 131

4.1 Relative risk regression 133

4.1.1 Partial likelihood and inference for regression coefficients .. 134

4.1.2 Estimation of cumulative hazards and survival probabilities . 141

4.1.3 Martingale residual processes and model check 142

4.1.4 Stratified models 148

4.1.5 Large sample properties of $\hat{\beta}$ 149

4.1.6 Large sample properties of estimators of cumulative hazards and survival functions 152

4.2 Additive regression models 154

4.2.1 Estimation in the additive hazard model 157

4.2.2 Interpreting changes over time 163

4.2.3 Martingale tests and a generalized log-rank test 164

4.2.4 Martingale residual processes and model check 167

4.2.5 Combining the Cox and the additive models 171

4.2.6 Adjusted monotone survival curves for comparing groups .. 172

4.2.7 Adjusted Kaplan-Meier curves under dependent censoring .. 175

4.2.8 Excess mortality models and the relative survival function .. 179

4.2.9 Estimation of Markov transition probabilities 181

4.3 Nested case-control studies 190

4.3.1 A general framework for nested-case control sampling 192

4.3.2 Two important nested case-control designs 194

4.3.3 Counting process formulation of nested case-control sampling 195

4.3.4 Relative risk regression for nested case-control data 196

4.3.5 Additive regression for nested case-control data: results 200

4.3.6 Additive regression for nested case-control data: theory 202

4.4 Exercises 203

5 Parametric counting process models 207

5.1 Likelihood inference 208

5.1.1 Parametric models for survival times 208

5.1.2 Likelihood for censored survival times 209

5.1.3 Likelihood for counting process models 210

5.1.4 The maximum likelihood estimator and related tests 213

5.1.5 Some applications 214

5.2 Parametric regression models 223

5.2.1 Poisson regression 223

5.3 Proof of large sample properties 226

5.4 Exercises 228

6 Unobserved heterogeneity: The odd effects of frailty 231

6.1 What is randomness in survival models? 233

6.2 The proportional frailty model 234

6.2.1 Basic properties 234

6.2.2	The Gamma frailty distribution	235
6.2.3	The PVF family of frailty distributions	238
6.2.4	Lévy-type frailty distributions	242
6.3	Hazard and frailty of survivors	243
6.3.1	Results for the PVF distribution	243
6.3.2	Cure models	244
6.3.3	Asymptotic distribution of survivors	245
6.4	Parametric models derived from frailty distributions	246
6.4.1	A model based on Gamma frailty: the Burr distribution	246
6.4.2	A model based on PVF frailty	247
6.4.3	The Weibull distribution derived from stable frailty	248
6.4.4	Frailty and estimation	249
6.5	The effect of frailty on hazard ratio	250
6.5.1	Decreasing relative risk and crossover	250
6.5.2	The effect of discontinuing treatment	253
6.5.3	Practical implications of artifacts	255
6.5.4	Frailty models yielding proportional hazards	257
6.6	Competing risks and false protectivity	260
6.7	A frailty model for the speed of a process	262
6.8	Frailty and association between individuals	264
6.9	Case study: A frailty model for testicular cancer	265
6.10	Exercises	268
7	Multivariate frailty models	271
7.1	Censoring in the multivariate case	272
7.1.1	Censoring for recurrent event data	273
7.1.2	Censoring for clustered survival data	274
7.2	Shared frailty models	275
7.2.1	Joint distribution	276
7.2.2	Likelihood	276
7.2.3	Empirical Bayes estimate of individual frailty	278
7.2.4	Gamma distributed frailty	279
7.2.5	Other frailty distributions suitable for the shared frailty model	284
7.3	Frailty and counting processes	286
7.4	Hierarchical multivariate frailty models	288
7.4.1	A multivariate model based on Lévy-type distributions	289
7.4.2	A multivariate stable model	290
7.4.3	The PVF distribution with $m = 1$	290
7.4.4	A trivariate model	290
7.4.5	A simple genetic model	291
7.5	Case study: A hierarchical frailty model for testicular cancer	293
7.6	Random effects models for transformed times	296
7.6.1	Likelihood function	296
7.6.2	General case	298

7.6.3	Comparing frailty and random effects models	299
7.7	Exercises	299
8	Marginal and dynamic models for recurrent events and clustered survival data	301
8.1	Intensity models and rate models	302
8.1.1	Dynamic covariates	304
8.1.2	Connecting intensity and rate models in the additive case	305
8.2	Nonparametric statistical analysis	308
8.2.1	A marginal Nelson-Aalen estimator for clustered survival data	308
8.2.2	A dynamic Nelson-Aalen estimator for recurrent event data	309
8.3	Regression analysis of recurrent events and clustered survival data	311
8.3.1	Relative risk models	313
8.3.2	Additive models	315
8.4	Dynamic path analysis of recurrent event data	324
8.4.1	General considerations	325
8.5	Contrasting dynamic and frailty models	331
8.6	Dynamic models – theoretical considerations	333
8.6.1	A dynamic view of the frailty model for Poisson processes	333
8.6.2	General view on the connection between dynamic and frailty models	334
8.6.3	Are dynamic models well defined?	336
8.7	Case study: Protection from natural infections with enterotoxigenic <i>Escherichia coli</i>	340
8.8	Exercises	346
9	Causality	347
9.1	Statistics and causality	347
9.1.1	Schools of statistical causality	349
9.1.2	Some philosophical aspects	351
9.1.3	Traditional approaches to causality in epidemiology	353
9.1.4	The great theory still missing?	353
9.2	Graphical models for event history analysis	354
9.2.1	Time-dependent covariates	356
9.3	Local characteristics - dynamic model	361
9.3.1	Dynamic path analysis – a general view	363
9.3.2	Direct and indirect effects – a general concept	365
9.4	Granger-Schweder causality and local dependence	367
9.4.1	Local dependence	367
9.4.2	A general definition of Granger-Schweder causality	370
9.4.3	Statistical analysis of local dependence	371
9.5	Counterfactual causality	373
9.5.1	Standard survival analysis and counterfactuals	376
9.5.2	Censored and missing data	377

9.5.3	Dynamic treatment regimes	378
9.5.4	Marginal versus joint modeling	380
9.6	Marginal modeling	380
9.6.1	Marginal structural models	380
9.6.2	G-computation: A Markov modeling approach	382
9.7	Joint modeling	383
9.7.1	Joint modeling as an alternative to marginal structural models	384
9.7.2	Modeling dynamic systems	385
9.8	Exercises	385
10	First passage time models: Understanding the shape of the hazard rate	387
10.1	First hitting time; phase type distributions	389
10.1.1	Finite birth-death process with absorbing state	389
10.1.2	First hitting time as the time to event	390
10.1.3	The risk distribution of survivors	392
10.1.4	Reversibility and progressive models	393
10.2	Quasi-stationary distributions	395
10.2.1	Infinite birth-death process (infinite random walk)	397
10.2.2	Interpretation	398
10.3	Wiener process models	399
10.3.1	The inverse Gaussian hitting time distribution	400
10.3.2	Comparison of hazard rates	402
10.3.3	The distribution of survivors	404
10.3.4	Quasi-stationary distributions for the Wiener process with absorption	405
10.3.5	Wiener process with a random initial value	407
10.3.6	Wiener process with lower absorbing and upper reflecting barriers	408
10.3.7	Wiener process with randomized drift	408
10.3.8	Analyzing the effect of covariates for the randomized Wiener process	410
10.4	Diffusion process models	416
10.4.1	The Kolmogorov equations and a formula for the hazard rate	418
10.4.2	An equation for the quasi-stationary distribution	419
10.4.3	The Ornstein-Uhlenbeck process	421
10.5	Exercises	424
11	Diffusion and Lévy process models for dynamic frailty	425
11.1	Population versus individual survival	426
11.2	Diffusion models for the hazard	428
11.2.1	A simple Wiener process model	428

11.2.2	The hazard rate as the square of an Ornstein-Uhlenbeck process	430
11.2.3	More general diffusion processes	431
11.3	Models based on Lévy processes	432
11.4	Lévy processes and subordinators	433
11.4.1	Laplace exponent	433
11.4.2	Compound Poisson processes and the PVF process	434
11.4.3	Other examples of subordinators	435
11.4.4	Lévy measure	436
11.5	A Lévy process model for the hazard	438
11.5.1	Population survival	440
11.5.2	The distribution of h conditional on no event	440
11.5.3	Standard frailty models	441
11.5.4	Moving average	441
11.5.5	Accelerated failure times	443
11.6	Results for the PVF processes	444
11.6.1	Distribution of survivors for the PVF processes	445
11.6.2	Moving average and the PVF process	446
11.7	Parameterization and estimation	448
11.8	Limit results and quasi-stationary distributions	450
11.8.1	Limits for the PVF process	452
11.9	Exercises	453
A	Markov processes and the product-integral	457
A.1	Hazard, survival, and the product-integral	458
A.2	Markov chains, transition intensities, and the Kolmogorov equations	461
A.2.1	Discrete time-homogeneous Markov chains	463
A.2.2	Continuous time-homogeneous Markov chains	465
A.2.3	The Kolmogorov equations for homogeneous Markov chains	467
A.2.4	Inhomogeneous Markov chains and the product-integral	468
A.2.5	Common multistate models	471
A.3	Stationary and quasi-stationary distributions	475
A.3.1	The stationary distribution of a discrete Markov chain	475
A.3.2	The quasi-stationary distribution of a Markov chain with an absorbing state	477
A.4	Diffusion processes and stochastic differential equations	479
A.4.1	The Wiener process	480
A.4.2	Stochastic differential equations	482
A.4.3	The Ornstein-Uhlenbeck process	484
A.4.4	The infinitesimal generator and the Kolmogorov equations for a diffusion process	486
A.4.5	The Feynman-Kac formula	488
A.5	Lévy processes and subordinators	490

A.5.1	The Lévy process	491
A.5.2	The Laplace exponent	493
B	Vector-valued counting processes, martingales and stochastic integrals	495
B.1	Counting processes, intensity processes and martingales	495
B.2	Stochastic integrals	496
B.3	Martingale central limit theorem	497
	References	499
	Author index	521
	Index	529