

Contents

1	Introduction	1
1.1	Summary	1
1.2	Motivation	1
1.3	Objectives	6
1.4	Controversies	7
1.5	Data sets	10
1.6	Use as a graduate text	10
1.7	Exercises and class projects	11
2	Background: Probability	13
2.1	Summary	13
2.2	Motivation	14
2.3	Integration and measure	14
2.4	Random variables and probability measure	17
2.5	Distributions and densities	19
2.6	Expectation	24
2.7	Order statistics and their expectations	26
2.8	Entropy and variance	32
2.9	Approximations	36
2.10	Stochastic processes	41
2.11	Brownian motion	42
2.12	Counting processes and martingales	51
2.13	Exercises and class projects	59
3	Background: General inference	63
3.1	Summary	63
3.2	Motivation	64
3.3	Limit theorems for sums of random variables	64
3.4	Functional Central Limit Theorem	68

3.5	Empirical distribution function	71
3.6	Inference for martingales and stochastic integrals . . .	74
3.7	Estimating equations	80
3.8	Inference using resampling techniques	88
3.9	Explained variation	91
3.10	Exercises and class projects	99
4	Background: Survival analysis	103
4.1	Summary	103
4.2	Motivation	103
4.3	Basic tools	104
4.4	Some potential models	112
4.5	Censoring	120
4.6	Competing risks as a particular type of censoring . . .	124
4.7	Exercises and class projects	125
5	Marginal survival	129
5.1	Summary	129
5.2	Motivation	129
5.3	Maximum likelihood estimation	130
5.4	Empirical estimate (no censoring)	136
5.5	Empirical estimate (with censoring)	138
5.6	Exercises and class projects	148
6	Regression models and subject heterogeneity	151
6.1	Summary	151
6.2	Motivation	152
6.3	General or nonproportional hazards model	153
6.4	Proportional hazards model	154
6.5	The Cox regression model	155
6.6	Modeling multivariate problems	165
6.7	Partially proportional hazards models	172
6.8	Non proportional hazards model with intercept	183
6.9	Time-dependent covariates	186
6.10	Time-dependent covariates and non proportional hazards models	188
6.11	Proportional hazards models in epidemiology	189
6.12	Exercises and class projects	199

7	Inference: Estimating equations	203
7.1	Summary	203
7.2	Motivation	204
7.3	The observations	205
7.4	Main theorem	207
7.5	The estimating equations	219
7.6	Consistency and asymptotic normality of $\tilde{\beta}$	225
7.7	Interpretation for β^* as average effect	226
7.8	Exercises and class projects	228
8	Inference: Functions of Brownian motion	231
8.1	Summary	231
8.2	Motivation	232
8.3	Brownian motion approximations	233
8.4	Non and partially proportional hazards models	238
8.5	Tests based on functions of Brownian motion	239
8.6	Multivariate model	249
8.7	Graphical representation of regression effects	254
8.8	Operating characteristics of tests	258
8.9	Goodness-of-fit tests	260
8.10	Exercises and class projects	263
9	Inference: Likelihood	267
9.1	Summary	267
9.2	Motivation	267
9.3	Likelihood solution for parametric models	268
9.4	Likelihood solution for exponential models	270
9.5	Semi-parametric likelihood solution	275
9.6	Other likelihood expressions	280
9.7	Goodness-of-fit of likelihood estimates	287
9.8	Exercises and class projects	292
10	Inference: Stochastic integrals	295
10.1	Summary	295
10.2	Motivation	295
10.3	Counting process framework to the model	296
10.4	Some nonparametric statistics	298
10.5	Stochastic integral representation of score statistic	299
10.6	Exercises and class projects	308

11 Inference: Small samples	311
11.1 Summary	311
11.2 Motivation	311
11.3 Additive and multiplicative models	312
11.4 Estimation: First two moments	315
11.5 Edgeworth and saddlepoint approximations	316
11.6 Distribution of estimating equation	318
11.7 Simulation studies	323
11.8 Example	327
11.9 Further points	329
11.10 Exercises and class projects	330
12 Inference: Changepoint models	331
12.1 Summary	331
12.2 Motivation	331
12.3 Some changepoint models	333
12.4 Inference when γ is known	334
12.5 Inference when γ is unknown	335
12.6 Maximum of log-rank type tests	338
12.7 Computational aspects	339
12.8 Two groups with crossing hazards	342
12.9 Illustrations	347
12.10 Some guidelines in model and test selection	354
12.11 Exercises and class projects	358
13 Explained variation	359
13.1 Summary	359
13.2 Motivation	359
13.3 Finding a suitable measure of R^2	361
13.4 An R^2 measure based on Schoenfeld residuals	366
13.5 Finite sample properties of R^2 and $R_{\mathcal{E}}^2$	370
13.6 Large sample properties	371
13.7 Interpretation	375
13.8 Simulation results	379
13.9 Extensions	381
13.10 Theoretical construction for distance measures	386
13.11 Isolation method for bias-reduction	390
13.12 Illustrations from studies in cancer	395
13.13 Exercises and class projects	403

14 Explained randomness	407
14.1 Summary	407
14.2 Motivation	407
14.3 Information gain and explained randomness	409
14.4 Explained randomness in Z given T	411
14.5 Approximation of ρ_1^2 by ρ_2^2	416
14.6 Simple working approximation of ρ_1^2 and ρ_2^2	417
14.7 Multiple coefficient of explained randomness	420
14.8 Partially explained randomness	421
14.9 Isolation method for bias-reduction	422
14.10 Simulations	423
14.11 Illustrations	426
14.12 Further extensions	429
14.13 Exercises and class projects	435
15 Survival given covariates	437
15.1 Summary	437
15.2 Motivation	438
15.3 Probability that T_i is greater than T_j	438
15.4 Estimating conditional survival given that $Z = z$	441
15.5 Estimating conditional survival given $Z \in H$	442
15.6 Estimating the variance of $\hat{S}(t Z \in H)$	444
15.7 Relative merits of competing estimators	448
15.8 Illustrations	448
15.9 Generalization under multiplicative relative risk	451
15.10 Informative censoring	454
15.11 Exercises and class projects	460
16 Proofs of theorems, lemmas and corollaries	463
Bibliography	477
Index	539