
Contents

I MODELS

1	Model building	3
1.1	Data generating mechanisms	3
1.2	Probability models	7
1.3	Regression models	12
1.4	Stochastic processes	15
1.5	Inferences about models	17
1.6	Characteristics of modelling	20
1.7	Exercises	22
2	Exponential family of probability distributions	25
2.1	General theory	25
2.1.1	Exponential representation	26
2.1.2	Fundamental parametrizations	28
2.1.3	Generating functions	29
2.1.4	One-parameter families	32
2.1.5	Regularity conditions	33
2.2	Distributions for random samples	34
2.2.1	Product families	34
2.2.2	Marginality	35
2.2.3	Conditionality	36
2.3	Special cases	37
2.3.1	Truncated distributions	37
2.3.2	Bivariate distributions	38
2.3.3	Double exponential family	39
2.3.4	Conditional exponential family	39
2.3.5	Curved exponential family	40
2.4	Parametrizations	41
2.4.1	Types of parameters	41
2.4.2	Parameter transformations	41
2.5	Related families	43
2.5.1	Exponential dispersion family	43
2.5.2	Extra parameters	46

2.6	Conjugate distributions	47
2.6.1	Exponential family	47
2.6.2	Exponential dispersion family	49
2.7	Other families	50
2.7.1	Transformation families	51
2.7.2	Stable distributions	53
2.8	Appendix: technical results	56
2.8.1	Exponential representation	56
2.8.2	Fundamental parametrizations	57
2.8.3	Generating functions	58
2.8.4	Regularity conditions	58
2.9	Exercises	60

II INFERENCE

3	Likelihood	69
3.1	Statistical inference	69
3.1.1	The problem	69
3.1.2	Inference and decision-making	71
3.1.3	A likelihood principle	73
3.2	Likelihood function	75
3.2.1	Definition	75
3.2.2	Maximum likelihood estimate	80
3.2.3	Normed likelihood and deviance	82
3.3	Samples	85
3.3.1	Sample size calculation	86
3.3.2	Stopping rules	87
3.4	Special statistics	89
3.4.1	Sufficient statistics	89
3.4.2	Ancillary statistics	92
3.5	Derived functions	94
3.5.1	Score	94
3.5.2	Observed and expected information	96
3.5.3	Kullback–Leibler information	101
3.5.4	Pivotal quantities	101
3.6	More complex models	103
3.6.1	Model selection	103
3.6.2	Parametrization invariance	109
3.6.3	Normed likelihood regions	110
3.6.4	Prior information	114
3.6.5	Prediction	116
3.6.6	Paradoxes	118
3.7	Exponential family	120

3.7.1	Maximum likelihood estimates	121
3.7.2	Deviance	121
3.7.3	Estimation	122
3.7.4	Sample size	123
3.7.5	Sufficient and ancillary statistics	124
3.7.6	Score and information	124
3.7.7	Appendix	128
3.8	Exercises	130
4	Goodness of fit	141
4.1	Likelihood comparisons	142
4.1.1	Embedding	142
4.1.2	Robustness	146
4.1.3	Contaminated observations	147
4.2	Tests of significance	148
4.2.1	Pure tests	148
4.2.2	Exact tests	153
4.2.3	Sufficient and ancillary statistics	156
4.3	Diagnostics	158
4.3.1	Departures from models	158
4.3.2	Residuals	160
4.3.3	Isolated departures	171
4.3.4	Systematic departures	175
4.4	Exercises	176

III APPROXIMATIONS

5	Asymptotics	185
5.1	Preliminaries	186
5.1.1	Regularity	186
5.1.2	Mean and variance of the score statistic	188
5.2	Small sample asymptotics	190
5.2.1	Large sample asymptotics	190
5.2.2	Small dispersion asymptotics	191
5.2.3	High precision asymptotics	192
5.2.4	Criteria for approximations	193
5.3	Maximum likelihood estimate statistic	194
5.3.1	Exact distribution	194
5.3.2	Consistency	196
5.3.3	Normal approximation and efficiency	198
5.3.4	Standard errors	203
5.3.5	Modified profile likelihood approximation	206
5.4	Deviance statistic	207

5.4.1	Exact distribution	207
5.4.2	Normal approximation	208
5.4.3	Chi-squared distribution	209
5.4.4	Likelihoods as martingales	213
5.5	Score statistic	215
5.5.1	Exact distribution	215
5.5.2	Asymptotic distribution	215
5.5.3	Score statistics as martingales	217
5.6	Asymptotic significance tests	218
5.6.1	Likelihood ratio test	218
5.6.2	Wald and score tests	219
5.7	Appendix: large sample approximations	220
5.7.1	Laplace approximation	220
5.7.2	Edgeworth expansions	221
5.7.3	Saddlepoint approximations	222
5.8	Exercises	226
6	Factoring the likelihood function	233
6.1	Parametrization orthogonality	234
6.1.1	Estimation orthogonality	235
6.1.2	Design orthogonality	235
6.1.3	Information orthogonality	236
6.1.4	Likelihood orthogonality	239
6.2	Partial sufficiency	240
6.2.1	Sufficiency and ancillarity	240
6.2.2	L-sufficiency	242
6.2.3	Cuts	243
6.2.4	S-sufficiency and S-ancillarity	244
6.2.5	Marginal and conditional likelihoods	245
6.3	Correcting the profile likelihood	248
6.3.1	p^* formula	248
6.3.2	Modified profile likelihood	249
6.3.3	Adjusted profile likelihood	251
6.3.4	Directed deviance	254
6.3.5	Pivotal likelihood	256
6.4	Special models	257
6.4.1	Random effects	257
6.4.2	Missing values	260
6.4.3	Coarsened data	262
6.5	Pseudo-likelihood	263
6.5.1	Partial likelihood	263
6.5.2	Composite likelihood	264
6.5.3	Quasi-likelihood	264
6.5.4	Generalized estimating equations (GEE)	265

6.6	Exercises	267
-----	-----------	-----

IV DECISIONS

7	Frequentist decision-making	279
7.1	Some principles for decision-making	279
7.1.1	Utility and risk	280
7.1.2	Repeated sampling	280
7.2	Tests: simple null hypothesis	282
7.2.1	Types of hypotheses	283
7.2.2	Simple alternative hypothesis	284
7.2.3	Power function	289
7.2.4	Sample size	291
7.2.5	Composite alternative hypothesis	292
7.3	Tests: composite null hypothesis	295
7.3.1	Types of null hypotheses	295
7.3.2	Similar regions	295
7.3.3	Invariant tests	298
7.3.4	Likelihood ratio tests	300
7.3.5	Sequential tests	301
7.4	Point estimation	303
7.4.1	Bias and variance	303
7.4.2	Decision theory	309
7.5	Interval estimation	312
7.5.1	Confidence limits	312
7.5.2	Confidence regions	314
7.6	Exercises	317
8	Bayesian decision-making	321
8.1	Interpretation	322
8.1.1	Prior probabilities	322
8.1.2	Bayes' formula	323
8.1.3	Rational Bayes	326
8.1.4	Subjective Bayes	326
8.1.5	Decision theoretic Bayes	328
8.1.6	Robust Bayes	328
8.2	Applications to statistical model building	329
8.2.1	Population frequencies	329
8.2.2	Empirical Bayes	330
8.2.3	Kalman filter	331
8.3	Types of prior distributions	332
8.3.1	Non-informative priors	334
8.3.2	Vague priors	336

8.3.3	Invariant priors	336
8.3.4	Reference priors	337
8.3.5	Conjugate priors	339
8.3.6	Hierarchical priors	341
8.3.7	Robust priors	342
8.4	Point estimation	343
8.4.1	Posterior estimates	343
8.4.2	Bayes risk	345
8.5	Interval estimation	347
8.5.1	Intervals by integration	347
8.5.2	Likelihood regions	349
8.5.3	Nuisance parameters	349
8.6	Testing	352
8.6.1	Posterior odds and Bayes factors	352
8.6.2	Sharp hypotheses	353
8.6.3	Jeffreys–Lindley paradox	355
8.6.4	Bayesian information criterion (BIC)	357
8.7	Bayesian and frequentist decision-making	357
8.8	Exercises	360

V EXAMPLES

9	Poisson regression	367
9.1	Models	367
9.2	Likelihood function	368
9.3	Goodness of fit	373
9.4	Asymptotic inference	379
9.5	Factoring the likelihood function	379
9.6	Frequentist decision-making	379
9.7	Bayesian decision-making	380
9.8	Exercises	381
10	Binomial regression	385
10.1	Models	385
10.2	Likelihood function	387
10.3	Goodness of fit	394
10.4	Asymptotic inference	397
10.5	Factoring the likelihood function	397
10.6	Frequentist decision-making	399
10.7	Bayesian decision-making	399
10.8	Exercises	399

VI APPENDICES

A	Elements of measure theory	407
	A.1 Measure spaces	407
	A.2 Integration	408
	A.3 Convexity	408
B	Review of probability theory	411
	B.1 Probability spaces	411
	B.1.1 Probability	411
	B.1.2 Random variables	413
	B.1.3 Probability laws	413
	B.1.4 Independence	414
	B.2 Integration	415
	B.2.1 Expectation	415
	B.2.2 Radon–Nikodým theorem	416
	B.3 Probability functions	416
	B.3.1 Probability density function	416
	B.3.2 Joint probability	417
	B.3.3 Marginal probability	418
	B.3.4 Convolutions	418
	B.3.5 Conditional probability	418
	B.3.6 Bayes' formula	420
	B.3.7 Paradoxes	420
	B.3.8 Statistics	423
	B.4 Martingales	424
	B.4.1 Filtrations	424
	B.4.2 Martingales	424
	B.4.3 Stopping times	425
	B.4.4 Transforms and the Doob decomposition	426
	B.5 Generating functions	427
	B.5.1 Probability generating function	427
	B.5.2 Moment generating function	428
	B.5.3 Cumulant generating function	428
	B.5.4 Characteristic function	429
	B.6 Asymptotic results	430
	B.6.1 Useful inequalities	430
	B.6.2 Modes of convergence	430
	B.6.3 Asymptotic laws	431
	B.6.4 O and o notation	433
C	Normal distribution statistics	435
	C.1 Chi-squared distribution	435
	C.2 Student t distribution	435

C.3	F distribution	437
D	Numerical methods	441
D.1	Maximization	441
D.1.1	Newton–Raphson	441
D.1.2	Derivative-free	442
D.1.3	EM algorithm	443
D.1.4	Simulated annealing	444
D.2	Integration	445
D.2.1	Quadrature	445
D.2.2	Monte Carlo	445
	Bibliography	449
	Index	469