

# Maximum Likelihood Estimation and Inference

With Examples in R, SAS and ADMB

**Russell B. Millar**

*Department of Statistics, University of Auckland, New Zealand*



A John Wiley & Sons, Ltd., Publication

# Contents

<b>Preface</b>	<b>xiii</b>
<b>Part I PRELIMINARIES</b>	<b>1</b>
<b>1 A taste of likelihood</b>	<b>3</b>
1.1 Introduction . . . . .	3
1.2 Motivating example . . . . .	4
1.2.1 ML estimation and inference for the binomial . . . . .	4
1.2.2 Approximate normality versus likelihood ratio . . . . .	8
1.3 Using SAS, R and ADMB . . . . .	9
1.3.1 Software resources . . . . .	11
1.4 Implementation of the motivating example . . . . .	11
1.4.1 Binomial example in SAS . . . . .	11
1.4.2 Binomial example in R . . . . .	14
1.4.3 Binomial example in ADMB . . . . .	15
1.5 Exercises . . . . .	17
<b>2 Essential concepts and iid examples</b>	<b>18</b>
2.1 Introduction . . . . .	18
2.2 Some necessary notation . . . . .	19
2.2.1 MLEs of functions of the parameters . . . . .	22
2.3 Interpretation of likelihood . . . . .	23
2.4 IID examples . . . . .	25
2.4.1 IID Bernoulli (i.e. binomial) . . . . .	25
2.4.2 IID normal . . . . .	26
2.4.3 IID uniform . . . . .	28
2.4.4 IID Cauchy . . . . .	29
2.4.5 IID binormal mixture model . . . . .	31
2.5 Exercises . . . . .	33
<b>Part II PRAGMATICS</b>	<b>37</b>
<b>3 Hypothesis tests and confidence intervals or regions</b>	<b>39</b>
3.1 Introduction . . . . .	39

3.2	Approximate normality of MLEs . . . . .	40
3.2.1	Estimating the variance of $\hat{\theta}$ . . . . .	41
3.3	Wald tests, confidence intervals and regions . . . . .	43
3.3.1	Test for a single parameter . . . . .	43
3.3.2	Test of a function of the parameters . . . . .	43
3.3.3	Joint test of two or more parameters . . . . .	44
3.3.4	In R and SAS: Old Faithful revisited . . . . .	45
3.4	Likelihood ratio tests, confidence intervals and regions . . . . .	49
3.4.1	Using R and SAS: Another visit to Old Faithful . . . . .	51
3.5	Likelihood ratio examples . . . . .	54
3.5.1	LR inference from a two-dimensional contour plot . . . . .	54
3.5.2	The $G$ -test for contingency tables . . . . .	56
3.6	Profile likelihood . . . . .	57
3.6.1	Profile likelihood for Old Faithful . . . . .	58
3.7	Exercises . . . . .	59
<b>4</b>	<b>What you really need to know</b>	<b>64</b>
4.1	Introduction . . . . .	64
4.2	Inference about $g(\theta)$ . . . . .	65
4.2.1	The delta method . . . . .	65
4.2.2	The delta method applied to MLEs . . . . .	67
4.2.3	The delta method using R, SAS and ADMB . . . . .	70
4.2.4	Delta method examples . . . . .	72
4.3	Wald statistics – quick and dirty? . . . . .	75
4.3.1	Wald versus likelihood ratio revisited . . . . .	77
4.3.2	Pragmatic considerations . . . . .	78
4.4	Model selection . . . . .	79
4.4.1	AIC . . . . .	79
4.5	Bootstrapping . . . . .	81
4.5.1	Bootstrap simulation . . . . .	82
4.5.2	Bootstrap confidence intervals . . . . .	83
4.5.3	Bootstrap estimate of variance . . . . .	85
4.5.4	Bootstrapping test statistics . . . . .	85
4.5.5	Bootstrap pragmatics . . . . .	86
4.5.6	Bootstrapping Old Faithful . . . . .	86
4.5.7	How many bootstrap simulations is enough? . . . . .	90
4.6	Prediction . . . . .	91
4.6.1	The plug-in approach . . . . .	92
4.6.2	Predictive likelihood . . . . .	93
4.6.3	Bayesian prediction . . . . .	93
4.6.4	Pseudo-Bayesian prediction . . . . .	94
4.6.5	Bootstrap prediction . . . . .	94
4.7	Things that can mess you up . . . . .	95
4.7.1	Multiple maxima of the likelihood . . . . .	95
4.7.2	Lack of convergence . . . . .	96

4.7.3	Parameters on the boundary of the parameter space . . . . .	96
4.7.4	Insufficient sample size . . . . .	98
4.8	Exercises . . . . .	98
<b>5</b>	<b>Maximizing the likelihood</b>	<b>101</b>
5.1	Introduction . . . . .	101
5.2	The Newton-Raphson algorithm . . . . .	103
5.3	The EM (Expectation–Maximization) algorithm . . . . .	104
5.3.1	The simple form of the EM algorithm . . . . .	105
5.3.2	Properties of the EM algorithm . . . . .	107
5.3.3	Accelerating the EM algorithm . . . . .	111
5.3.4	Inference . . . . .	112
5.4	Multi-stage maximization . . . . .	113
5.4.1	Efficient maximization via profile likelihood . . . . .	114
5.4.2	Multi-stage optimization . . . . .	116
5.5	Exercises . . . . .	118
<b>6</b>	<b>Some widely used applications of maximum likelihood</b>	<b>121</b>
6.1	Introduction . . . . .	121
6.2	Box-Cox transformations . . . . .	122
6.2.1	Example: the Box and Cox poison data . . . . .	123
6.3	Models for survival-time data . . . . .	125
6.3.1	Notation . . . . .	126
6.3.2	Accelerated failure-time model . . . . .	127
6.3.3	Parametric proportional hazards model . . . . .	128
6.3.4	Cox’s proportional hazards model . . . . .	130
6.3.5	Example in R and SAS: Leukaemia data . . . . .	132
6.4	Mark–recapture models . . . . .	134
6.4.1	Hypergeometric likelihood for integer valued $N$ . . . . .	136
6.4.2	Hypergeometric likelihood for $N \in \mathbb{R}^+$ . . . . .	137
6.4.3	Multinomial likelihood . . . . .	138
6.4.4	Closing remarks . . . . .	140
6.5	Exercises . . . . .	141
<b>7</b>	<b>Generalized linear models and extensions</b>	<b>143</b>
7.1	Introduction . . . . .	143
7.2	Specification of a GLM . . . . .	144
7.2.1	Exponential family distribution . . . . .	144
7.2.2	GLM formulation . . . . .	146
7.3	Likelihood calculations . . . . .	148
7.4	Model evaluation . . . . .	149
7.4.1	Deviance . . . . .	149
7.4.2	Model selection . . . . .	151
7.4.3	Residuals . . . . .	151
7.4.4	Goodness of fit . . . . .	152

7.5	Case study 1: Logistic regression and inverse prediction in R . . . .	154
7.5.1	Size-selectivity modelling in R . . . . .	155
7.6	Beyond binomial and Poisson models . . . . .	161
7.6.1	Quasi-likelihood and quasi-AIC . . . . .	162
7.6.2	Zero inflation and the negative binomial . . . . .	164
7.7	Case study 2: Multiplicative vs additive models of over-dispersed counts in SAS . . . . .	167
7.7.1	Background . . . . .	167
7.7.2	Poisson and quasi-Poisson fits . . . . .	169
7.7.3	Negative binomial fits . . . . .	171
7.8	Exercises . . . . .	173
<b>8</b>	<b>Quasi-likelihood and generalized estimating equations</b>	<b>175</b>
8.1	Introduction . . . . .	175
8.2	Wedderburn's quasi-likelihood . . . . .	177
8.2.1	Quasi-likelihood analysis of barley blotch data in R . . . .	177
8.3	Generalized estimating equations . . . . .	181
8.3.1	GEE analysis of multi-centre data in SAS . . . . .	183
8.4	Exercises . . . . .	187
<b>9</b>	<b>ML inference in the presence of incidental parameters</b>	<b>188</b>
9.1	Introduction . . . . .	188
9.1.1	Analysis of paired data: an intuitive use of conditional likelihood . . . . .	190
9.2	Conditional likelihood . . . . .	192
9.2.1	Restricted maximum likelihood . . . . .	197
9.3	Integrated likelihood . . . . .	198
9.3.1	Justification . . . . .	199
9.3.2	Uses of integrated likelihood . . . . .	200
9.4	Exercises . . . . .	201
<b>10</b>	<b>Latent variable models</b>	<b>202</b>
10.1	Introduction . . . . .	202
10.2	Developing the likelihood . . . . .	203
10.3	Software . . . . .	204
10.3.1	Background . . . . .	204
10.3.2	The Laplace approximation and Gauss-Hermite quadrature	206
10.3.3	Importance sampling . . . . .	208
10.3.4	Separability . . . . .	209
10.3.5	Overview of examples . . . . .	210
10.4	One-way linear random-effects model . . . . .	210
10.4.1	SAS . . . . .	212
10.4.2	R . . . . .	215
10.4.3	ADMB . . . . .	216
10.5	Nonlinear mixed-effects model . . . . .	217

10.5.1	SAS . . . . .	219
10.5.2	ADMB . . . . .	220
10.6	Generalized linear mixed-effects model . . . . .	221
10.6.1	R . . . . .	222
10.6.2	SAS . . . . .	224
10.6.3	ADMB . . . . .	224
10.6.4	GLMM vs GEE . . . . .	225
10.7	State-space model for count data . . . . .	227
10.8	ADMB template files . . . . .	228
10.8.1	One-way linear random-effects model using REML . . . . .	228
10.8.2	Nonlinear crossed mixed-effects model . . . . .	229
10.8.3	Generalized linear mixed model using GREML . . . . .	230
10.8.4	State-space model for count data . . . . .	231
10.9	Exercises . . . . .	232

## **Part III THEORETICAL FOUNDATIONS 233**

### **11 Cramér-Rao inequality and Fisher information 235**

11.1	Introduction . . . . .	235
11.1.1	Notation . . . . .	236
11.2	The Cramér-Rao inequality for $\theta \in \mathbb{R}$ . . . . .	236
11.3	Cramér-Rao inequality for functions of $\theta$ . . . . .	239
11.4	Alternative formulae for $\mathcal{I}(\theta)$ . . . . .	241
11.5	The iid data case . . . . .	243
11.6	The multi-dimensional case, $\theta \in \Theta \subset \mathbb{R}^s$ . . . . .	243
11.6.1	Parameter orthogonality . . . . .	244
11.6.2	Alternative formulae for $\mathcal{I}(\theta)$ . . . . .	245
11.6.3	Fisher information for re-parameterized models . . . . .	246
11.7	Examples of Fisher information calculation . . . . .	247
11.7.1	Normal( $\mu, \sigma^2$ ) . . . . .	247
11.7.2	Exponential family distributions . . . . .	249
11.7.3	Linear regression model . . . . .	251
11.7.4	Nonlinear regression model . . . . .	252
11.7.5	Generalized linear model with canonical link function . . . . .	252
11.7.6	Gamma( $\alpha, \beta$ ) . . . . .	252
11.8	Exercises . . . . .	253

### **12 Asymptotic theory and approximate normality 256**

12.1	Introduction . . . . .	256
12.2	Consistency and asymptotic normality . . . . .	257
12.2.1	Asymptotic normality, $\theta \in \mathbb{R}$ . . . . .	261
12.2.2	Asymptotic normality: $\theta \in \mathbb{R}^s$ . . . . .	264
12.2.3	Asymptotic normality of $g(\theta) \in \mathbb{R}^p$ . . . . .	266
12.2.4	Asymptotic normality under model misspecification . . . . .	267

12.2.5	Asymptotic normality of M-estimators . . . . .	268
12.2.6	The non-iid case . . . . .	271
12.3	Approximate normality . . . . .	271
12.3.1	Estimation of the approximate variance . . . . .	273
12.3.2	Approximate normality of M-estimators . . . . .	274
12.4	Wald tests and confidence regions . . . . .	276
12.4.1	Wald test statistics . . . . .	276
12.4.2	Wald confidence intervals and regions . . . . .	278
12.5	Likelihood ratio test statistic . . . . .	280
12.5.1	Likelihood ratio test: $\theta \in \mathbb{R}$ . . . . .	280
12.5.2	Likelihood ratio test for $\theta \in \mathbb{R}^s$ , and $g(\theta) \in \mathbb{R}^p$ . . . . .	281
12.6	Rao-score test statistic . . . . .	281
12.7	Exercises . . . . .	283
<b>13</b>	<b>Tools of the trade</b>	<b>286</b>
13.1	Introduction . . . . .	286
13.2	Equivalence of tests and confidence intervals . . . . .	286
13.3	Transformation of variables . . . . .	287
13.4	Mean and variance conditional identities . . . . .	288
13.5	Relevant inequalities . . . . .	289
13.5.1	Jensen's inequality for convex functions . . . . .	289
13.5.2	Cauchy-Schwarz inequality . . . . .	291
13.6	Asymptotic probability theory . . . . .	291
13.6.1	Convergence in distribution and probability . . . . .	292
13.6.2	Properties . . . . .	294
13.6.3	Slutsky's theorem . . . . .	295
13.6.4	Delta theorem . . . . .	295
13.7	Exercises . . . . .	297
<b>14</b>	<b>Fundamental paradigms and principles of inference</b>	<b>299</b>
14.1	Introduction . . . . .	299
14.2	Sufficiency principle . . . . .	300
14.2.1	Finding sufficient statistics . . . . .	301
14.2.2	Examples of the sufficiency principle . . . . .	303
14.3	Conditionality principle . . . . .	304
14.4	The likelihood principle . . . . .	306
14.4.1	Relationship with sufficiency and conditionality . . . . .	308
14.5	Statistical significance versus statistical evidence . . . . .	309
14.6	Exercises . . . . .	311
<b>15</b>	<b>Miscellanea</b>	<b>313</b>
15.1	Notation . . . . .	313
15.2	Acronyms . . . . .	315
15.3	Do you think like a frequentist or a Bayesian? . . . . .	315
15.4	Some useful distributions . . . . .	316

15.4.1	Discrete distributions . . . . .	316
15.4.2	Continuous distributions . . . . .	318
15.5	Software extras . . . . .	321
15.5.1	R function <b>Plkhci</b> for likelihood ratio confidence intervals . . . . .	321
15.5.2	R function <b>Profile</b> for calculation of profile likelihoods . . . . .	322
15.5.3	SAS macro <b>Plkhci</b> for likelihood ratio confidence intervals . . . . .	322
15.5.4	SAS macro <b>Profile</b> for calculation of profile likelihoods . . . . .	323
15.5.5	SAS macro <b>DeltaMethod</b> for application of the delta method . . . . .	323
15.6	Automatic differentiation . . . . .	323
	<b>Appendix: Partial solutions to selected exercises</b>	<b>325</b>
	<b>Bibliography</b>	<b>337</b>
	<b>Index</b>	<b>345</b>