



**MODEL SELECTION AND MULTIMODEL INFERENCE**  
*A Practical Information-Theoretic Approach*

SECOND EDITION

KENNETH P. BURNHAM • DAVID R. ANDERSON



# Contents

<b>Preface</b>	<b>vii</b>
<b>About the Authors</b>	<b>xxi</b>
<b>Glossary</b>	<b>xxiii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Objectives of the Book . . . . .	1
1.2 Background Material . . . . .	5
1.2.1 Inference from Data, Given a Model . . . . .	5
1.2.2 Likelihood and Least Squares Theory . . . . .	6
1.2.3 The Critical Issue: “What Is the Best Model to Use?” . . . . .	13
1.2.4 Science Inputs: Formulation of the Set of Candidate Models . . . . .	15
1.2.5 Models Versus Full Reality . . . . .	20
1.2.6 An Ideal Approximating Model . . . . .	22
1.3 Model Fundamentals and Notation . . . . .	23
1.3.1 Truth or Full Reality $f$ . . . . .	23
1.3.2 Approximating Models $g_i(x \theta)$ . . . . .	23
1.3.3 The Kullback–Leibler Best Model $g_i(x \theta_0)$ . . . . .	25
1.3.4 Estimated Models $g_i(x \hat{\theta})$ . . . . .	25
1.3.5 Generating Models . . . . .	26
1.3.6 Global Model . . . . .	26

1.3.7	Overview of Stochastic Models in the Biological Sciences . . . . .	27
1.4	Inference and the Principle of Parsimony . . . . .	29
1.4.1	Avoid Overfitting to Achieve a Good Model Fit . . . . .	29
1.4.2	The Principle of Parsimony . . . . .	31
1.4.3	Model Selection Methods . . . . .	35
1.5	Data Dredging, Overanalysis of Data, and Spurious Effects . . . . .	37
1.5.1	Overanalysis of Data . . . . .	38
1.5.2	Some Trends . . . . .	40
1.6	Model Selection Bias . . . . .	43
1.7	Model Selection Uncertainty . . . . .	45
1.8	Summary . . . . .	47

**2 Information and Likelihood Theory: A Basis for Model Selection and Inference 49**

2.1	Kullback–Leibler Information or Distance Between Two Models . . . . .	50
2.1.1	Examples of Kullback–Leibler Distance . . . . .	54
2.1.2	Truth, $f$ , Drops Out as a Constant . . . . .	58
2.2	Akaike’s Information Criterion: 1973 . . . . .	60
2.3	Takeuchi’s Information Criterion: 1976 . . . . .	65
2.4	Second-Order Information Criterion: 1978 . . . . .	66
2.5	Modification of Information Criterion for Overdispersed Count Data . . . . .	67
2.6	AIC Differences, $\Delta_i$ . . . . .	70
2.7	A Useful Analogy . . . . .	72
2.8	Likelihood of a Model, $\mathcal{L}(g_i data)$ . . . . .	74
2.9	Akaike Weights, $w_i$ . . . . .	75
2.9.1	Basic Formula . . . . .	75
2.9.2	An Extension . . . . .	76
2.10	Evidence Ratios . . . . .	77
2.11	Important Analysis Details . . . . .	80
2.11.1	AIC Cannot Be Used to Compare Models of Different Data Sets . . . . .	80
2.11.2	Order Not Important in Computing AIC Values . . . . .	81
2.11.3	Transformations of the Response Variable . . . . .	81
2.11.4	Regression Models with Differing Error Structures . . . . .	82
2.11.5	Do Not Mix Null Hypothesis Testing with Information-Theoretic Criteria . . . . .	83
2.11.6	Null Hypothesis Testing Is Still Important in Strict Experiments . . . . .	83
2.11.7	Information-Theoretic Criteria Are Not a “Test” . . . . .	84
2.11.8	Exploratory Data Analysis . . . . .	84

2.12	Some History and Further Insights . . . . .	85
2.12.1	Entropy . . . . .	86
2.12.2	A Heuristic Interpretation . . . . .	87
2.12.3	More on Interpreting Information-Theoretic Criteria . . . . .	87
2.12.4	Nonnested Models . . . . .	88
2.12.5	Further Insights . . . . .	89
2.13	Bootstrap Methods and Model Selection Frequencies $\pi_i$ . . . . .	90
2.13.1	Introduction . . . . .	91
2.13.2	The Bootstrap in Model Selection: The Basic Idea . . . . .	93
2.14	Return to Flather's Models . . . . .	94
2.15	Summary . . . . .	96
<b>3</b>	<b>Basic Use of the Information-Theoretic Approach</b>	<b>98</b>
3.1	Introduction . . . . .	98
3.2	Example 1: Cement Hardening Data . . . . .	100
3.2.1	Set of Candidate Models . . . . .	101
3.2.2	Some Results and Comparisons . . . . .	102
3.2.3	A Summary . . . . .	106
3.3	Example 2: Time Distribution of an Insecticide Added to a Simulated Ecosystem . . . . .	106
3.3.1	Set of Candidate Models . . . . .	108
3.3.2	Some Results . . . . .	110
3.4	Example 3: Nestling Starlings . . . . .	111
3.4.1	Experimental Scenario . . . . .	112
3.4.2	Monte Carlo Data . . . . .	113
3.4.3	Set of Candidate Models . . . . .	113
3.4.4	Data Analysis Results . . . . .	117
3.4.5	Further Insights into the First Fourteen Nested Models . . . . .	120
3.4.6	Hypothesis Testing and Information-Theoretic Approaches Have Different Selection Frequencies . . . . .	121
3.4.7	Further Insights Following Final Model Selection . . . . .	124
3.4.8	Why Not Always Use the Global Model for Inference? . . . . .	125
3.5	Example 4: Sage Grouse Survival . . . . .	126
3.5.1	Introduction . . . . .	126
3.5.2	Set of Candidate Models . . . . .	127
3.5.3	Model Selection . . . . .	129
3.5.4	Hypothesis Tests for Year-Dependent Survival Probabilities . . . . .	131

3.5.5	Hypothesis Testing Versus AIC in Model Selection . . . . .	132
3.5.6	A Class of Intermediate Models . . . . .	134
3.6	Example 5: Resource Utilization of <i>Anolis</i> Lizards . . . . .	137
3.6.1	Set of Candidate Models . . . . .	138
3.6.2	Comments on Analytic Method . . . . .	138
3.6.3	Some Tentative Results . . . . .	139
3.7	Example 6: Sakamoto et al.'s (1986) Simulated Data . . . . .	141
3.8	Example 7: Models of Fish Growth . . . . .	142
3.9	Summary . . . . .	143
<b>4</b>	<b>Formal Inference From More Than One Model: Multimodel Inference (MMI)</b>	<b>149</b>
4.1	Introduction to Multimodel Inference . . . . .	149
4.2	Model Averaging . . . . .	150
4.2.1	Prediction . . . . .	150
4.2.2	Averaging Across Model Parameters . . . . .	151
4.3	Model Selection Uncertainty . . . . .	153
4.3.1	Concepts of Parameter Estimation and Model Selection Uncertainty . . . . .	155
4.3.2	Including Model Selection Uncertainty in Estimator Sampling Variance . . . . .	158
4.3.3	Unconditional Confidence Intervals . . . . .	164
4.4	Estimating the Relative Importance of Variables . . . . .	167
4.5	Confidence Set for the K-L Best Model . . . . .	169
4.5.1	Introduction . . . . .	169
4.5.2	$\Delta_i$ , Model Selection Probabilities, and the Bootstrap . . . . .	171
4.6	Model Redundancy . . . . .	173
4.7	Recommendations . . . . .	176
4.8	Cement Data . . . . .	177
4.9	Pine Wood Data . . . . .	183
4.10	The Durban Storm Data . . . . .	187
4.10.1	Models Considered . . . . .	188
4.10.2	Consideration of Model Fit . . . . .	190
4.10.3	Confidence Intervals on Predicted Storm Probability . . . . .	191
4.10.4	Comparisons of Estimator Precision . . . . .	193
4.11	Flour Beetle Mortality: A Logistic Regression Example . . . . .	195
4.12	Publication of Research Results . . . . .	201
4.13	Summary . . . . .	203
<b>5</b>	<b>Monte Carlo Insights and Extended Examples</b>	<b>206</b>
5.1	Introduction . . . . .	206
5.2	Survival Models . . . . .	207

5.2.1	A Chain Binomial Survival Model . . . . .	207
5.2.2	An Example . . . . .	210
5.2.3	An Extended Survival Model . . . . .	215
5.2.4	Model Selection if Sample Size Is Huge, or Truth Known . . . . .	219
5.2.5	A Further Chain Binomial Model . . . . .	221
5.3	Examples and Ideas Illustrated with Linear Regression . . . . .	224
5.3.1	All-Subsets Selection: A GPA Example . . . . .	225
5.3.2	A Monte Carlo Extension of the GPA Example . . . . .	229
5.3.3	An Improved Set of GPA Prediction Models . . . . .	235
5.3.4	More Monte Carlo Results . . . . .	238
5.3.5	Linear Regression and Variable Selection . . . . .	244
5.3.6	Discussion . . . . .	248
5.4	Estimation of Density from Line Transect Sampling . . . . .	255
5.4.1	Density Estimation Background . . . . .	255
5.4.2	Line Transect Sampling of Kangaroos at Wallaby Creek . . . . .	256
5.4.3	Analysis of Wallaby Creek Data . . . . .	256
5.4.4	Bootstrap Analysis . . . . .	258
5.4.5	Confidence Interval on $D$ . . . . .	258
5.4.6	Bootstrap Samples: 1,000 Versus 10,000 . . . . .	260
5.4.7	Bootstrap Versus Akaike Weights: A Lesson on QAIC <sub>c</sub> . . . . .	261
5.5	Summary . . . . .	264
<b>6</b>	<b>Advanced Issues and Deeper Insights</b>	<b>267</b>
6.1	Introduction . . . . .	267
6.2	An Example with 13 Predictor Variables and 8,191 Models . . . . .	268
6.2.1	Body Fat Data . . . . .	268
6.2.2	The Global Model . . . . .	269
6.2.3	Classical Stepwise Selection . . . . .	269
6.2.4	Model Selection Uncertainty for AIC <sub>c</sub> and BIC . . . . .	271
6.2.5	An A Priori Approach . . . . .	274
6.2.6	Bootstrap Evaluation of Model Uncertainty . . . . .	276
6.2.7	Monte Carlo Simulations . . . . .	279
6.2.8	Summary Messages . . . . .	281
6.3	Overview of Model Selection Criteria . . . . .	284
6.3.1	Criteria That Are Estimates of K-L Information . . . . .	284
6.3.2	Criteria That Are Consistent for $K$ . . . . .	286
6.3.3	Contrasts . . . . .	288
6.3.4	Consistent Selection in Practice: Quasi-true Models . . . . .	289
6.4	Contrasting AIC and BIC . . . . .	293
6.4.1	A Heuristic Derivation of BIC . . . . .	293

6.4.2	A K-L-Based Conceptual Comparison of AIC and BIC . . . . .	295
6.4.3	Performance Comparison . . . . .	298
6.4.4	Exact Bayesian Model Selection Formulas . . . . .	301
6.4.5	Akaike Weights as Bayesian Posterior Model Probabilities . . . . .	302
6.5	Goodness-of-Fit and Overdispersion Revisited . . . . .	305
6.5.1	Overdispersion $\hat{c}$ and Goodness-of-Fit: A General Strategy . . . . .	305
6.5.2	Overdispersion Modeling: More Than One $\hat{c}$ . . . . .	307
6.5.3	Model Goodness-of-Fit After Selection . . . . .	309
6.6	AIC and Random Coefficient Models . . . . .	310
6.6.1	Basic Concepts and Marginal Likelihood Approach . . . . .	310
6.6.2	A Shrinkage Approach to AIC and Random Effects . . . . .	313
6.6.3	On Extensions . . . . .	316
6.7	Selection When Probability Distributions Differ by Model . . . . .	317
6.7.1	Keep All the Parts . . . . .	317
6.7.2	A Normal Versus Log-Normal Example . . . . .	318
6.7.3	Comparing Across Several Distributions: An Example . . . . .	320
6.8	Lessons from the Literature and Other Matters . . . . .	323
6.8.1	Use $AIC_c$ , Not AIC, with Small Sample Sizes . . . . .	323
6.8.2	Use $AIC_c$ , Not AIC, When $K$ Is Large . . . . .	325
6.8.3	When Is $AIC_c$ Suitable: A Gamma Distribution Example . . . . .	326
6.8.4	Inference from a Less Than Best Model . . . . .	328
6.8.5	Are Parameters Real? . . . . .	330
6.8.6	Sample Size Is Often Not a Simple Issue . . . . .	332
6.8.7	Judgment Has a Role . . . . .	333
6.9	Tidbits About AIC . . . . .	334
6.9.1	Irrelevance of Between-Sample Variation of AIC . . . . .	334
6.9.2	The G-Statistic and K-L Information . . . . .	336
6.9.3	AIC Versus Hypothesis Testing: Results Can Be Very Different . . . . .	337
6.9.4	A Subtle Model Selection Bias Issue . . . . .	339
6.9.5	The Dimensional Unit of AIC . . . . .	340
6.9.6	AIC and Finite Mixture Models . . . . .	342
6.9.7	Unconditional Variance . . . . .	344
6.9.8	A Baseline for $w_+(i)$ . . . . .	345
6.10	Summary . . . . .	347

<b>7</b>	<b>Statistical Theory and Numerical Results</b>	<b>352</b>
7.1	Useful Preliminaries . . . . .	352
7.2	A General Derivation of AIC . . . . .	362
7.3	General K-L–Based Model Selection: TIC . . . . .	371
7.3.1	Analytical Computation of TIC . . . . .	371
7.3.2	Bootstrap Estimation of TIC . . . . .	372
7.4	AIC <sub>c</sub> : A Second-Order Improvement . . . . .	374
7.4.1	Derivation of AIC <sub>c</sub> . . . . .	374
7.4.2	Lack of Uniqueness of AIC <sub>c</sub> . . . . .	379
7.5	Derivation of AIC for the Exponential Family of Distributions . . . . .	380
7.6	Evaluation of $\text{tr}(J(\underline{\theta}_o)[I(\underline{\theta}_o)]^{-1})$ and Its Estimator . . . . .	384
7.6.1	Comparison of AIC Versus TIC in a Very Simple Setting . . . . .	385
7.6.2	Evaluation Under Logistic Regression . . . . .	390
7.6.3	Evaluation Under Multinomially Distributed Count Data . . . . .	397
7.6.4	Evaluation Under Poisson-Distributed Data . . . . .	405
7.6.5	Evaluation for Fixed-Effects Normality-Based Linear Models . . . . .	406
7.7	Additional Results and Considerations . . . . .	412
7.7.1	Selection Simulation for Nested Models . . . . .	412
7.7.2	Simulation of the Distribution of $\Delta_p$ . . . . .	415
7.7.3	Does AIC Overfit? . . . . .	417
7.7.4	Can Selection Be Improved Based on All the $\Delta_i$ ? . . . . .	419
7.7.5	Linear Regression, AIC, and Mean Square Error . . . . .	421
7.7.6	AIC <sub>c</sub> and Models for Multivariate Data . . . . .	424
7.7.7	There Is No True TIC <sub>c</sub> . . . . .	426
7.7.8	Kullback–Leibler Information Relationship to the Fisher Information Matrix . . . . .	426
7.7.9	Entropy and Jaynes Maxent Principle . . . . .	427
7.7.10	Akaike Weights $w_i$ Versus Selection Probabilities $\pi_i$ . . . . .	428
7.8	Kullback–Leibler Information Is Always $\geq 0$ . . . . .	429
7.9	Summary . . . . .	434
<b>8</b>	<b>Summary</b>	<b>437</b>
8.1	The Scientific Question and the Collection of Data . . . . .	439
8.2	Actual Thinking and A Priori Modeling . . . . .	440
8.3	The Basis for Objective Model Selection . . . . .	442
8.4	The Principle of Parsimony . . . . .	443
8.5	Information Criteria as Estimates of Expected Relative Kullback–Leibler Information . . . . .	444
8.6	Ranking Alternative Models . . . . .	446



8.7	Scaling Alternative Models . . . . .	447
8.8	MMI: Inference Based on Model Averaging . . . . .	448
8.9	MMI: Model Selection Uncertainty . . . . .	449
8.10	MMI: Relative Importance of Predictor Variables . . . . .	451
8.11	More on Inferences . . . . .	451
8.12	Final Thoughts . . . . .	454
	<b>References</b>	<b>455</b>
	<b>Index</b>	<b>485</b>