
Contents

<i>Preface</i>	<i>page xi</i>
<i>A guide to notation</i>	<i>xiv</i>
1 Model selection: data examples and introduction	1
1.1 Introduction	1
1.2 Egyptian skull development	3
1.3 Who wrote ‘The Quiet Don’?	7
1.4 Survival data on primary biliary cirrhosis	10
1.5 Low birthweight data	13
1.6 Football match prediction	15
1.7 Speedskating	17
1.8 Preview of the following chapters	19
1.9 Notes on the literature	20
2 Akaike’s information criterion	22
2.1 Information criteria for balancing fit with complexity	22
2.2 Maximum likelihood and the Kullback–Leibler distance	23
2.3 AIC and the Kullback–Leibler distance	28
2.4 Examples and illustrations	32
2.5 Takeuchi’s model-robust information criterion	43
2.6 Corrected AIC for linear regression and autoregressive time series	44
2.7 AIC, corrected AIC and bootstrap-AIC for generalised linear models*	46
2.8 Behaviour of AIC for moderately misspecified models*	49
2.9 Cross-validation	51
2.10 Outlier-robust methods	55
2.11 Notes on the literature	64
Exercises	66

3	The Bayesian information criterion	70
3.1	Examples and illustrations of the BIC	70
3.2	Derivation of the BIC	78
3.3	Who wrote ‘The Quiet Don’?	82
3.4	The BIC and AIC for hazard regression models	85
3.5	The deviance information criterion	90
3.6	Minimum description length	94
3.7	Notes on the literature	96
	Exercises	97
4	A comparison of some selection methods	99
4.1	Comparing selectors: consistency, efficiency and parsimony	99
4.2	Prototype example: choosing between two normal models	102
4.3	Strong consistency and the Hannan–Quinn criterion	106
4.4	Mallows’s C_p and its outlier-robust versions	107
4.5	Efficiency of a criterion	108
4.6	Efficient order selection in an autoregressive process and the FPE	110
4.7	Efficient selection of regression variables	111
4.8	Rates of convergence*	112
4.9	Taking the best of both worlds?*	113
4.10	Notes on the literature	114
	Exercises	115
5	Bigger is not always better	117
5.1	Some concrete examples	117
5.2	Large-sample framework for the problem	119
5.3	A precise tolerance limit	124
5.4	Tolerance regions around parametric models	126
5.5	Computing tolerance thresholds and radii	128
5.6	How the 5000-m time influences the 10,000-m time	130
5.7	Large-sample calculus for AIC	137
5.8	Notes on the literature	140
	Exercises	140
6	The focussed information criterion	145
6.1	Estimators and notation in submodels	145
6.2	The focussed information criterion, FIC	146
6.3	Limit distributions and mean squared errors in submodels	148
6.4	A bias-modified FIC	150
6.5	Calculation of the FIC	153
6.6	Illustrations and applications	154
6.7	Exact mean squared error calculations for linear regression*	172

<i>Contents</i>		ix
6.8	The FIC for Cox proportional hazard regression models	174
6.9	Average-FIC	179
6.10	A Bayesian focussed information criterion*	183
6.11	Notes on the literature	188
	Exercises	189
7	Frequentist and Bayesian model averaging	192
7.1	Estimators-post-selection	192
7.2	Smooth AIC, smooth BIC and smooth FIC weights	193
7.3	Distribution of model average estimators	195
7.4	What goes wrong when we ignore model selection?	199
7.5	Better confidence intervals	206
7.6	Shrinkage, ridge estimation and thresholding	211
7.7	Bayesian model averaging	216
7.8	A frequentist view of Bayesian model averaging*	220
7.9	Bayesian model selection with canonical normal priors*	223
7.10	Notes on the literature	224
	Exercises	225
8	Lack-of-fit and goodness-of-fit tests	227
8.1	The principle of order selection	227
8.2	Asymptotic distribution of the order selection test	229
8.3	The probability of overfitting*	232
8.4	Score-based tests	236
8.5	Two or more covariates	238
8.6	Neyman's smooth tests and generalisations	240
8.7	A comparison between AIC and the BIC for model testing*	242
8.8	Goodness-of-fit monitoring processes for regression models*	243
8.9	Notes on the literature	245
	Exercises	246
9	Model selection and averaging schemes in action	248
9.1	AIC and BIC selection for Egyptian skull development data	248
9.2	Low birthweight data: FIC plots and FIC selection per stratum	252
9.3	Survival data on PBC: FIC plots and FIC selection	256
9.4	Speedskating data: averaging over covariance structure models	259
	Exercises	266
10	Further topics	269
10.1	Model selection in mixed models	269
10.2	Boundary parameters	273
10.3	Finite-sample corrections*	281

10.4	Model selection with missing data	282
10.5	When p and q grow with n	284
10.6	Notes on the literature	285
	<i>Overview of data examples</i>	287
	<i>References</i>	293
	<i>Author index</i>	306
	<i>Subject index</i>	310