

2. Information	57
2.1 Shannon Information	57
2.1.1 Binary Codes	59
2.1.2 Optimal Codes	63
2.1.3 Measurement of Information	66
2.1.4 The Construction of Optimal Codes	69
2.1.5 Coding Multi-Word Messages	72
2.1.6 Arithmetic Coding	73
2.1.7 Some Properties of Optimal Codes	76
2.1.8 Non-Binary Codes: The Nit	77
2.1.9 The Subjective Nature of Information	79
2.1.10 The Information Content of a Multinomial Distribution	81
2.1.11 Entropy	87
2.1.12 Codes for Infinite Sets	91
2.1.13 Unary and Punctuated Binary Codes	92
2.1.14 Optimal Codes for Integers	93
2.1.15 Feasible Codes for Infinite Sets	96
2.1.16 Universal Codes	98
2.2 Algorithmic Complexity	100
2.2.1 Turing Machines	101
2.2.2 Start and Stop Conditions	102
2.2.3 Dependence on the Choice of Turing Machine	103
2.2.4 Turing Probability Distributions	103
2.2.5 Universal Turing Machines	105
2.2.6 Algorithmic Complexity vs. Shannon Information	107
2.3 Information, Inference and Explanation	110
2.3.1 The Second Part of an Explanation	110
2.3.2 The First Part of an Explanation	112
2.3.3 Theory Description Codes as Priors	114
2.3.4 Universal Codes in Theory Descriptions	115
2.3.5 Relation to Bayesian Inference	116
2.3.6 Explanations and Algorithmic Complexity	118
2.3.7 The Second Part	120
2.3.8 The First Part	121
2.3.9 An Alternative Construction	123
2.3.10 Universal Turing Machines as Priors	124
2.3.11 Differences among UTMs	130
2.3.12 The Origins of Priors Revisited	133
2.3.13 The Evolution of Priors	135
3. Strict Minimum Message Length (SM)	
3.1 Problem Definition	
3.1.1 The Set X of Possible Data	
3.1.2 The Probabilistic Model of Data	
3.1.3 Coding of the Data	
3.1.4 The Set of Possible Inferences	
3.1.5 Coding the Inference $\hat{\theta}$	
3.1.6 Prior Probability Density	
3.1.7 Meaning of the Assertion	
3.2 The Strict Minimum Message Length Discrete Data	
3.2.1 Discrete Hypothesis Sets	
3.2.2 Minimizing Relations for SMM	
3.2.3 Binomial Example	
3.2.4 Significance of $I_1 - I_0$	
3.2.5 Non-Uniqueness of Θ^*	
3.2.6 Sufficient Statistics	
3.2.7 Binomial Example Using a Sum	
3.2.8 An Exact Algorithm for the Binomial	
3.2.9 A Solution for the Trinomial	
3.3 The SMM Explanation for Continuous Data	
3.3.1 Mean of a Normal	
3.3.2 A Boundary Rule for Growing Means	
3.3.3 Estimation of Normal Mean with Unknown Variance	
3.3.4 Mean of a Multivariate Normal Distribution	
3.3.5 Summary of Multivariate Mean Estimation	
3.3.6 Mean of a Uniform Distribution	
3.4 Some General Properties of SMM Estimation	
3.4.1 Property 1: Data Representativeness	
3.4.2 Property 2: Model Representativeness	
3.4.3 Property 3: Generality	
3.4.4 Property 4: Dependence on Sample Size	
3.4.5 Property 5: Efficiency	
3.4.6 Discrimination	
3.4.7 Example: Discrimination of a Coin	
3.5 Summary	
4. Approximations to SMM	
4.1 The "Ideal Group" (IG) Estimator	
4.1.1 SMM-like codes	
4.1.2 Ideal Data Groups	
4.1.3 The Estimator	
4.2 The Neyman-Scott Problem	
4.3 The Ideal Group Estimator for Neyman- Scott	
4.4 Other Estimators for Neyman-Scott	

4.5	Maximum Likelihood for Neyman-Scott	203
4.5.1	Marginal Maximum Likelihood	203
4.6	Kullback-Leibler Distance	204
4.7	Minimum Expected K-L Distance (MEKL)	205
4.8	Minimum Expected K-L Distance for Neyman-Scott	206
4.9	Blurred Images	208
4.10	Dowe's Approximation IID to the Message Length	209
4.10.1	Random Coding of Estimates	210
4.10.2	Choosing a Region in Θ	211
4.11	Partitions of the Hypothesis Space	213
4.12	The Meaning of Uncertainty Regions	215
4.12.1	Uncertainty via Limited Precision	216
4.12.2	Uncertainty via Dowe's IID Construction	216
4.12.3	What Uncertainty Is Described by a Region?	216
4.13	Summary	218
5.	MML: Quadratic Approximations to SMML	221
5.1	The MML Coding Scheme	222
5.1.1	Assumptions of the Quadratic MML Scheme	226
5.1.2	A Trap for the Unwary	227
5.2	Properties of the MML Estimator	228
5.2.1	An Alternative Expression for Fisher Information	228
5.2.2	Data Invariance and Sufficiency	229
5.2.3	Model Invariance	229
5.2.4	Efficiency	230
5.2.5	Multiple Parameters	232
5.2.6	MML Multi-Parameter Properties	234
5.2.7	The MML Message Length Formulae	235
5.2.8	Standard Formulae	235
5.2.9	Small-Sample Message Length	235
5.2.10	Curved-Prior Message Length	236
5.2.11	Singularities in the Prior	237
5.2.12	Large- D Message Length	237
5.2.13	Approximation Based on I_0	237
5.2.14	Precision of Estimate Spacing	238
5.3	Empirical Fisher Information	240
5.3.1	Formula IIA for Many Parameters	240
5.3.2	Irregular Likelihood Functions	242
5.3.3	Transformation of Empirical Fisher Information	243
5.3.4	A Safer? Empirical Approximation to Fisher Information	244
5.4	A Binomial Example	246
5.4.1	The Multinomial Distribution	247
5.4.2	Irregularities in the Binomial and Multinomial Distributions	248
5.5	Limitations
5.6	The Normal Distribution
5.6.1	Extension to the Neyman-Scott Problem
5.7	Negative Binomial Distribution
5.8	The Likelihood Principle
6.	MML Details in Some Interesting Cases
6.1	Geometric Constants
6.2	Conjugate Priors for the Normal Distribution
6.2.1	Conjugate Priors for the Multivariate Distribution
6.3	Normal Distribution with Perturbed Data
6.4	Normal Distribution with Coarse Data
6.5	von Mises-Fisher Distribution
6.5.1	Circular von Mises-Fisher distribution
6.5.2	Spherical von Mises-Fisher Distribution
6.6	Poisson Distribution
6.7	Linear Regression and Function Approximation
6.7.1	Linear Regression
6.7.2	Function Approximation
6.8	Mixture Models
6.8.1	ML Mixture Estimation: The EM Algorithm
6.8.2	A Message Format for Mixtures
6.8.3	A Coding Trick
6.8.4	Imprecise Assertion of Discrete Parameters
6.8.5	The Code Length of Imprecise Disc
6.8.6	A Surrogate Class Label "Estimate"
6.8.7	The Fisher Information for Mixtures
6.8.8	The Fisher Information with Class I
6.8.9	Summary of the Classified Model
6.8.10	Classified vs. Unclassified Models
6.9	A "Latent Factor" Model
6.9.1	Multiple Latent Factors
7.	Structural Models
7.1	Inference of a Regular Grammar
7.1.1	A Mealey Machine Representation
7.1.2	Probabilistic FSMs
7.1.3	An Assertion Code for PFSMs
7.1.4	A Less Redundant FSM Code
7.1.5	Transparency and Redundancy
7.1.6	Coding Transitions
7.1.7	An Example
7.2	Classification Trees and Nets
7.2.1	A Decision Tree Explanation

7.2.2	Coding the Tree Structure	316
7.2.3	Coding the Class Distributions at the Leaves	317
7.2.4	Decision Graphs and Other Elaborations	318
7.3	A Binary Sequence Segmentation Problem	321
7.3.1	The Kearns <i>et al.</i> "MDL" Criterion	322
7.3.2	Correcting the Message Length	323
7.3.3	Results Using the MML Criterion	324
7.3.4	An SMMML Approximation to the Sequence Problem	325
7.4	Learning Causal Nets	326
7.4.1	The Model Space	327
7.4.2	The Message Format	328
7.4.3	Equivalence Sets	329
7.4.4	Insignificant Effects	329
7.4.5	Partial Order Equivalence	330
7.4.6	Structural Equivalence	330
7.4.7	Explanation Length	331
7.4.8	Finding Good Models	331
7.4.9	Prior Constraints	335
7.4.10	Test Results	335
8.	The Feathers on the Arrow of Time	337
8.1	Closed Systems and Their States	339
8.2	Reversible Laws	340
8.3	Entropy as a Measure of Disorder	341
8.4	Why Entropy Will Increase	343
8.5	A Paradox?	344
8.6	Deducing the Past	345
8.6.1	Macroscopic Deduction	345
8.6.2	Deduction with Deterministic Laws, Exact View	346
8.6.3	Deduction with Deterministic Laws, Inexact View	347
8.6.4	Deduction with Non-deterministic Laws	348
8.6.5	Alternative Priors	350
8.6.6	A Tale of Two Clocks	353
8.7	Records and Memories	355
8.8	Induction of the Past (<i>A la recherche du temps perdu</i>)	356
8.8.1	Induction of the Past by Maximum Likelihood	357
8.8.2	Induction of the Past by MML	358
8.8.3	The Uses of Deduction	361
8.8.4	The Inexplicable	362
8.8.5	Induction of the Past with Deterministic Laws	363
8.9	Causal and Teleological Explanations	365
8.10	Reasons for Asymmetry	367
8.11	Summary: The Past Regained?	369
8.12	Gas Simulations	370
8.12.1	Realism of the Simulation	370
8.12.2	Backtracking to the Past	370
8.12.3	Diatomic Molecules	370
8.12.4	The Past of a Computer Process	370
8.13	Addendum: Why Entropy Will Increase (A Simulation Details)	370
8.13.1	Simulation of the Past	370
8.13.2	A Non-Adiabatic Experiment	370
9.	MML as a Descriptive Theory	370
9.1	The Grand Theories	370
9.2	Primitive Inductive Inferences	370
9.3	The Hypotheses of Natural Languages	370
9.3.1	The Efficiencies of Natural Languages	370
9.4	Some Inefficiencies of Natural Languages	370
9.5	Scientific Languages	370
9.6	The Practice of MML Induction	370
9.6.1	Human Induction	370
9.6.2	Evolutionary Induction	370
9.6.3	Experiment	370
10.	Related Work	370
10.1	Solomonoff	370
10.1.1	Prediction with Generalized Scoring	370
10.1.2	Is Prediction Inductive?	370
10.1.3	A Final Quibble	370
10.2	Rissanen, MDL and NML	370
10.2.1	Normalized Maximum Likelihood	370
10.2.2	Has NML Any Advantage over MDL	370
	Bibliography	370
	Index	370