

---

# Contents

<b>0</b>	<b>Meta Contents</b>	iii
	Preface	v
	Contents	ix
	Tables, Figures, Theorems, ...	xv
	Notation	xvii
<b>1</b>	<b>A Short Tour Through the Book</b>	1
1.1	Introduction	2
1.2	Simplicity & Uncertainty	3
	1.2.1 Introduction	3
	1.2.2 Algorithmic Information Theory	4
	1.2.3 Uncertainty & Probabilities	5
	1.2.4 Algorithmic Probability & Universal Induction	6
	1.2.5 Generalized Universal (Semi)Measures	7
1.3	Universal Sequence Prediction	7
	1.3.1 Setup & Convergence	8
	1.3.2 Loss Bounds	8
	1.3.3 Optimality Properties	9
	1.3.4 Miscellaneous	10
1.4	Rational Agents in Known Probabilistic Environments	11
	1.4.1 The Agent Model	11
	1.4.2 Value Functions & Optimal Policies	11
	1.4.3 Sequential Decision Theory & Reinforcement Learning	12
1.5	The Universal Algorithmic Agent AIXI	13
	1.5.1 The Universal AIXI Model	13
	1.5.2 On the Optimality of AIXI	14
	1.5.3 Value-Related Optimality Results	15
	1.5.4 Markov Decision Processes	17
	1.5.5 The Choice of the Horizon	18
1.6	Important Environmental Classes	18
	1.6.1 Introduction	18
	1.6.2 Sequence Prediction (SP)	19
	1.6.3 Strategic Games (SG)	19
	1.6.4 Function Minimization (FM)	19
	1.6.5 Supervised Learning from Examples (EX)	19
	1.6.6 Other Aspects of Intelligence	20
1.7	Computational Aspects	20

1.7.1	The Fastest & Shortest Algorithm for All Problems . . . .	20
1.7.2	Time-Bounded AIXI Model . . . . .	22
1.8	Discussion . . . . .	24
1.9	History & References . . . . .	26
<b>2</b>	<b>Simplicity &amp; Uncertainty</b> . . . . .	<b>29</b>
2.1	Introduction . . . . .	30
2.1.1	Examples of Induction Problems . . . . .	30
2.1.2	Ockham, Epicurus, Hume, Bayes, Solomonoff . . . . .	31
2.1.3	Problem Setup . . . . .	32
2.2	Algorithmic Information Theory . . . . .	33
2.2.1	Definitions and Notation . . . . .	33
2.2.2	Turing Machines . . . . .	34
2.2.3	Kolmogorov Complexity . . . . .	36
2.2.4	Computability Concepts . . . . .	38
2.3	Uncertainty & Probabilities . . . . .	40
2.3.1	Frequency Interpretation: Counting . . . . .	40
2.3.2	Objective Interpretation: Uncertain Events . . . . .	41
2.3.3	Subjective Interpretation: Degrees of Belief . . . . .	43
2.3.4	Determining Priors . . . . .	45
2.4	Algorithmic Probability & Universal Induction . . . . .	45
2.4.1	The Universal Prior $M$ . . . . .	45
2.4.2	Universal Sequence Prediction . . . . .	47
2.4.3	Universal (Semi)Measures . . . . .	48
2.4.4	Martin-Löf Randomness . . . . .	54
2.5	History & References . . . . .	55
2.6	Problems . . . . .	60
<b>3</b>	<b>Universal Sequence Prediction</b> . . . . .	<b>65</b>
3.1	Introduction . . . . .	67
3.2	Setup and Convergence . . . . .	68
3.2.1	Random Sequences . . . . .	68
3.2.2	Universal Prior Probability Distribution . . . . .	69
3.2.3	Universal Posterior Probability Distribution . . . . .	70
3.2.4	Convergence of Random Sequences . . . . .	71
3.2.5	Distance Measures between Probability Distributions . .	72
3.2.6	Convergence of $\xi$ to $\mu$ . . . . .	74
3.2.7	Convergence in Martin-Löf Sense . . . . .	76
3.2.8	The Case where $\mu \notin \mathcal{M}$ . . . . .	80
3.2.9	Probability Classes $\mathcal{M}$ . . . . .	81
3.3	Error Bounds . . . . .	82
3.3.1	Bayes Optimal Predictors . . . . .	82
3.3.2	Total Expected Numbers of Errors . . . . .	82
3.3.3	Proof of Theorem 3.36 . . . . .	84
3.4	Loss Bounds . . . . .	86

3.4.1	Unit Loss Function .....	86
3.4.2	Loss Bound of Merhav & Feder .....	88
3.4.3	Example Loss Functions .....	89
3.4.4	Proof of Theorem 3.48 .....	89
3.4.5	Convergence of Instantaneous Losses .....	91
3.4.6	General Loss .....	92
3.5	Application to Games of Chance .....	93
3.5.1	Introduction .....	93
3.5.2	Games of Chance .....	94
3.5.3	Example .....	95
3.5.4	Information-Theoretic Interpretation .....	95
3.6	Optimality Properties .....	96
3.6.1	Lower Error Bound .....	96
3.6.2	Pareto Optimality of $\xi$ .....	99
3.6.3	Balanced Pareto Optimality of $\xi$ .....	101
3.6.4	On the Optimal Choice of Weights .....	102
3.6.5	Occam's razor versus No Free Lunches .....	103
3.7	Miscellaneous .....	103
3.7.1	Multistep Predictions .....	104
3.7.2	Continuous Probability Classes $\mathcal{M}$ .....	106
3.7.3	Further Applications .....	108
3.7.4	Prediction with Expert Advice .....	108
3.7.5	Outlook .....	110
3.8	Summary .....	111
3.9	Technical Proofs .....	112
3.9.1	How to Deal with $\mu=0$ .....	112
3.9.2	Entropy Inequalities (Lemma 3.11) .....	113
3.9.3	Error Inequality (Theorem 3.36) .....	115
3.9.4	Binary Loss Inequality for $z \leq \frac{1}{2}$ (3.57) .....	116
3.9.5	Binary Loss Inequality for $z \geq \frac{1}{2}$ (3.58) .....	117
3.9.6	General Loss Inequality (3.53) .....	117
3.10	History & References .....	119
3.11	Problems .....	119
4	<b>Agents in Known Probabilistic Environments</b> .....	125
4.1	The $AI_\mu$ Model in Functional Form .....	126
4.1.1	The Cybernetic Agent Model .....	126
4.1.2	Strings .....	128
4.1.3	AI Model for Known Deterministic Environment .....	128
4.1.4	AI Model for Known Prior Probability .....	130
4.2	The $AI_\mu$ Model in Recursive and Iterative Form .....	132
4.2.1	Probability Distributions .....	132
4.2.2	Explicit Form of the $AI_\mu$ Model .....	133
4.2.3	Equivalence of Functional and Explicit AI Model .....	134
4.3	Special Aspects of the $AI_\mu$ Model .....	135

4.3.1	Factorizable Environments	135
4.3.2	Constants and Limits	138
4.3.3	Sequential Decision Theory	139
4.4	Problems	140
<b>5</b>	<b>The Universal Algorithmic Agent AIXI</b>	<b>141</b>
5.1	The Universal AIXI Model	142
5.1.1	Definition of the AIXI Model	142
5.1.2	Universality of $M^{\text{AI}}$ and $\xi^{\text{AI}}$	144
5.1.3	Convergence of $\xi^{\text{AI}}$ to $\mu^{\text{AI}}$	145
5.1.4	Intelligence Order Relation	146
5.2	On the Optimality of AIXI	147
5.3	Value Bounds and Separability Concepts	149
5.3.1	Introduction	149
5.3.2	(Pseudo) Passive $\mu$ and the HeavenHell Example	149
5.3.3	The OnlyOne Example	150
5.3.4	Asymptotic Learnability	151
5.3.5	Uniform $\mu$	152
5.3.6	Other Concepts	152
5.3.7	Summary	153
5.4	Value-Related Optimality Results	153
5.4.1	The $\text{AI}\rho$ Models: Preliminaries	153
5.4.2	Pareto Optimality of $\text{AI}\xi$	154
5.4.3	Self-Optimizing Policy $p^\xi$ w.r.t. Average Value	156
5.5	Discounted Future Value Function	159
5.6	Markov Decision Processes (MDP)	165
5.7	The Choice of the Horizon	169
5.8	Outlook	172
5.9	Conclusions	173
5.10	Functions $\leadsto$ Chronological Semimeasures	173
5.11	Proof of the Entropy Inequality	175
5.12	History & References	177
5.13	Problems	178
<b>6</b>	<b>Important Environmental Classes</b>	<b>185</b>
6.1	Repetition of the $\text{AI}\mu/\xi$ Models	186
6.2	Sequence Prediction (SP)	187
6.2.1	Using the $\text{AI}\mu$ Model for Sequence Prediction	188
6.2.2	Using the $\text{AI}\xi$ Model for Sequence Prediction	190
6.3	Strategic Games (SG)	192
6.3.1	Introduction	192
6.3.2	Strictly Competitive Strategic Games	193
6.3.3	Using the $\text{AI}\mu$ Model for Game Playing	193
6.3.4	Games of Variable Length	195
6.3.5	Using the $\text{AI}\xi$ Model for Game Playing	195

6.4	Function Minimization (FM) . . . . .	197
6.4.1	Applications/Examples . . . . .	197
6.4.2	The Greedy Model $FMG_{\mu}$ . . . . .	198
6.4.3	The General $FM_{\mu}/\xi$ Model . . . . .	199
6.4.4	Is the General Model Inventive? . . . . .	201
6.4.5	Using the AI Models for Function Minimization . . . . .	202
6.4.6	Remark on TSP . . . . .	203
6.5	Supervised Learning from Examples (EX) . . . . .	204
6.5.1	Applications/Examples . . . . .	204
6.5.2	Supervised Learning with the $AI_{\mu}/\xi$ Model . . . . .	204
6.6	Other Aspects of Intelligence . . . . .	206
6.7	Problems . . . . .	207
<b>7</b>	<b>Computational Aspects</b> . . . . .	<b>209</b>
7.1	The Fastest & Shortest Algorithm for All Problems . . . . .	210
7.1.1	Introduction & Main Result . . . . .	210
7.1.2	Levin Search . . . . .	212
7.1.3	Fast Matrix Multiplication . . . . .	213
7.1.4	Applicability of the Fast Algorithm $M_p^{\varepsilon}$ . . . . .	214
7.1.5	The Fast Algorithm $M_p^{\varepsilon}$ . . . . .	215
7.1.6	Time Analysis . . . . .	216
7.1.7	Assumptions on the Machine Model . . . . .	218
7.1.8	Algorithmic Complexity and the Shortest Algorithm . . . . .	218
7.1.9	Generalizations . . . . .	220
7.1.10	Summary & Outlook . . . . .	220
7.2	Time-Bounded AIXI Model . . . . .	221
7.2.1	Introduction . . . . .	221
7.2.2	Time-Limited Probability Distributions . . . . .	222
7.2.3	The Idea of the Best Vote Algorithm . . . . .	224
7.2.4	Extended Chronological Programs . . . . .	224
7.2.5	Valid Approximations . . . . .	225
7.2.6	Effective Intelligence Order Relation . . . . .	226
7.2.7	The Universal Time-Bounded AIXI $_{tl}$ Agent . . . . .	226
7.2.8	Limitations and Open Questions . . . . .	227
7.2.9	Remarks . . . . .	228
<b>8</b>	<b>Discussion</b> . . . . .	<b>231</b>
8.1	What has been Achieved . . . . .	232
8.1.1	Results . . . . .	232
8.1.2	Comparison to Other Approaches . . . . .	234
8.2	General Remarks . . . . .	235
8.2.1	Miscellaneous . . . . .	235
8.2.2	Prior Knowledge . . . . .	236
8.2.3	Universal Prior Knowledge . . . . .	237
8.2.4	How AIXI( $tl$ ) Deals with Encrypted Information . . . . .	237

8.2.5	Mortal Embodied Agents . . . . .	238
8.3	Personal Remarks . . . . .	239
8.3.1	On the Foundations of Machine Learning . . . . .	239
8.3.2	In a World Without Occam . . . . .	240
8.4	Outlook & Open Questions . . . . .	241
8.5	Assumptions, Problems, Limitations . . . . .	242
8.5.1	Assumptions . . . . .	243
8.5.2	Problems . . . . .	244
8.5.3	Limitations . . . . .	244
8.6	Philosophical Issues . . . . .	245
8.6.1	Turing Test . . . . .	245
8.6.2	On the Existence of Objective Probabilities . . . . .	245
8.6.3	Free Will versus Determinism . . . . .	246
8.6.4	The Big Questions . . . . .	248
8.7	Conclusions . . . . .	248
<b>Bibliography . . . . .</b>		<b>251</b>
<b>Index . . . . .</b>		<b>265</b>

Universal Artificial Intelligence

Sequential Decisions Based on Algorithmic Probability

Hutter, M.

2005, XX, 278 p., Hardcover

ISBN: 978-3-540-22139-5