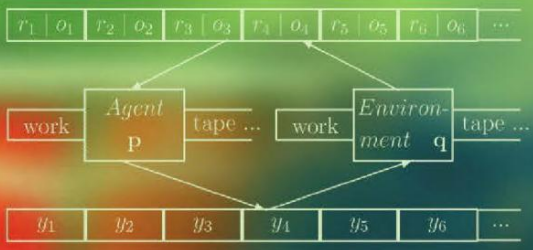
Marcus Hutter



Universal Artificial Intelligence

Sequential Decisions Based on Algorithmic Probability

Springer

Marcus Hutter

Universal Artificial Intelligence

Sequential Decisions Based on Algorithmic Probability

Springer

Author

Dr. Marcus Hutter
Istituto Dalle Molle
di Studi sull'Intelligenza
Artificiale (IDSIA)
Galleria 2
CH-6928 Manno-Lugano
Switzerland
marcus@idsia.ch
www.idsia.ch/~marcus

Series Editors

Prof. Dr. Wilfried Brauer
Institut für Informatik der TUM
Boltzmannstr. 3, 85748 Garching, Germany
Brauer@informatik.tu-muenchen.de

Prof. Dr. Grzegorz Rozenberg
Leiden Institute of Advanced Computer Science
University of Leiden
Niels Bohrweg 1, 2333 CA Leiden, The Netherlands
rozenber@liacs.nl

Prof. Dr. Arto Salomaa
Turku Centre for Computer Science
Lemminkäisenkatu 14 A, 20520 Turku, Finland
asalomaa@utu.fi

Library of Congress Control Number: 2004112980

ACM Computing Classification (1998): I.3, I.2.6, F.0, F.1.3, F.4.1, E.4, G.3

ISBN 3-540-22139-5 Springer Berlin Heidelberg New York

Cover design: KünkelLopka, Heidelberg
Typesetting: by the author
Production: LE-TeX Jelonek, Schmidt & Vöckler GbR, Leipzig
Printed on acid-free paper 45/3142/YL - 5 4 3 2 1 0

Contents

Tables, Figures, Theorems, ...

Notation

The following is a list of commonly used notation. The first entry is the symbol itself, followed by its meaning or name (if any) and the page number where the definition appears. Some standard symbols like $\mathbb{R}$ are not defined in the text. There appears a * in place of the page number for these symbols.

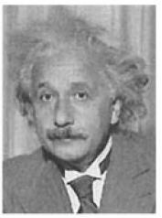
Symbol	Explanation	Page
[C35s]	classification of problems	viii
[Hut04b]	paper, book or other reference	*
(5.3)	label/reference for a formula/theorem/definition/...	*
∞	infinity	*
$\{a,...,z\}$	set containing elements a,b,...,y,z. $\{\}$ is the empty set	*
$[a,b)$	interval on the real line, closed at a and open at b	*
$\cap$,$\cup$,$\setminus$,$\in$	set intersection, union, difference, membership	*
$\wedge$,$\vee$,$\neg$	Boolean conjunction (and), disjunction (or), negation (not)	*
$\subseteq$,$\subset$	subset, proper subset	*
$\Rightarrow$	implies	*
$\Leftrightarrow$	equivalence, if and only if, iff	*
$\square$	q.e.d. (Latin), which was to be demonstrated	*
$\forall$,$\exists$	for all, there exists	*
$\approx$,$\lesssim$,$\gtrsim$	approximately equal, less equal, greater equal	33
$\ll$,$\gg$	much smaller/greater than	*
$\equiv$	equivalent, identical, equal by definition	*
$\cong$	isomorphic	*
$:=$	define as	*
$\widehat{=}$	corresponds to, informal equality	*
$\sim$	asymptotically proportional to	33
$\propto$	proportional to	*
$=$,$\neq$	equal to, not equal to	*
$+$,$-$,$\cdot$,$/$	standard arithmetic operations: sum, difference, product, ratio	*
$\sqrt{}$	square root	*
$\leq$,$\geq$,$<$,$>$	standard inequalities	*
$\lvert\mathcal{S}\rvert$,$\lvert a\rvert$	size/cardinality of set $\mathcal{S}$, absolute value of a	*

$\rightarrow$	mapping, approaches, Boolean implication	*
$\rightarrow$	converge to each other	33
$\lim_{n\rightarrow\infty}$	limiting value of argument for n tending to infinity	*
$\rightsquigarrow$	replace with	*
$\lceil x \rceil$	ceiling of x: smallest integer larger or equal than x	*
$\lfloor x \rfloor$	floor of x: largest integer smaller or equal than x	*
δ_{ab}	Kronecker symbol, $\delta_{ab}=1$ if $a=b$ and 0 otherwise	*
$\sum_{k=1}^{n}$	summation from $k=1$ to n	*
$\sum_{x}'$	summation over x for which $\mu(x)\neq 0$	69
$\prod_{k=1}^{n}$	product from $k=1$ to n	*
$\int, \int_a^b dx$	Lebesgue integral, integral from a to b over x	*
i,k,n,t	natural numbers	33
x,y,z	finite strings	33
min/max	min-/maximal element of set: $\min_{x\in\mathcal{X}} f(x)=\min\{f(x):x\in\mathcal{X}\}$	*
argmin	$\operatorname{argmin}_x f(x)$ is the x minimizing $f(x)$ (ties broken arbitrarily)	*
l.h.s.	left-hand side	*
r.h.s.	right-hand side	*
w.r.t.	with respect to	*
e.g.	exempli gratia (Latin), for example	*
i.e.	id est (Latin), that is	*
etc.	et cetera (Latin), and so forth	*
cf.	confer (Latin, imperative of conferre), compare with	*
et al.	et alii (Latin), and others	*
q.e.d.	quod erat demonstrandum (Latin), which was to be shown	*
i.i.d.	independent identically distributed (random variables)	*
iff	if and only if	*
w.p.1/i.p.	with probability 1 / in probability	71
i.m./i.m.s.	in the mean / in mean sum	71
log	logarithm to some basis	*
$\log_b$	logarithm to basis b	*
ln	natural logarithm to basis e$=2.71828...$	*
e	base of natural logarithm e$=2.71828...$	*
$\mathbb{R}$	set of real numbers	*
$\mathbb{R}^+$	set of nonnegative real numbers	*
$\mathbb{N}$	set of natural numbers $\{1,2,3,...\}$	33
$\mathbb{N}_0$	set of natural numbers including zero $\{0,1,2,3,...\}$	33
$\mathbb{Z}$	set of integers $\{...,-2,-1,0,1,2,3,...\}$	*
$\mathbb{Q}$	set of rational numbers $\{\frac{n}{d}\}$	*

$\mathbb{B}=\{0,1\}$	binary alphabet	*
$y_t \in \mathcal{Y}$	action (output of agent) in cycle t, followed by ...	128
$x_t \in \mathcal{X}$	perception (feedback/input to agent) in cycle t	45, 128
$o_t \in \mathcal{O}$	informative input/observation in cycle t	128
$r_t \in \mathcal{R} \subset \mathbb{R}$	reward in cycle t	128
ε	some small positive real number	*
ϵ	empty string	33
$*$	wildcard for some string (prefix, finite, or infinite)	33
$x_{1:n}$	$=x_1...x_n=$ string of length n	45, 68, 128
$x_{<t}$	$=x_1...x_{t-1}=$ string of length $t-1$	45, 68, 128
$yx_{k:n}$	action-perception sequence $y_k x_k ... y_n x_n$	128
$\dot{y}\dot{x}_{<k}$	actually realized action-perception sequence $\dot{y}_1\dot{x}_1...\dot{y}_{k-1}\dot{x}_{k-1}$	130
ω	infinite sequence, elementary event	33
Ω	sample space	42, 68
$\Gamma_{x_{1:n}}$	$=\{\omega:\omega_{1:n}=x_{1:n}\}=$ cylinder set	46, 68
$\ell(x)$	length of string x	33
$\langle o \rangle$	coding of object o	33
$\langle x,y \rangle$	uniquely decodable pairing of x and y	33
x'	prefix coding of x	33
$O(),o()$	big and small oh-notation	33
$a \stackrel{+}{<} b$	less within an additive const., i.e. $a \leq b+O(1)$. Similarly $\stackrel{+}{=}$	33
$a \stackrel{\times}{<} b$	less within a multiplicative const., i.e. $a=O(b)$. Similarly $\stackrel{\times}{=}$	33
$K(x)$	prefix Kolmogorov complexity of string x	37
$Km(x_{1:n})$	monotone (Kolmogorov) complexity of string $x_{1:n}$	47, 190
$K(o_1\|o_2)$	Kolmogorov complexity of object o_1, given object o_2	37
$M \stackrel{\times}{=} \xi_U$	Solomonoff-Levin's universal semimeasure	46, 48
$\mathcal{M}=\{\nu\}$	(usually countable) set of (semi)measures	48, 81
EC	$\in$ {AI, SP, FM, EX, SG, ...} is an environmental class	*
AI	artificial or algorithmic intelligence,	2
	most general computational environmental class	130, 154
SP	sequence prediction	187
CF	classification	108
SG	strategic two-player informed zero-sum games	192
FM	function minimization	197
EX	supervised learning (by examples)	204
pd	probability density function / distribution / measure	*
$\rho(x_{1:n})$	probability of string/sequence starting with $x_{1:n}$	46, 68
$\mu \in \mathcal{M}$	true generating environmental pd	68

$\mathbf{E}$	expectation value, usually w.r.t. the true distribution μ	68
$\mathbf{P}$	probability, usually w.r.t. the true distribution μ	68
$\mu(x_1\underline{x}_2x_3\underline{x}_4)$	μ probability that the 2^{nd} and 4^{th} symbols of a string are x_2 and x_4, given the 1^{st} and 3^{rd} symbols are x_1 and x_3	132
$\nu \in \mathcal{M}$	any pd in $\mathcal{M}$	70
ρ	any pd not necessarily in $\mathcal{M}$ usually specifying a policy	68
ξ	$=\sum_{\nu\in\mathcal{M}} w_\nu \nu =$ mixture (belief) pd	48, 70
w_ν	prior degree of belief in ν –or– weight of ν	48, 70
ρ^{EC}	pd of environmental argument type EC	185
ξ^{EC}	mixture distribution of type EC for class EC	185
$\ell_{x_t y_t}$	incurred loss when predicting y_t and x_t is next symbol	86
$l^\Lambda_{t\nu}$	ν-expected instantaneous loss in step t of predictor Λ	99, 87
$L^\Lambda_{n\nu}$	ν-expected cumulative loss of steps $1...n$ of predictor Λ	100
Θ_ρ	predictor with minimal number of ρ-expected errors	82
Λ_ρ	predictor that minimizes the ρ-expected loss	87
$e^\Theta_{t\nu}$	ν-probability that Θ-predictor errs in step t	83
$E^\Theta_{n\nu}$	ν-expected number of errors in steps $1...n$ of predictor Θ	83
$L^\Lambda_n \equiv L^\Lambda_{n\mu}$	abbreviation for true μ-expected loss	86
$V^{p\nu}_{km}(\dot{y}\dot{x}_{<k})$	value of policy p in environment ν given history $\dot{y}\dot{x}_{<k}$	153
y^Λ_t	prediction/decision/action of predictor Λ in step t	87
y^p_k	action of policy p in cycle k	*
γ_k	discounting sequence	159
Γ_k	value function normalization $(\sum_{i=k}^\infty \gamma_k)$	159
m,h	agent's lifespan, horizon	129, 169
p	agent's policy	126
q	deterministic environment	126
p^ν	policy that maximizes value V^p_ν	130
$V^*_\mu \equiv V^{p^\mu\mu}_{1m}$	true or generating value	130
$V^*_\xi \equiv V^{p^\xi\xi}_{1m}$	universal value	146
$D_n \equiv D^\xi_{n\mu}$	relative entropy between μ and ξ for the first n cycles	73

I have no particular talent. I am merely inquisitive.
— Albert Einstein



Albert Einstein
(1879–1955)

1 A Short Tour Through the Book

This Chapter represents a short tour through the book. It is not meant as a gentle introduction for novices, but as a condensed presentation of the most important concepts and results of the book. The price for this brevity is that in this chapter we mostly forgo mathematical rigor, subtleties, proofs, discussions, references and comparisons to other work. More seriously, some sections demand high background knowledge. Readers unfamiliar with algorithmic information theory should first read Chapter 2 or consult the textbooks [LV97, Cal02]. Readers unfamiliar with sequential decision theory should first read Chapter 4 or consult the textbooks [BT96, SB98]. Before becoming discouraged by the complexity of some of the sections, it is better to skip them completely.

1.1 Introduction

Artificial Intelligence. The science of artificial intelligence (AI) might be defined as the construction of intelligent systems and their analysis. A natural definition of a *system* is anything that has an input and an output stream. Intelligence is more complicated. It can have many faces like creativity, solving problems, pattern recognition, classification, learning, induction, deduction, building analogies, optimization, surviving in an environment, language processing, knowledge and many more. A formal definition incorporating every aspect of intelligence, however, seems difficult. Further, intelligence is graded: There is a smooth transition between systems, which everyone would agree to be not intelligent, and truly intelligent systems. One simply has to look in nature, starting with, for instance, inanimate crystals, then amino acids, then some RNA fragments, then viruses, bacteria, plants, animals, apes, followed by the truly intelligent homo sapiens, and possibly continued by AI systems or ETs. So, the best we can expect to find is a partial or total order relation on the set of systems, which orders them w.r.t. their degree of intelligence (like

intelligence tests do for human systems, but for a limited class of problems). Having this order we are, of course, interested in large elements, i.e. highly intelligent systems. If a largest element exists, it would correspond to the most intelligent system which could exist.

Most, if not all, known facets of intelligence can be formulated as goal driven or, more precisely, as maximizing some utility function. It is therefore sufficient to study goal-driven AI. For example, the (biological) goal of animals and humans is to survive and spread. The goal of AI systems should be to be useful to humans. The problem is that, except for special cases, we know neither the utility function nor the environment in which the agent will operate in advance.

Main idea. This book presents a theory that formally[1] solves the problem of unknown goal and environment. It might be viewed as a unification of the ideas of universal induction, probabilistic planning and reinforcement learning, or as a unification of sequential decision theory with algorithmic information theory. We apply this model to some of the facets of intelligence, including induction, game playing, optimization, reinforcement and supervised learning, and show how it solves these problem classes. This, together with general convergence theorems, supports the belief that the constructed universal AI system is the best one in a sense to be clarified in the following, i.e. that it is the most intelligent environment-independent system possible. The intention of this book is to introduce the universal AI model and give an extensive analysis.

1.2 Simplicity & Uncertainty

This section introduces Occam's razor principle, Kolmogorov complexity, and objective/subjective probabilities. We finally arrive at the problem of universal prediction, and its solution by Solomonoff.

1.2.1 Introduction

An important and nontrivial aspect of intelligence is inductive inference. Simply speaking, induction is the process of predicting the future from the past, or, more precisely, it is the process of finding rules in (past) data and using these rules to guess future data. Weather or stock-market forecasting or continuing number series in an IQ test are nontrivial examples. Making good predictions plays a central role in natural and artificial intelligence in general, and in machine learning in particular. All induction problems can be phrased

[1] With a formal solution we mean a rigorous mathematically definition, uniquely specifying the solution. In the following, a solution is always meant in this formal sense.

as sequence prediction tasks. This is, for instance, obvious for time-series prediction, but also includes classification tasks. Having observed data x_t at times $t<n$, the task is to predict the n^{th} symbol x_n from sequence $x_1...x_{n-1}$. This *prequential approach* [Daw84] skips over the intermediate step of learning a model based on observed data $x_1...x_{n-1}$ and then using this model to predict x_n. The prequential approach avoids problems of model consistency, how to separate noise from useful data, and many other issues. The goal is to make "good" predictions, where the prediction quality is usually measured by a loss function, which shall be minimized. The key concept to well-defining and solving induction problems is *Occam's razor* (simplicity) principle, which says that "*Entities should not be multiplied beyond necessity.*" This may be interpreted as keeping the simplest theory consistent with the observations $x_1...x_{n-1}$ and using this theory to predict x_n. Before we can present Solomonoff's formal solution, we have to quantify Occam's razor in terms of Kolmogorov complexity, and introduce the notions of subjective and objective probabilities.

1.2.2 Algorithmic Information Theory

Intuitively, a string is simple if it can be described in a few words, like "the string of one million ones", and is complex if there is no such short description, like for a random string whose shortest description is specifying it bit by bit. We can restrict the discussion to binary strings, since for other (non-stringy mathematical) objects we may assume some default coding as binary strings. Furthermore, we are only interested in effective descriptions, and hence restrict decoders to be Turing machines. Let us choose some universal (so-called prefix) *Turing machine* U with unidirectional binary input and output tapes and a bidirectional work tape. We can then define the *prefix Kolmogorov complexity* [Cha75, Gác74, Kol65, Lev74] of a binary string x as the length ℓ of the shortest program p for which U outputs the binary string x

$$K(x) \;:=\; \min_p\{\ell(p) : U(p) = x\}.$$

Simple strings like 000...0 can be generated by short programs, and, hence have low Kolmogorov complexity, but irregular (e.g. random) strings are their own shortest description, and hence have high Kolmogorov complexity. An important property of K is that it is nearly independent of the choice of U. Furthermore, it shares many properties with Shannon's entropy (information measure) S, but K is superior to S in many respects. Figure 2.11 on page 38 contains a schematic graph of K. To be brief, K is an excellent universal complexity measure, suitable for quantifying Occam's razor. There is (only) one severe disadvantage: K is not finitely computable. More precisely, a function f is said to be *finitely computable* (or *recursive*) if there exists a Turing machine which, given x, computes $f(x)$ and then halts. Some functions are not finitely computable but still *approximable* in the sense that there is a nonhalting Turing machine with an infinite output sequence $y_1,y_2,y_3,...$ with $\lim_{t\to\infty} y_t = f(x)$.

1.2.5 Generalized Universal (Semi)Measures

One can derive a universal prior in a different way: Solomonoff [Sol64, Eq.(13)] defines a somewhat problematic mixture over all computable probability distributions. Levin [ZL70] considers the larger class $\mathcal{M}_U := \{\nu_1,\nu_2,...\}$ of all so-called enumerable semimeasures. Let $\mu \in \mathcal{M}_U$, and assign (consistent with Occam's razor) a prior plausibility of $2^{-K(\nu_a)}$ to ν_a. Then the prior plausibility of $x_{1:n}$ is, by elementary probability theory,

$$\xi_U(x_{1:n}) := \sum_{\nu\in\mathcal{M}_U} 2^{-K(\nu)}\nu(x_{1:n}). \tag{1.3}$$

One can show that ξ_U coincides with M within an (irrelevant) multiplicative constant, i.e. $M(x) \stackrel{\times}{=} \xi_U(x)$, where $f(x) \stackrel{\times}{\leq} g(x)$ abbreviates $f(x)=O(g(x))$, and $\stackrel{\times}{=}$ denotes $\stackrel{\times}{\leq}$ *and* $\stackrel{\times}{\geq}$. Both ξ_U and M can be shown to be lower semicomputable. The dominance $M(x) \stackrel{\times}{=} \xi_U(x) \geq 2^{-K(\mu)}\mu(x)$ is the central ingredient in the proof of (1.2). The advantage of ξ_U over M is that the definition immediately generalizes to arbitrary weighted sums of (semi)measures in $\mathcal{M}$ for arbitrary countable $\mathcal{M}$. Most proofs in this book go through for generic $\mathcal{M}$ and weights.

So, what is so special about the class of all enumerable semimeasures $\mathcal{M}_U$? The larger we choose $\mathcal{M}$, the less restrictive is the assumption that $\mathcal{M}$ should contain the true distribution μ, which will be essential throughout the book. Why not restrict to the still rather general class of estimable or finitely computable (semi)measures? For *every* countable class $\mathcal{M}$, the mixture $\xi(x) := \xi_{\mathcal{M}}(x) := \sum_{\nu\in\mathcal{M}} w_\nu \nu(x)$ with $w_\nu > 0$, the important dominance $\xi(x) \geq w_\nu \nu(x)$ is satisfied. The question is, what properties does ξ possess. The distinguishing property of $\mathcal{M}_U$ is that ξ_U is itself an element of $\mathcal{M}_U$. On the other hand, in this book $\xi_{\mathcal{M}} \in \mathcal{M}$ is not by itself an important property. What matters is whether ξ is computable in one of the senses we defined above. There is an enumerable semimeasure (M) that dominates all enumerable semimeasures in $\mathcal{M}_U$. As we will see, there is *no* estimable semimeasure that dominates all computable measures, and there is *no* approximable semimeasure that dominates all approximable measures. From this it follows that for a universal (semi)measure which at least satisfies the weakest form of computability, namely being approximable, the largest dominated class among the classes considered in this book is the class of enumerable semimeasures, but there are even larger classes [Sch02a]. This is the reason why $\mathcal{M}_U$ and M play a special role in this (and other) works. In practice though, one has to restrict to a finite subset of finitely computable environments ν to get a finitely computable ξ.

1.3 Universal Sequence Prediction

In the following we more closely investigate sequence prediction (SP) schemes based on Solomonoff's universal prior $M \stackrel{\times}{=} \xi_U$ and on more general Bayes

mixtures ξ, mainly from a decision-theoretic perspective. In particular, we show that they are optimal w.r.t. various optimality criteria.

1.3.1 Setup & Convergence

Let $\mathcal{M}:=\{\nu_1,\nu_2,...\}$ be a countable set of candidate probability distributions on strings over the finite alphabet $\mathcal{X}$. We define a weighted average on $\mathcal{M}$:

$$\xi(x_{1:n}) \;:=\; \sum_{\nu\in\mathcal{M}} w_\nu\cdot\nu(x_{1:n}), \qquad \sum_{\nu\in\mathcal{M}} w_\nu = 1, \quad w_\nu > 0. \tag{1.4}$$

It is easy to see that ξ is a probability distribution as the weights w_ν are positive and normalized to 1 and the $\nu\in\mathcal{M}$ are probabilities. We call ξ universal relative to $\mathcal{M}$, as it multiplicatively dominates all distributions in $\mathcal{M}$ in the sense that $\xi(x_{1:n})\geq w_\nu\cdot\nu(x_{1:n})$ for all $\nu\in\mathcal{M}$. In the following, we assume that $\mathcal{M}$ is known and contains the true but unknown distribution μ, i.e. $\mu\in\mathcal{M}$, and $x_{1:\infty}$ is sampled from μ. We abbreviate expectations w.r.t. μ by $\mathbf{E}[..]$; for instance, $\mathbf{E}[f(x_{1:n})]=\sum_{x_{1:n}\in\mathcal{X}^n}\mu(x_{1:n})f(x_{1:n})$. We use the (total) relative entropy D_n and squared Euclidian distance S_n to measure the distance between μ and ξ:

$$D_n \;:=\; \mathbf{E}\left[\ln\frac{\mu(x_{1:n})}{\xi(x_{1:n})}\right], \quad S_n \;:=\sum_{t=1}^{n}\mathbf{E}\left[\sum_{x'_t\in\mathcal{X}}\Big(\mu(x'_t|x_{<t})-\xi(x'_t|x_{<t})\Big)^2\right]. \tag{1.5}$$

The following sequence of inequalities can be shown, which generalize Solomonoff's result (1.2): $S_n\leq D_n\leq\ln w_\mu^{-1}<\infty$. The finiteness of S_∞ implies $\xi(x'_t|x_{<t})-\mu(x'_t|x_{<t})\to 0$ for $t\to\infty$ w.μ.p.1 for any x'_t ($\sum_{t=1}^{\infty}s_t^2<\infty\Rightarrow s_t\to 0$). We also show that $\sum_{t=1}^{n}\mathbf{E}[(\sqrt{\xi(x_t|x_{<t})/\mu(x_t|x_{<t})}-1)^2]\leq D_n\leq\ln w_\mu^{-1}<\infty$, which implies $\xi(x_t|x_{<t})/\mu(x_t|x_{<t})\to 1$ for $t\to\infty$ w.μ.p.1. This convergence motivates the belief that predictions based on (the known) ξ are asymptotically as good as predictions based on (the unknown) μ, with rapid convergence.

1.3.2 Loss Bounds

Most predictions are eventually used as a basis for some decision or action, which itself leads to some reward or loss. Let $\ell_{x_ty_t}\in[0,1]\subset\mathbb{R}$ be the received loss when performing prediction/decision/action $y_t\in\mathcal{Y}$, and $x_t\in\mathcal{X}$ is the t^{th} symbol of the sequence. Let $y_t^\Lambda\in\mathcal{Y}$ be the prediction of a (causal) prediction scheme Λ. The true probability of the next symbol being x_t, given $x_{<t}$, is $\mu(x_t|x_{<t})$. The expected loss when predicting y_t is $\mathbf{E}[\ell_{x_ty_t}]$. The total μ-expected loss suffered by the Λ scheme in the first n predictions is

$$L_n^\Lambda \;:=\; \sum_{t=1}^{n}\mathbf{E}[\ell_{x_ty_t^\Lambda}].$$

The goal is to minimize the expected loss. More generally, we define the Λ_ρ sequence prediction scheme (later also called SPρ) $y_t^{\Lambda_\rho} := \text{argmin}_{y_t \in \mathcal{Y}} \sum_{x_t} \rho(x_t|x_{<t}) \ell_{x_t y_t}$, which minimizes the ρ-expected loss. If μ is known, Λ_μ is obviously the best prediction scheme in the sense of achieving minimal expected loss ($L_n^{\Lambda_\mu} \leq L_n^{\Lambda}$ for any Λ). We prove the following loss bound for the universal Λ_ξ predictor

$$0 \;\leq\; L_n^{\Lambda_\xi} - L_n^{\Lambda_\mu} \;\leq\; D_n + \sqrt{4L_n^{\Lambda_\mu} D_n + D_n^2} \;\leq\; 2D_n + 2\sqrt{L_n^{\Lambda_\mu} D_n}. \qquad (1.6)$$

Together with $L_n \leq n$ and $D_\infty \leq \ln w_\mu^{-1} < \infty$, this shows that $\frac{1}{n} L_n^{\Lambda_\xi} - \frac{1}{n} L_n^{\Lambda_\mu} = O(n^{-1/2})$, i.e. asymptotically Λ_ξ achieves the optimal average loss of Λ_μ with rapid convergence. Moreover, $L_\infty^{\Lambda_\xi}$ is finite if $L_\infty^{\Lambda_\mu}$ is finite, and $L_n^{\Lambda_\xi} / L_n^{\Lambda_\mu} \to 1$ if $L_\infty^{\Lambda_\mu}$ is not finite. Bound (1.6) also implies $L_n^{\Lambda} \geq L_n^{\Lambda_\xi} - 2\sqrt{L_n^{\Lambda_\xi} D_n}$, which shows that *no* (causal) predictor Λ whatsoever achieves significantly less (expected) loss than Λ_ξ. Note that for $w_\nu = 2^{-K(\nu)}$, $D_n \leq \ln 2 \cdot K(\mu)$ is of "reasonable" size. Instantaneous loss bounds can also be proven.

1.3.3 Optimality Properties

For any predictor Λ, a worst-case lower bound that asymptotically matches the upper bound (1.6) can be derived. More precisely, let Λ be any deterministic predictor not knowing from which distribution $\mu \in \mathcal{M}$ the observed sequence $x_1 x_2 ...$ is sampled. Predictor Λ knows (depends on) $\mathcal{M}$, w_ν, and ℓ, and has at time t access to the previous outcomes $x_{<t}$. Then for every n there is an $\mathcal{M}$ and $\mu \in \mathcal{M}$ and ℓ and weights w_ν such that

$$L_n^{\Lambda} - L_n^{\Lambda_\mu} \;\geq\; \tfrac{1}{2}[S_n + \sqrt{4L_n^{\Lambda_\mu} S_n + S_n^2}], \qquad \text{and} \quad D_n/S_n \to 1 \quad \text{for} \quad n \to \infty.$$

For the universal predictor $\Lambda = \Lambda_\xi$, the lower bound holds even without the factor $\frac{1}{2}$. This shows that bound (1.6) is quite tight in the sense that no other predictor can lead to significantly smaller bounds without making extra assumptions on $\mathcal{M}$, w_ν, or ℓ. For instance, for logarithmic and quadratic loss functions the regret $L_\infty^{\Lambda_\xi} - L_\infty^{\Lambda_\mu}$ is finite and bounded by $\ln w_\mu^{-1}$.

A different kind of optimality is *Pareto optimality*. Let $\mathcal{F}(\mu,\rho)$ be any performance measure of ρ relative to μ. The universal prior ξ is called Pareto optimal w.r.t. $\mathcal{F}$ if there is no ρ with $\mathcal{F}(\nu,\rho) \leq \mathcal{F}(\nu,\xi)$ for all $\nu \in \mathcal{M}$ and strict inequality for at least one ν. We show that the universal prior ξ is Pareto optimal w.r.t. the squared distance S_n, the relative entropy D_n, and the losses L_n. That is, for all performance measures that are relevant from a decision-theoretic point of view (i.e. for all loss functions ℓ) any improvement achieved by some predictor Λ_ρ over Λ_ξ in some environments ν is balanced by a deterioration in other environments. There are non-decision-theoretic performance measures w.r.t. which ξ is *not* Pareto optimal. Pareto optimality is a rather weak notion of optimality, but it emphasizes the distinctiveness of Bayes mixture strategies.

Pareto optimality of ξ still leaves open the question of how to choose the class $\mathcal{M}$ and the weights w_ν. We have argued that $\mathcal{M}_U$ is the largest $\mathcal{M}$ suitable from a computational point of view. $\mathcal{M}_U$ is also sufficiently large if we make the mild assumption that strings are sampled from a computable probability distribution. We show that within the class of enumerable weight functions with short program, the universal weights $w_\nu = 2^{-K(\nu)}$ lead to the smallest performance bounds within an additive (to $\ln w_\nu^{-1}$) constant in all enumerable environments. This argument justifies the selection of Solomonoff-Levin's prior (1.3) among all possible Bayes mixtures.[4]

1.3.4 Miscellaneous

Games of chance. The general loss bound (1.6) can, for instance, be used to estimate the time needed to reach the winning threshold in a game of chance (defined as a sequence of bets, observations and rewards). At time t we bet, depending on the history $x_{<t}$, a certain amount of money s_t, take some action y_t, observe outcome x_t, and receive reward r_t. Our net profit, which we want to maximize, is $p_t = r_t - s_t \in [p_{max} - p_\Delta, p_{max}]$. The loss, which we want to minimize, can be identified with the negative (scaled) profit, $\ell_{x_k y_t} = (p_{max} - p_t)/p_\Delta \in [0,1]$. The Λ_ρ-system acts as to maximize the ρ-expected profit. Let $\bar{p}_n^{\Lambda_\rho}$ be the average expected profit of the first n rounds. Bound (1.6) shows that the average profit of the Λ_ξ system converges to the best possible average profit $\bar{p}_n^{\Lambda_\mu}$ achieved by the Λ_μ scheme ($\bar{p}_n^{\Lambda_\xi} - \bar{p}_n^{\Lambda_\mu} = O(n^{-1/2}) \to 0$ for $n \to \infty$). If there is a profitable scheme at all, then asymptotically the universal Λ_ξ scheme will also become profitable with the same average profit. We further show using ξ_U that $(2p_\Delta/\bar{p}_n^{\Lambda_\mu})^2 \cdot \ln 2 \cdot K(\mu)$ is an upper bound on the number of bets n needed to reach the winning zone. The bound is proportional to the complexity of the environment μ.

Continuous probability classes $\mathcal{M}$. We have considered thus far countable probability classes $\mathcal{M}$, which makes sense from a computational point of view. On the other hand, in statistical parameter estimation one often has a continuous hypothesis class (e.g. a Bernoulli(θ) process with unknown $\theta \in [0,1]$). Let $\mathcal{M} := \{\mu_\theta : \theta \in \Theta \subseteq \mathbb{R}^d\}$ be a family of probability distributions parameterized by a d-dimensional continuous parameter θ. Let $\mu \equiv \mu_{\theta_0} \in \mathcal{M}$ be the true generating distribution. For a continuous weight density $w(\theta) > 0$ the sums in (1.4) are naturally replaced by integrals: $\xi(x_{1:n}) := \int_\Theta w(\theta) \cdot \mu_\theta(x_{1:n}) d\theta$ with $\int_\Theta w(\theta) d\theta = 1$. The most important property of ξ in the discrete case was the dominance $\xi(x_{1:n}) \geq w_\nu \cdot \nu(x_{1:n})$, which was obtained from (1.4) by dropping the sum over ν. The analogous construction here is to restrict the integral over Θ to a small vicinity N_δ of θ. For sufficiently smooth μ_θ and $w(\theta)$ we expect $\xi(x_{1:n}) \gtrsim |N_{\delta_n}| \cdot w(\theta) \cdot \mu_\theta(x_{1:n})$, where $|N_{\delta_n}|$ is the volume of

[4] Readers who smell some free lunch here [WM97] should appease their hunger with Section 3.6.5.

N_{δ_n}. This in turn leads to $D_n \lesssim \ln w_\mu^{-1} + \ln|N_{\delta_n}|^{-1}$, where $w_\mu := w(\theta_0)$. N_{δ_n} should be the largest possible region in which $\ln\mu_\theta$ is approximately flat on average. More precisely, generalizing [CB90] to the non-i.i.d. case, we show $D_n \leq \ln w_\mu^{-1} + \frac{d}{2}\ln\frac{n}{2\pi} + O(1)$, where the $O(1)$ term depends on the smoothness of μ_θ, measured by the Fisher information. D_n is no longer bounded by a constant, but still grows only logarithmically with n, the intuitive reason being the necessity to describe θ to an accuracy $O(n^{-1/2})$. So, bound (1.6) is also applicable to the case of continuously parameterized probability classes.

1.4 Rational Agents in Known Probabilistic Environments

1.4.1 The Agent Model

A very general framework for intelligent systems is that of rational agents [RN95]. In cycle k, an agent performs *action* $y_k \in \mathcal{Y}$ (output), which results in a *perception* $x_k \in \mathcal{X}$ (input), followed by cycle $k+1$, and so on. We assume that the action and perception spaces $\mathcal{X}$ and $\mathcal{Y}$ are finite. We write $p(x_{<k}) = y_{1:k}$ to denote the output $y_{1:k}$ of the agent's policy p on input $x_{<k}$, and similarly $q(y_{1:k}) = x_{1:k}$ for the environment q in the case of deterministic environments. We call policy p and environment q behaving in this way *chronological*. The figure on the book cover and on page 128 depicts this interaction in the case where p and q are modeled by Turing machines. Note that policy and environment are allowed to depend on the complete history. We do not make any MDP or POMDP assumption here, and we do not talk about states of the environment, only about observations. In the more general case of a *probabilistic environment*, given the history $yx_{<k}y_k \equiv yx_1...yx_{k-1}y_k \equiv y_1x_1...y_{k-1}x_{k-1}y_k$, the probability that the environment leads to perception x_k in cycle k is (by definition) $\mu(yx_{<k}y\underline{x}_k)$. The underlined argument $\underline{x}_k$ in μ is a random variable, and the other non-underlined arguments $yx_{<k}y_k$ represent conditions.[5] We call probability distributions like μ *chronological*. Since value-optimizing policies (see below) can always be chosen deterministic, there is no real need to generalize the setting to probabilistic policies.

1.4.2 Value Functions & Optimal Policies

The goal of the agent is to maximize future *rewards*, which are provided by the environment through the inputs x_k. The inputs $x_k \equiv r_k o_k$ are divided into a regular part o_k and some (possibly empty or delayed) reward $r_k \in [0\,, r_{max}]$.[6] We use the abbreviation

[5] The standard notation $\mu(x_k|yx_{<k}y_k)$ for conditional probabilities destroys the chronological order and would become confusing in later expressions.

[6] In the reinforcement learning literature when dealing with (PO)MDPs the reward is usually considered to be a function of the environmental state. The zero-

its behavior is completely defined by (1.10) and (1.11). It (slightly) depends on the choice of the universal Turing machine, because $K()$ and $\ell()$ depend on U and hence are defined only up to terms of order one. The AIXI model also depends on the choice of $\mathcal{X}$ and $\mathcal{Y}$, but we do not expect any bias when the spaces are chosen sufficiently large and simple, e.g. all strings of length 2^{16}. Choosing $\mathbb{N}$ as the I/O spaces would be ideal, but whether the maxima (or suprema) exist in this case has to be shown beforehand. The only nontrivial dependence is on the horizon m. Ideally, we would like to chose $m=\infty$, but there are several subtleties to be unraveled later, which prevent at least a naive limit $m\to\infty$. So apart from m and unimportant details, *the AIXI system is uniquely defined by (1.10) and (1.11) without adjustable parameters.*

1.5.2 On the Optimality of AIXI

Universality and convergence of ξ. One can show that also ξ defined in (1.10) is universal and rapidly converges to μ analogous to the induction (SP) case. If we take a finite product of conditional ξ's and use the chain rule, we see that also $\xi(yx_{<k}y\underline{x}_{k:k+h})$ converges to $\mu(yx_{<k}y\underline{x}_{k:k+h})$ for $k\to\infty$. This gives confidence that the outputs y_k^ξ of the AIXI model (1.11) could converge to the outputs y_k^μ of the AIμ model (1.8), at least for a bounded moving horizon h. The problems with a fixed horizon m and especially $m\to\infty$ will be discussed at the end of this section.

Universally optimal AI systems. We call an AI model *universal* if it is independent of the true environment μ (unbiased, model-free) and is able to solve any solvable problem and learn any learnable task. Further, we call a universal model *universally optimal* if there is no program that can solve or learn significantly faster (in terms of interaction cycles). As the AIXI model is parameter-free, ξ converges to μ, the AIμ model is itself optimal, and we expect no other model to converge faster to AIμ by analogy to the SP case,

we expect AIXI to be universally optimal.

This is our main claim. Further support is given below.

Intelligence order relation. We want to call a policy p *more or equally intelligent* than a policy p' and write $p\succeq p'$ if p yields in every cycle k and for every fixed history $yx_{<k}$ higher (future) ξ-expected reward sum than p'. It is a formal exercise to show that $p^\xi\succeq p$ for all p. The AIXI model is hence the most intelligent agent w.r.t. $\succeq$. Relation $\succeq$ is a universal order relation in the sense that it is free of any parameters (except m) or specific assumptions about the environment. A proof that $\succeq$ is a reasonable intelligence order (which we believe to be true) would prove that AIXI is universally optimal.

Value bounds. The values V_ρ^* associated with the AIρ systems correspond roughly to the negative total loss $-L_n^{\Lambda_\rho}$ (with $n=m$) of the SPρ ($=\Lambda_\rho$) systems.

In the SP case we were interested in small bounds for the regret $L_n^{\Lambda_\xi} - L_n^{\Lambda_\mu}$. Unfortunately, simple value bounds for AIXI or any other AI system in terms of V_ν^* analogous to the loss bound (1.6) cannot hold. We even have difficulties in specifying what we can expect to hold for AIXI or any AI system that claims to be universally optimal. In SP, the only important property of μ for proving loss bounds was its complexity $K(\mu)$. In the AI case, there are no useful bounds in terms of $K(\mu)$ only. We either have to study restricted problem or environmental classes or consider bounds depending on other properties of μ, rather than on its complexity only.

1.5.3 Value-Related Optimality Results

The mixture distribution ξ. In the following, we consider general Bayes mixtures ξ over classes $\mathcal{M}$ of chronological probability distributions ν:

$$\xi(\underline{yx}_{1:m}) = \sum_{\nu\in\mathcal{M}} w_\nu \nu(\underline{yx}_{1:m}) \quad \text{with} \quad \sum_{\nu\in\mathcal{M}} w_\nu = 1 \quad \text{and} \quad w_\nu > 0 \quad \forall \nu \in \mathcal{M}.$$

We define V_ξ^p, p^ξ, and V_ξ^* as in (1.7)–(1.9) with μ replaced by ξ. Policy p^ξ is called the AIξ model. For $\xi = \xi_U$ the AIXI≡AIξ_U model is recovered. If μ is unknown, but known to belong to the known class $\mathcal{M}$, it is natural to follow policy p^ξ, which maximizes V_ξ^p. The (true μ-)expected reward when following policy p^ξ is $V_\mu^{p^\xi}$. The optimal (but infeasible) policy p^μ yields reward $V_\mu^{p^\mu} \equiv V_\mu^*$. It is now of interest (a) whether there are policies with uniformly larger value than $V_\mu^{p^\xi}$ and (b) how close $V_\mu^{p^\xi}$ is to V_μ^*.

Linearity and convexity of V_ρ in ρ. The following properties of V_ρ are crucial. V_ρ^p is a linear function in ρ, and V_ρ^* is a convex function in ρ in the sense that

$$V_\xi^p = \sum_{\nu\in\mathcal{M}} w_\nu V_\nu^p \qquad \text{and} \qquad V_\xi^* \le \sum_{\nu\in\mathcal{M}} w_\nu V_\nu^*.$$

Linearity is obvious from the definition of V_ρ^p, and convexity follows easily from the convexity of $\max_p$ and nonnegativity of the weights w_ν. One loose interpretation of the convexity is that a mixture can never increase performance.

Pareto optimality of AIξ. Similarly to the SP case, one can show that p^ξ is *Pareto optimal* in the sense that there is no other policy p with $V_\nu^p \ge V_\nu^{p^\xi}$ for all $\nu \in \mathcal{M}$ and strict inequality for at least one ν. In particular, AIXI is Pareto optimal.

Self-optimizing policy p^ξ w.r.t. average value. Since we do not know the true environment μ in advance, we are interested under which circumstances[7]

[7] Here and elsewhere we interpret $a_m \to b_m$ as an abbreviation for $a_m - b_m \to 0$. $\lim_{m\to\infty} b_m$ may not exist.

$$\tfrac{1}{m}V_\nu^{p^\xi} \to \tfrac{1}{m}V_\nu^* \quad \text{for horizon} \quad m\to\infty \quad \text{for } \textit{all} \quad \nu\in\mathcal{M}. \tag{1.12}$$

Note that V_ν as well as $p^\xi = p^\xi_m$ depend on m. The least we must demand from $\mathcal{M}$ to have a chance that (1.12) is true is that there exists a policy (sequence) $\tilde{p}=\tilde{p}_m$ at all with this property, i.e.

$$\exists\tilde{p}: \ \tfrac{1}{m}V_\nu^{\tilde{p}} \to \tfrac{1}{m}V_\nu^* \quad \text{for horizon} \quad m\to\infty \quad \text{for } \textit{all} \quad \nu\in\mathcal{M}. \tag{1.13}$$

We show that this necessary condition is also sufficient, i.e. (1.13) implies (1.12). This is another (asymptotic) optimality property of policy p^ξ. *If* universal convergence in the sense of (1.13) is possible at all in a class of environments $\mathcal{M}$, then policy p^ξ converges in the same sense (1.12). We call policies $\tilde{p}$ with a property like (1.13) *self-optimizing* [KV86].

Unfortunately, the result is not an asymptotic convergence statement of a single policy p^ξ, since p^ξ depends on m. The result merely says that under the stated conditions the average value of p^ξ_m is arbitrarily close to optimum for sufficiently large (pre-chosen) horizon m. This weakness will be resolved in the following.

Discounted future value function. We now shift our focus from the total value to future values (value-to-go). First, we have to get rid of the horizon parameter m. We eliminate the horizon by discounting the rewards $r_k \rightsquigarrow \gamma_k r_k$ with $\gamma_k \geq 0$ and $\sum_{i=1}^\infty \gamma_i < \infty$ and taking $m\to\infty$. The analogue of m is now an effective horizon h_k^{eff}, which may be defined by $\sum_{i=k}^{k+h_k^{eff}}\gamma_i \approx \sum_{i=k+h_k^{eff}}^\infty \gamma_i$. Furthermore, we renormalize the value V by $\sum_{i=k}^\infty\gamma_i$ and denote it by $V_{k\gamma}$. Finally, we extend the definition to probabilistic policies π (which is not essential). We define the γ-discounted weighted-average future *value* of (probabilistic) policy π in environment ρ given history $yx_{<k}$, or shorter, the ρ-value of π given $yx_{<k}$, as

$$V_{k\gamma}^{\pi\rho}(yx_{<k}) \ := \ \frac{1}{\Gamma_k}\lim_{m\to\infty}\sum_{yx_{k:m}}(\gamma_k r_k + ... + \gamma_m r_m)\rho(yx_{<k}\underline{yx}_{k:m})\pi(yx_{<k}\underline{yx}_{k:m}),$$

with $\Gamma_k := \sum_{i=k}^\infty\gamma_i$. The policy p^ρ is defined as to maximize the future value $V_{k\gamma}^{\pi\rho}$:

$$p^\rho := \arg\max_\pi V_{k\gamma}^{\pi\rho}, \qquad V_{k\gamma}^{*\rho} := V_{k\gamma}^{p^\rho\rho} = \max_\pi V_{k\gamma}^{\pi\rho} \geq V_{k\gamma}^{\pi\rho}\ \forall\pi.$$

Setting $\gamma_k=1$ for $k\leq m$ and $\gamma_k=0$ for $k>m$ gives back the old undiscounted model with horizon m and $V_{1\gamma}^{p\rho} = \frac{1}{m}V_\rho^p$. Note that $V_{k\gamma}$ depends on the realized history $yx_{<k}$. More important, p^ρ can be shown to be independent of k. Similarly to the undiscounted case, one can prove that for every k and history $yx_{<k}$, $V_{k\gamma}^{\pi\rho}$ is a linear function in ρ, $V_{k\gamma}^{*\rho}$ is a convex function in ρ, and p^ξ is Pareto optimal in the sense that there is no other policy π with $V_{k\gamma}^{\pi\nu} \geq V_{k\gamma}^{p^\xi\nu}$ for all $\nu\in\mathcal{M}$ and strict inequality for at least one ν. Finally, p^ξ is self-optimizing (w.r.t. discounted value) if $\mathcal{M}$ admits self-optimizing policies:

1.5.5 The Choice of the Horizon

The only significant arbitrariness in the AIXI model lies in the choice of the lifespan m or in the discounted case in the discount sequence γ_k. We will not discuss ad hoc choices for specific problems. We are interested in universal choices. In many cases the time we are willing to run a system depends on the quality of its actions. Hence, the lifetime, if finite at all, is not known in advance. Geometric discounting $r_k \leadsto r_k \cdot \gamma^k$ solves the mathematical problem of $m \to \infty$ but is not a real solution, since an effective horizon $h^{eff} \sim \ln\gamma^{-1} < \infty$ has been introduced. The scale-invariant discounting $r_k \leadsto r_k \cdot k^{-\alpha}$ with $\alpha > 1$ has a dynamic horizon $h \sim k$. This choice has some appeal, as it seems that humans of age k years also usually do not plan their lives for more than the next $\sim k$ years. It also satisfies the condition $\frac{\gamma_{k+1}}{\gamma_k} \to 1$, necessary for AI$\xi$ being self-optimizing in ergodic MDPs. The largest lower semicomputable horizon with guaranteed finite reward sum $\Gamma_1 < \infty$ is obtained by the discount $r_k \leadsto r_k \cdot 2^{-K(k)}$, where $K(k)$ is the Kolmogorov complexity of k. This is maybe the most attractive universal discount. It is similar to a near-harmonic discount $r_k \leadsto r_k \cdot k^{-(1+\varepsilon)}$, since $2^{-K(k)} \leq 1/k$ for most k and $2^{-K(k)} \geq c/(k \log^2 k)$ for some constant c. We are not sure whether the choice of the horizon is of marginal importance, as long as it is chosen sufficiently large, or whether the choice will turn out to be a central topic for the AIXI model or for the planning aspect of any universal AI system in general. Most, if not all, problems in agent design of balancing exploration and exploitation vanish by a sufficiently large choice of the (effective) horizon and a sufficiently general prior.

1.6 Important Environmental Classes

In this and the next section we define $\xi = \xi_U \stackrel{\text{def}}{=} M$ be Solomonoff's prior, i.e. AIξ=AIXI. Each subsection represents an abstract on what will be done in the corresponding section of Chapter 6.

1.6.1 Introduction

In order to give further support for the universality and optimality of the AIξ theory, we apply AIξ to a number of problem classes. They include sequence prediction, strategic games, function minimization and, especially, how AIξ learns to learn supervised. For some classes we give concrete examples to illuminate the scope of the problem class. We first formulate each problem class in its natural way (when μ^{problem} is known) and then construct a formulation within the AIμ model and prove its equivalence. We then consider the consequences of replacing μ by ξ. The main goal is to understand why and how the problems are solved by AIξ. We only highlight special aspects of each problem class. The goal is to give a better picture of the flexibility of the AIξ model.

time bounds for practical problems can often be computed quickly, i.e. $time_{t_p}(x)/time_p(x)$ often converges very quickly to zero. Furthermore, from a practical point of view, the provability restrictions are often rather weak. Hence, we have constructed for all those problems a solution that is asymptotically only a factor $1+\varepsilon$ slower than the (provably) fastest algorithm. On the flip side, for realistically sized problems, the lower-order terms usually dominate, which limits the practical use of $M^{\varepsilon}_{p^*}$.

Algorithmic complexity and the shortest algorithm. A natural definition for the (Kolmogorov) complexity of a function f is the length of the shortest program computing f: $K'(f) := \min_p\{\ell(p) : U(p,x) = f(x)\,\forall x\}$. Unfortunately, K' suffers from not even being approximable, since functional equality of programs is in general undecidable. Let p^* be a formal specification or a program for f. Using $K(p^*)$ is also not a suitable alternative, since it essentially depends on the choice of p^* because, e.g. "dead code" in p^* contributes to $K(p^*)$. A satisfactory solution is to take the length of the shortest program *provably* equivalent to p^*:

$$K''(p^*) \;:=\; \min_p\{\ell(p) : \text{a proof of } [\forall y{:}U(p,y) = U(p^*,y)] \text{ exists}\}.$$

K'' (like K) is upper semicomputable. Let p' be some short description of p^*. We are now concerned with the computation time of p'. Could we get slower and slower algorithms by compressing p^* more and more? Interestingly, this is not the case. Inventing complex (long) programs is *not* necessary to construct asymptotically fast algorithms, under the stated provability assumptions, in contrast to Blum's theorem [Blu67, Blu71]. We show that there exists a program $\tilde{p}$, equivalent to p^* with

$$\begin{aligned} (i) \quad & \ell(\tilde{p}) && \leq K''(p^*) + O(1), \\ (ii) \quad & time_{\tilde{p}}(x) && \leq (1+\varepsilon)\cdot t_p(x) + \tfrac{d_p}{\varepsilon}\cdot time_{t_p}(x) + \tfrac{c_p}{\varepsilon}, \end{aligned}$$

where p is any program provably equivalent to p^* with computation time provably less than $t_p(x)$. That is, $\tilde{p}$ is simultaneously among the shortest *and* fastest programs.

Generalizations. Algorithm $M^{\varepsilon}_{p^*}$ can be modified to handle I/O streams, definable by a Turing machine with unidirectional input and output tapes (and bidirectional work tapes) receiving an input stream and producing an output stream, as is the case in the agent setup.

1.7.2 Time-Bounded AIXI Model

The major drawback of the AIXI model is that it is uncomputable. To overcome this problem, we construct a modified algorithm AIXItl, which is still superior to any other time t and length l bounded agent. The computation