Studies in Logic, Vol. 7, No. 4 (2014): 48–68
PII: 1674-3202(2014)-04-0048-21

Why Inductive Logic Needs Solomonoff Prior?*

Xi Li
Department of Philosophy, Peking University

xli@pku.edu.cn

Abstract. In 1950s, Carnap develops inductive logic to express the degree of confirmation of some hypothesis relative to some evidence. In 1960s, Somlomonoff invents the universal induction method to make prediction. In this paper we integrate the two methods to extent universal induction's expressive power and to enhance inductive logic's predictive power. We introduce Solomonoff prior into inductive logic, and prove the monadic first order logic version of Solomonoff completeness theorem. Then we make some comparison of the two quite different induction methods in the framework of the modified "inductive logic". As is well known, Carnap's λ-continuum fails to confirm universal generalizations. However, in the modified "inductive logic", the proposition "all ravens are black" can be confirmed in any computable world as long as all ravens are really black in that world. If we want to prove the completeness theorem by the method of Solomonoff's universal induction, we have to keep the complete information of the past history in memory. In the modified "inductive logic", we can neglect all of the irrelevant information to concentrate only on some specific pattern, and prove similar convergence results. Even without complete record of relevant information of the specific pattern, we can still build our belief through random sampling.

Background In early 20th century, Keynes tries to assign to inductive generalizations, according to available evidence, probabilities that should converge to 1 as the generalizations are supported by more and more independent events.

In 1950s, Carnap develops the antecedent of the modern inductive logic, in which he tries to use logic to distinguish alternative states of affairs that can be expressed in a given formal language, then define inductive probabilities for sentences by taking advantage of symmetry assumptions concerning such states of affairs. ([1]) In a deductively valid argument, the conclusion is true in every possible world in which the premises are true; while in a good inductive argument, the conclusion can be false in some possible world, but the set of worlds in which the premises are true and the conclusion false should be "small" enough that we can say that the premises confirm the conclusion in some sense.

In 1960s, inspired by Carnap, Solomonoff formalizes Occam's Razor by means of algorithmic information theory, and uses it to construct a universal Bayesian prior for sequence prediction. ([9]) Rathmanner and Hutter give a philosophical analysis of

Received 2014-03-07 *Revision Received* 2014-12-25
*Thanks to Jan Leike for lots of helpful discussions, especially for helping me with the last theorem, and thanks to the anonymous reviewers for valuable feedback.

Solomonoff's universal induction, and claim that "all ravens are black" can be confirmed in the framework of Solomonoff's universal induction if we interpret "black ravens" and "non-black ravens" as a sequence of digits. ([8])

Contribution and Structure of the paper We can't make induction/prediction without taking causality into consideration. Solomonoff formalizes causality with computable functions. If we want to extend Solomonoff's universal induction's expressive power and Carnap's inductive logic's predictive power, we have to combine them together.

We introduce Solomonoff prior into inductive logic to take advantage of its predictive power and extend its expressive power. In the following, we first give a brief introduction to Carnap's inductive logic, and then we introduce Solomonoff prior into inductive logic. After that, We prove the monadic first order logic version of Solomonoff's convergence theorem and make some comparison of the two quite different induction methods in the framework of the modified "inductive logic". As is well known, Carnap's λ-continuum fails to confirm universal generalizations. However, in the modified "inductive logic", "all ravens are black" can be confirmed in any computable world as long as all ravens are really black in that world. You have to keep the complete information of the past history in memory to prove the completeness theorem in the method of Solomonoff's universal induction. In our modified "inductive logic", we can neglect all of the irrelevant information to concentrate only on some specific pattern, and we can still prove similar convergence results. Even without complete record of relevant information of the specific pattern, we can still build our belief by random sampling.

1 An Introduction to Carnap's Inductive Logic

Assume the monadic first order language $\mathscr{L}$ contains countable constants $\mathscr{C}$ and m monadic predicates $\mathscr{R} = \{R_1, R_2, \ldots, R_m\}$ with no function symbols nor equality. The constants $\mathscr{C}$ name all the individuals in some Universe though there is no prior assumption that they necessarily name different individuals.

Notation: $f\restriction_X := \{(x, y) \in f : x \in X\}$, and $f_{-1}(X) := \{x : f(x) \in X\}$.

Definition 1 (Probability on Sentences) A probability on sentences is a non-negative function $w\colon \mathscr{S} \to [0, 1]$ such that

P_1. $\models \psi \Longrightarrow w(\psi) = 1$

P_2. $\psi_1 \models \neg\psi_2 \Longrightarrow w(\psi_1 \vee \psi_2) = w(\psi_1) + w(\psi_2)$

P_3. $w(\exists x \psi(x)) = \lim_{n\to\infty} w\left(\bigvee_{i=1}^{n} \psi(a_i)\right)$

Theorem 1 (Extension Theorem) Suppose $w\colon \mathscr{S}_{QF} \to [0,1]$ over quantifier-free sentences satisfies P_1, P_2, then w has an unique extension to $w^+\colon \mathscr{S} \to [0,1]$ satisfying P_1, P_2, P_3.

Let $Q_i \equiv \bigwedge_{j=1}^{m} \pm R_j$ for $1 \le i \le 2^m =: r$, where $\pm R$ means one of $\{R, \neg R\}$, then $\mathscr{Q} = \{Q_1, \cdots, Q_r\}$ is a r-fold classification system of some Universe with domain $\mathscr{C}$, and every individual in the universe has to satisfy one and only one Q-predicate that is determined by the state description function $h\colon a_i \mapsto Q_{h_i}$. The state description of $\vec{a} = (a_1, \ldots, a_n)$ is $\Theta(\vec{a}) \equiv \bigwedge_{i=1}^{n} Q_{h_i}(a_i)$. The set of state descriptions of $\vec{a}$ is $\mathcal{H}_{\vec{a}} := \{\Theta(\vec{a}) : h\colon \{1, \ldots, n\} \to \{1, \ldots, r\}\}$.

Sometimes we write $n_i := |h\restriction_{\{1,\ldots,n\}}{}_{-1}(i)|$ to denote the number of times that event Q_i occurs in n trials $\bigwedge_{j=1}^{n} Q_{h_j}(a_j)$. Carnap takes $\{n_i : 1 \le i \le r\}$ as the structure description.

Carnap's aim is to find the right w. Carnap believes that the right w should satisfy some symmetry principle. For example, it should be invariant under finite permutations of names.

For any permutation σ of $\mathbb{N}^+$,

$$w(\psi(a_1, \ldots, a_n)) = w(\psi(a_{\sigma(1)}, \ldots, a_{\sigma(n)})) \tag{Ex}$$

For any permutation τ of $\{1, 2, \ldots, r\}$,

$$w\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right) = w\left(\bigwedge_{i=1}^{n} Q_{\tau(h_i)}(a_i)\right) \tag{Ax}$$

Besides the above symmetry principles there is a stronger postulate—*sufficientness postulate*, which asserts that there exists a series of functions $\{f_i : 1 \le i \le r\}$ such that

$$w\left(Q_j(a_{n+1}) \,\middle|\, \bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right) = f_j(n_j, n) \tag{SP}$$

Principle (Ex) asserts that $w\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right)$ depends only on the vector $\langle n_{h_i} : 1 \le i \le n\rangle$, so that it is independent on the order of observing the individuals, while in the presence of (Ex), principle (Ax) asserts that $w\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right)$ depends only on $\{n_i : 1 \le i \le r\}$, and $w(Q_i(a_1)) = 1/r$ for all $1 \le i \le r$.

Considering the principle of indifference—all possibilities that can't be distinguished should be assigned equal probability, there are two intuitive ways to assign prior probability.

(A). All state descriptions have equal weight.
(B). All structure descriptions have equal weight.

Given n individuals, there are r^n possible state descriptions and $\binom{n+r-1}{r-1}$ possible structure descriptions.

According to (A),

$$m^{\dagger}\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right) = \frac{1}{r^n}$$

and

$$\mathfrak{c}^{\dagger}\left(Q_j(a_{n+1}) \,\middle|\, \bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right) = \frac{m^{\dagger}\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i) \wedge Q_j(a_{n+1})\right)}{m^{\dagger}\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right)} = \frac{1}{r}$$

It is independent of the history $\bigwedge_{i=1}^{n} Q_{h_i}(a_i)$, which means it violates the principle of learning from experience and hence is unacceptable.

According to (B),

$$m^*(n_1, \ldots, n_r) = \frac{1}{\binom{n+r-1}{r-1}}$$

Since each structure description can be seen as $\binom{n}{n_1,\ldots,n_r}$ possible state descriptions, and according to principle (Ex), every possible state description shares equal portion of its structure description, so we have

$$m^*\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right) = \frac{m^*(n_1, \ldots, n_r)}{\binom{n}{n_1,\ldots,n_r}} = \frac{1}{\binom{n+r-1}{r-1}\binom{n}{n_1,\ldots,n_r}}$$

which depends only on structure description.

Carnap defines his favorite "degree of confirmation" as

$$\mathfrak{c}^*\left(Q_j(a_{n+1}) \,\middle|\, \bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right) = \frac{m^*\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i) \wedge Q_j(a_{n+1})\right)}{m^*\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right)} = \frac{n_j + 1}{n + r}$$

Carnap's λ-continuum $\mathfrak{c}_{\lambda}$ is a generalization of $\mathfrak{c}^*$.

Suppose $(Q_1, \cdots, Q_r)$ are defined so that they have different relative widths γ_i such that $\sum_{i=1}^{r} \gamma_i = 1$, Carnap's λ-continuum is

$$\mathfrak{c}_{\lambda}\left(Q_j(a_{n+1}) \,\middle|\, \bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right) = \frac{n_j + \lambda\gamma_j}{n + \lambda} = \frac{n}{n+\lambda}\frac{n_j}{n} + \frac{\lambda}{n+\lambda}\gamma_j$$

relative to a free parameter $0 < \lambda \leq \infty$ which indicates the weight given to logical or language-dependent factors over and above purely empirical factors (observed frequencies). The parameter λ serves as an index of caution in singular inductive inference.

Carnap's λ-continuum is invariant under the two symmetry principles (Ex) and (Ax).

The symmetry principles (Ex) and (Ax) says that the temporal order of events is irrelevant, but in reality, the temporal order is of great significance. Carnap's degree of confirmation function is a kind of smoothing method. It approximates frequency when there are enough data. But the frequency philosophy presupposes an independently identical distribution(i.i.d), so it does not work for higher order Markov Chains. Temporal order can't be neglected for any "universal" inductive method. If the temporal order is taken into consideration, the subscript of a should represent the time stamp, and the conjunction $\bigwedge_{i=1}^{n} Q_{h_i}(a_i)$ indicates the time series of observations $\langle Q_{h_1}(a_1), \ldots, Q_{h_n}(a_n)\rangle$.

2 Introduction of Solomonoff Prior into Inductive Logic

In 1963, Putnam takes Carnap's inductive logic as a design for a 'learning machine' —a design for a computing machine that can extrapolate certain kinds of empirical regularities from the data with which it is supplied, and the task of inductive logic is to construct a 'universal learning machine'. ([6, 7]) If there is such a thing as a correct definition of 'degree of confirmation' that can be fixed once and for all, then a machine that predicts in accordance with the degree of confirmation would be a cleverest possible learning machine. Thus any argument against the existence of a cleverest possible learning machine must show either that Carnap's program can't be completely fulfilled or that the correct $\mathfrak{c}$-function must be one that can't be computed by a machine. Either there are better and better $\mathfrak{c}$-functions, but no 'best possible', or else there is a 'best possible' but it is not computable by a machine.

But, Carnap's $\mathfrak{c}$-function depends on the language, then what is the correct language? What is the correct 'degree of confirmation'? How much evidence is strong enough to hold our belief?

By choosing the smallest model class that contains the true environment and the universal (mixture) prior beliefs of the environments that reflect the simplicity criterion, Solomonoff solves the problem. Carnap's "degree of confirmation" function can be made better and better, while Solomonoff's "degree of confirmation" function is the best but not computable. ([9])

Solonomoff Prior

We assume the reader is familiar with the basics of Kolmogorov Complexity. Preliminaries can be found in [5] or [3]. Some basic concepts are given below.

Notation

$$[\![\phi]\!] := \begin{cases} 1 & \text{if } \phi \text{ is true} \\ 0 & \text{otherwise} \end{cases}$$

Definition 2 (**Kolmogorov Complexity**)

$$\begin{aligned} K(x|y) &:= \min_p \{|p| : U(\langle p, y \rangle) = x\} \\ K(x) &:= K(x|\epsilon) \\ K(f) &:= \min_p \{|p| : \forall x \, (U(\langle p, x \rangle) \simeq f(x))\} \end{aligned}$$

where U is a universal prefix Turing machine, $\min \emptyset = \infty$.

For non-string objects o we define $K(o) := K(\langle o \rangle)$, where $\langle o \rangle \in \mathcal{X}^*$ is some standard code for o.

Definition 3 (**Monotone Complexity**)

$$Km(x) := \min_p \{|p| : U(p) = x*\} \tag{1}$$

where U is a universal monotone Turing machine.

Let $\mathcal{M}_U$ be the set of *enumerable semimeasures*.

Every specific state description function h determines an unique state description — an unique universe, in other word, every universe is generated by some program. We identify $h_{1:n}$ with history $\bigwedge_{i=1}^{n} Q_{h_i}(a_i)$ and identify p with the universe $h_{1:\infty}$ if $U(p) = h_{1:\infty}$ without confusion.

Generally speaking, our beliefs and hence probabilities are a result of our personal history. In order to update our beliefs consistently we must first generate the set of all explanations that may be possible. The actual universe is just one of a large number of possible universes. Every universe develops in a sequence of possible states; the probability assigned to each state should be the proportion of the possible universes in which that state is attained, if we weigh all of the possible universes equally. Each new observation/measurement eliminates some fraction of the universes in a given state, depending on how likely or unlikely that state was to actually produce that observation/measurement; the surviving universes then gain a new posterior probability distribution, which is related to the prior distribution by Bayes' formula.

Definition 4 (**Universal Probability**)

$$\mathfrak{c}_M\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right) = \sum_{p:U(p)=h_{1:n}*} 2^{-|p|}$$

where U is a universal monotone Turing machine.

It can be regarded as the limit of the relative frequency of the consistent possible worlds over all possible worlds:

$$\begin{aligned}\mathfrak{c}_M\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right) &= \sum_{p} 2^{-|p|}[\![U(p) = h_{1:n}*]\!] \\ &= \lim_{n\to\infty} \frac{\sum\limits_{p:|p|\leq n} 2^{n-|p|}\left[\!\!\left[\bigwedge\limits_{i=1}^{n} Q_{\langle U(p)\rangle_i}(a_i) \equiv \bigwedge\limits_{i=1}^{n} Q_{h_i}(a_i)\right]\!\!\right]}{2^n} \\ &\approx \lim_{n\to\infty} \frac{\left|\left\{p : |p| = n \ \ \& \ \ \bigwedge\limits_{i=1}^{n} Q_{\langle U(p)\rangle_i}(a_i) \equiv \bigwedge\limits_{i=1}^{n} Q_{h_i}(a_i)\right\}\right|}{2^n}\end{aligned}$$

It means that $\mathfrak{c}_M\left(\bigwedge\limits_{i=1}^{n} Q_{h_i}(a_i)\right)$ is the frequentist probability that the program of a universal monotone Turing machine U generates $\bigwedge\limits_{i=1}^{n} Q_{h_i}(a_i)$ when provided with uniform random noise (fair coin flips) on the input tape.

$$probability = \frac{|\text{consistent universes}|}{|\text{all possible universes}|}$$

Just like Carnap's (A) and (B), we also turn to the principle of indifference for help. However, we use the principle of indifference on the level of the causes of the phenomena rather than on the level of the phenomena themselves.

Lemma 1 ([3]) For every $\nu \in \mathcal{M}_U$ there exists some monotone Turing machine T such that

$$\nu\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right) = \sum_{p:T(p)=h_{1:n}*} 2^{-|p|} \quad \text{and} \quad K(\nu) \stackrel{+}{=} |\langle T\rangle|$$

where $T(p) = U(\langle T\rangle p)$.

Lemma 2 For $\nu \in \mathcal{M}_U$,

$$\mathfrak{c}_M\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right) \stackrel{+}{\geq} 2^{-K(\nu)}\nu\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right)$$

Proof

$$\mathfrak{c}_M\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right) = \sum_{p:U(p)=h_{1:n}*} 2^{-|p|}$$

$$\begin{aligned}
&\geq \sum_{q:U(\langle T\rangle q)=h_{1:n}*} 2^{-|\langle T\rangle q|} \\
&= 2^{-|\langle T\rangle|} \sum_{q:T(q)=h_{1:n}*} 2^{-|q|} \\
&\stackrel{\times}{=} 2^{-K(\nu)} \nu \left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i) \right) \qquad \text{[Lemma 1]}
\end{aligned}$$

□

Define

$$\mathfrak{c}'_M(\top) := 1$$

$$\mathfrak{c}'_M\left(\bigwedge_{i=1}^{t} Q_{h_i}(a_i)\right) := \mathfrak{c}'_M\left(\bigwedge_{i=1}^{t-1} Q_{h_i}(a_i)\right) \frac{\mathfrak{c}_M\left(\bigwedge_{i=1}^{t} Q_{h_i}(a_i)\right)}{\sum_{1\leq k\leq r} \mathfrak{c}_M\left(\bigwedge_{i=1}^{t-1} Q_{h_i}(a_i) \wedge Q_k(a_t)\right)}$$

$$= \frac{\mathfrak{c}_M\left(\bigwedge_{i=1}^{t} Q_{h_i}(a_i)\right)}{\mathfrak{c}_M(\top)} \prod_{i=1}^{t} \frac{\mathfrak{c}_M\left(\bigwedge_{j=1}^{i-1} Q_{h_j}(a_j)\right)}{\sum_{1\leq k\leq r} \mathfrak{c}_M\left(\bigwedge_{j=1}^{i-1} Q_{h_j}(a_j) \wedge Q_k(a_i)\right)}$$

Obviously, for any state description Θ, Θ',

(i). $\mathfrak{c}'_M\left(\Theta(a_1, \ldots, a_n)\right) \geq 0$
(ii). $\mathfrak{c}'_M(\top) = 1$
(iii). $\mathfrak{c}'_M\left(\Theta(a_1, \ldots, a_n)\right) = \sum_{\Theta'(a_1,\ldots,a_{n+1})\models\Theta(a_1,\ldots,a_n)} \mathfrak{c}'_M\left(\Theta'(a_1, \ldots, a_{n+1})\right)$

For any quantifier-free sentence $\psi(\vec{a})$, let

$$\mathfrak{c}'_M\left(\psi(\vec{a})\right) := \sum_{\Theta(\vec{b})\models\psi(\vec{a})} \mathfrak{c}'_M\left(\Theta(\vec{b})\right)$$

where $\vec{b}$ is sufficiently large that all of the $\vec{a}$ are amongst $\vec{b}$, and $\bigvee_{\Theta(\vec{b})\models\psi(\vec{a})} \Theta(\vec{b})$ is the *full disjunctive normal form* of $\psi(\vec{a})$.

$$\psi(\vec{a}) \equiv \bigvee_{\Theta(\vec{b})\models\psi(\vec{a})} \Theta(\vec{b}) \qquad \text{(DNF)}$$

It is easy to see, $\mathfrak{c}'_M$ satisfies P_1, P_2, and according to Theorem (1), $\mathfrak{c}'_M$ has an unique extension over all of the sentences $\mathscr{S}$ of $\mathscr{L}$. Then $\mathfrak{c}'_M$ induces a confirmation function by the conditional probability

$$c'_M(\psi_H|\psi_E) = \frac{\mathfrak{c}'_M(\psi_E \wedge \psi_H)}{\mathfrak{c}'_M(\psi_E)}$$

In fact, any w satisfying (i),(ii),(iii) can extend to a probability function on $\mathscr{L}$.

Following the binary digital version of completeness theorem ([3, 9]), we can prove the following monadic first order version of completeness theorem (2,3). There will be a finite bound on the number of prediction errors we have to pay out over an infinite sequence prediction. In other words, given enough computing resources, an agent will learn to predict very well as long as the universe is computable (2) or it is generated by some computable measure (3).

Theorem 2 If the universe is deterministic, $\mathfrak{c}'_M$ can predict the future very well with few errors.

$$\sum_{t=1}^{\infty}\left|1-\mathfrak{c}'_M\left(Q_{h_t}(a_t)\,\middle|\,\bigwedge_{i=1}^{t-1}Q_{h_i}(a_i)\right)\right| \leq Km(h_{1:\infty})\ln 2$$

Proof

$$\begin{aligned}
&\sum_{t=1}^{\infty}\left|1-\mathfrak{c}'_M\left(Q_{h_t}(a_t)\,\middle|\,\bigwedge_{i=1}^{t-1}Q_{h_i}(a_i)\right)\right| \\
&\overset{(a)}{\leq} -\sum_{t=1}^{\infty}\ln\mathfrak{c}'_M\left(Q_{h_t}(a_t)\,\middle|\,\bigwedge_{i=1}^{t-1}Q_{h_i}(a_i)\right) \\
&= -\ln\prod_{t=1}^{\infty}\mathfrak{c}'_M\left(Q_{h_t}(a_t)\,\middle|\,\bigwedge_{i=1}^{t-1}Q_{h_i}(a_i)\right) \\
&= -\lim_{n\to\infty}\ln\mathfrak{c}'_M\left(\bigwedge_{i=1}^{n}Q_{h_i}(a_i)\right) \\
&\leq -\lim_{n\to\infty}\ln\mathfrak{c}_M\left(\bigwedge_{i=1}^{n}Q_{h_i}(a_i)\right) \\
&\leq Km(h_{1:\infty})\ln 2
\end{aligned}$$

where $\overset{(a)}{\leq}$ follows from $1-x+\ln x\leq 0$ for $x\in[0,1]$. □

Theorem 3 (**Completeness Theorem**) For histories generated by a computable stochastic process μ, the following bound holds.

$$\sum_{l=1}^{\infty}\sum_{h_{1:l}\in\{1,\ldots,r\}^l}\mu\left(\bigwedge_{i=1}^{l}Q_{h_i}(a_i)\right)\cdot$$

$$\left(\sum_{h_{l+1:t}\in\{1,\ldots,r\}^{t-l}}\left|\mathfrak{c}'_M\left(\bigwedge_{i=l+1}^{t} Q_{h_i}(a_i)\,\middle|\,\bigwedge_{i=1}^{l} Q_{h_i}(a_i)\right)-\mu\left(\bigwedge_{i=l+1}^{t} Q_{h_i}(a_i)\,\middle|\,\bigwedge_{i=1}^{l} Q_{h_i}(a_i)\right)\right|\right)^2$$
$$\stackrel{+}{\leq} 2(t-l)K(\mu)\ln 2$$
$$<\infty$$

Proof

$$\sum_{l=1}^{\infty}\sum_{h_{1:l}\in\{1,\ldots,r\}^l}\mu\left(\bigwedge_{i=1}^{l} Q_{h_i}(a_i)\right)\cdot$$
$$\left(\sum_{h_{l+1:t}\in\{1,\ldots,r\}^{t-l}}\left|\mathfrak{c}'_M\left(\bigwedge_{i=l+1}^{t} Q_{h_i}(a_i)\,\middle|\,\bigwedge_{i=1}^{l} Q_{h_i}(a_i)\right)-\mu\left(\bigwedge_{i=l+1}^{t} Q_{h_i}(a_i)\,\middle|\,\bigwedge_{i=1}^{l} Q_{h_i}(a_i)\right)\right|\right)^2$$
$$\stackrel{(a)}{\leq} 2\sum_{l=1}^{\infty}\sum_{h_{1:l}\in\{1,\ldots,r\}^l}\mu\left(\bigwedge_{i=1}^{l} Q_{h_i}(a_i)\right)\cdot$$
$$\sum_{h_{l+1:t}\in\{1,\ldots,r\}^{t-l}}\mu\left(\bigwedge_{i=l+1}^{t} Q_{h_i}(a_i)\,\middle|\,\bigwedge_{i=1}^{l} Q_{h_i}(a_i)\right)\ln\frac{\mu\left(\bigwedge_{i=l+1}^{t} Q_{h_i}(a_i)\,\middle|\,\bigwedge_{i=1}^{l} Q_{h_i}(a_i)\right)}{\mathfrak{c}'_M\left(\bigwedge_{i=l+1}^{t} Q_{h_i}(a_i)\,\middle|\,\bigwedge_{i=1}^{l} Q_{h_i}(a_i)\right)}$$
$$=2\sum_{l=1}^{\infty}\sum_{h_{1:t}\in\{1,\ldots,r\}^l}\mu\left(\bigwedge_{i=1}^{t} Q_{h_i}(a_i)\right)\sum_{m=l}^{t-1}\ln\frac{\mu\left(Q_{h_{m+1}}(a_{m+1})\,\middle|\,\bigwedge_{i=1}^{m} Q_{h_i}(a_i)\right)}{\mathfrak{c}'_M\left(Q_{h_{m+1}}(a_{m+1})\,\middle|\,\bigwedge_{i=1}^{m} Q_{h_i}(a_i)\right)}$$
$$=2\sum_{l=1}^{\infty}\sum_{m=l}^{t-1}\sum_{h_{1:m+1}\in\{1,\ldots,r\}^{m+1}}\mu\left(\bigwedge_{i=1}^{m+1} Q_{h_i}(a_i)\right)\cdot$$
$$\left(\sum_{h_{m+2:t}\in\{1,\ldots,r\}^{t-m-1}}\mu\left(\bigwedge_{i=m+2}^{t} Q_{h_i}(a_i)\,\middle|\,\bigwedge_{i=1}^{m+1} Q_{h_i}(a_i)\right)\right)\ln\frac{\mu\left(Q_{h_{m+1}}(a_{m+1})\,\middle|\,\bigwedge_{i=1}^{m} Q_{h_i}(a_i)\right)}{\mathfrak{c}'_M\left(Q_{h_{m+1}}(a_{m+1})\,\middle|\,\bigwedge_{i=1}^{m} Q_{h_i}(a_i)\right)}$$
$$\leq 2(t-l)\sum_{m=1}^{\infty}\sum_{h_{1:m+1}\in\{1,\ldots,r\}^{m+1}}\mu\left(\bigwedge_{i=1}^{m+1} Q_{h_i}(a_i)\right)\ln\frac{\mu\left(Q_{h_{m+1}}(a_{m+1})\,\middle|\,\bigwedge_{i=1}^{m} Q_{h_i}(a_i)\right)}{\mathfrak{c}'_M\left(Q_{h_{m+1}}(a_{m+1})\,\middle|\,\bigwedge_{i=1}^{m} Q_{h_i}(a_i)\right)}$$
$$=2(t-l)\lim_{n\to\infty}\sum_{m=1}^{n}\sum_{h_{1:m+1}\in\{1,\ldots,r\}^{m+1}}\left(\sum_{h_{m+2:n}\in\{1,\ldots,r\}^{n-m-1}}\mu\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right)\right)\cdot$$

$$
\begin{aligned}
&\ln \frac{\mu\left(Q_{h_{m+1}}(a_{m+1}) \,\middle|\, \bigwedge_{i=1}^{m} Q_{h_i}(a_i)\right)}{\mathfrak{c}'_M\left(Q_{h_{m+1}}(a_{m+1}) \,\middle|\, \bigwedge_{i=1}^{m} Q_{h_i}(a_i)\right)} \\
&= 2(t-l) \lim_{n\to\infty} \sum_{m=1}^{n} \sum_{h_{1:n}\in\{1,\ldots,r\}^n} \mu\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right) \ln \frac{\mu\left(Q_{h_{m+1}}(a_{m+1}) \,\middle|\, \bigwedge_{i=1}^{m} Q_{h_i}(a_i)\right)}{\mathfrak{c}'_M\left(Q_{h_{m+1}}(a_{m+1}) \,\middle|\, \bigwedge_{i=1}^{m} Q_{h_i}(a_i)\right)} \\
&= 2(t-l) \lim_{n\to\infty} \sum_{h_{1:n}\in\{1,\ldots,r\}^n} \mu\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right) \sum_{m=1}^{n} \ln \frac{\mu\left(Q_{h_{m+1}}(a_{m+1}) \,\middle|\, \bigwedge_{i=1}^{m} Q_{h_i}(a_i)\right)}{\mathfrak{c}'_M\left(Q_{h_{m+1}}(a_{m+1}) \,\middle|\, \bigwedge_{i=1}^{m} Q_{h_i}(a_i)\right)} \\
&= 2(t-l) \lim_{n\to\infty} \sum_{h_{1:n}\in\{1,\ldots,r\}^n} \mu\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right) \ln \prod_{m=1}^{n} \frac{\mu\left(Q_{h_{m+1}}(a_{m+1}) \,\middle|\, \bigwedge_{i=1}^{m} Q_{h_i}(a_i)\right)}{\mathfrak{c}'_M\left(Q_{h_{m+1}}(a_{m+1}) \,\middle|\, \bigwedge_{i=1}^{m} Q_{h_i}(a_i)\right)} \\
&= 2(t-l) \lim_{n\to\infty} \sum_{h_{1:n}\in\{1,\ldots,r\}^n} \mu\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right) \ln \frac{\mu\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right)}{\mathfrak{c}'_M\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right)} \\
&\leq 2(t-l) \lim_{n\to\infty} \sum_{h_{1:n}\in\{1,\ldots,r\}^n} \mu\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right) \ln \frac{\mu\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right)}{\mathfrak{c}_M\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right)} \\
&\overset{+}{\leq} 2(t-l)K(\mu)\ln 2 \\
&< \infty
\end{aligned}
$$

where $\overset{(a)}{\leq}$ follows from Entropy Inequality (Theorem 3.19-vi, in [3]), and the last inequality $\overset{+}{\leq}$ follows from Lemma (2). □

Compare Carnap's λ-continuum $\mathfrak{c}_\lambda$ with $\mathfrak{c}'_M$. With zero-knowledge ($n = 0$), Carnap would use

$$\mathfrak{c}_\lambda\left(Q_j(a_1)\right) = \frac{0+\lambda\gamma_j}{0+\lambda} = \gamma_j$$

to estimate the future, while Solomonoff would prefer

$$\mathfrak{c}'_M\left(Q_{h_1}(a_1)\right) = \frac{\mathfrak{c}_M\left(Q_{h_1}(a_1)\right)}{\sum_{1\leq j\leq r} \mathfrak{c}_M\left(Q_j(a_1)\right)}$$

with sufficient experiences (n large enough), Carnap would use

$$\mathfrak{c}_\lambda\left(Q_j(a_{n+1}) \,\middle|\, \bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right) = \frac{n_j + \lambda\gamma_j}{n+\lambda} \approx \frac{n_j}{n}$$

—the frequency of the phenomena, while Solomonoff would still use

$$\mathfrak{c}'_M\left(Q_j(a_{n+1}) \,\middle|\, \bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right)$$

—the normalization of the frequency of the consistent universes/causes. The frequency of the phenomena does not always converge to the true probability. It depends on the environment. If the true environment is independent and identically distributed, c'_M can also converge to the limit of frequency according to the completeness theorem (3). In other words, Carnap would like to *know how* while Solomonoff would like to *know why*.

What's more? Carnap's λ-continuum $\mathfrak{c}_\lambda$ fails to confirm "all ravens are black" while $\mathfrak{c}'_M$ is qualified as a solution.

3 How to Confirm *"All Ravens are Black"*?

In our monadic first order language, "All ravens are black" can be expressed by $\forall x(R(x) \to B(x))$.

Since

$$\begin{aligned} w(\forall x\psi(x)) &= 1 - w(\exists x \neg\psi(x)) \\ &= 1 - \lim_{n\to\infty} w\left(\bigvee_{i=1}^{n} \neg\psi(a_i)\right) \\ &= \lim_{n\to\infty}\left(1 - w\left(\bigvee_{i=1}^{n} \neg\psi(a_i)\right)\right) \\ &= \lim_{n\to\infty} w\left(\bigwedge_{i=1}^{n} \psi(a_i)\right) \end{aligned}$$

Hence, to solve the Raven Paradox, we only have to make sure that

$$\lim_{n\to\infty} w\left(\bigwedge_{i=1}^{n}(R(a_i) \to B(a_i))\right) > 0$$

Now we prove the above inequality holds for $\mathfrak{c}'_M$.

Lemma 3 If there exists a computable universe that contains only black ravens, then $\mathfrak{c}'_M\,(\forall x(R(x) \to B(x)) > 0$ is true.

Proof

$$
\begin{aligned}
&\mathfrak{c}'_M\left(\forall x(R(x) \to B(x)\right)\\
&= \lim_{n\to\infty} \mathfrak{c}'_M\left(\bigwedge_{i=1}^{n}(R(a_i) \to B(a_i))\right)\\
&= \lim_{n\to\infty} \mathfrak{c}'_M\left(\bigwedge_{i=1}^{n}(\neg R(a_i) \vee B(a_i))\right)\\
&= \lim_{n\to\infty} \mathfrak{c}'_M\left(\bigvee_{j=1}^{2^n}\left(\bigwedge_{i=1}^{n}(\neg R(a_i)/B(a_i))\right)_j\right)\\
&= \lim_{n\to\infty} \mathfrak{c}'_M\left(\bigvee_{\bigwedge_{i=1}^{n} Q_{h_i}(a_i) \models \bigvee_{j=1}^{2^n}\left(\bigwedge_{i=1}^{n}(\neg R(a_i)/B(a_i))\right)_j} \bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right)\\
&= \lim_{n\to\infty} \sum_{\bigwedge_{i=1}^{n} Q_{h_i}(a_i) \models \bigvee_{j=1}^{2^n}\left(\bigwedge_{i=1}^{n}(\neg R(a_i)/B(a_i))\right)_j} \mathfrak{c}'_M\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right)\\
&= \lim_{n\to\infty} \sum_{\bigwedge_{i=1}^{n} Q_{h_i}(a_i) \models \bigvee_{j=1}^{2^n}\left(\bigwedge_{i=1}^{n}(\neg R(a_i)/B(a_i))\right)_j} \frac{\mathfrak{c}_M\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right)}{\mathfrak{c}_M(\top)}.\\
&\prod_{i=1}^{n} \frac{\mathfrak{c}_M\left(\bigwedge_{j=1}^{i-1} Q_{h_j}(a_j)\right)}{\sum_{1\le k\le r} \mathfrak{c}_M\left(\bigwedge_{j=1}^{i-1} Q_{h_j}(a_j) \wedge Q_k(a_i)\right)}\\
&= \lim_{n\to\infty} \sum_{\bigwedge_{i=1}^{n} Q_{h_i}(a_i) \models \bigvee_{j=1}^{2^n}\left(\bigwedge_{i=1}^{n}(\neg R(a_i)/B(a_i))\right)_j} \frac{\sum_{p:U(p)=h_{1:n}*} 2^{-|p|}}{\sum_{p\in dom(U)} 2^{-|p|}}\\
&\prod_{i=1}^{n} \frac{\sum_{p:U(p)=h_{<i}*} 2^{-|p|}}{\sum_{1\le k\le r} \sum_{p:U(p)=h_{<i}k*} 2^{-|p|}}
\end{aligned}
$$

where ψ_1/ψ_2 means that we mutual exclusively choose either ψ_1 or ψ_2.

Since

$$\forall i: \quad \frac{\sum\limits_{p:U(p)=h_{<i}*} 2^{-|p|}}{\sum\limits_{1\leq k\leq r}\sum\limits_{p:U(p)=h_{<i}k*} 2^{-|p|}} \geq 1$$

hence, if there exists some computable universe $h_{1:\infty}$ such that

$$\forall n\left(\bigwedge_{i=1}^{n} Q_{h_i}(a_i) \models \bigwedge_{i=1}^{n}(\neg R(a_i)/B(a_i))\right)$$

then $\mathfrak{c}'_M\left(\forall x(R(x)\rightarrow B(x)\right) > 0$ is true. □

In other words, if all of the ravens in a computable universe are black, $\mathfrak{c}'_M$ can confirm that “all ravens are black”.

$$\begin{aligned}
&\lim_{n\to\infty}\mathfrak{c}'_M\left(\forall x(R(x)\rightarrow B(x)) \,\middle|\, \bigwedge_{i=1}^{n}(\neg R(a_i)\vee B(a_i))\right)\\
&=\lim_{n\to\infty}\frac{\mathfrak{c}'_M\left(\forall x(R(x)\rightarrow B(x)\right)}{\mathfrak{c}'_M\left(\bigwedge\limits_{i=1}^{n}(R(a_i)\rightarrow B(a_i))\right)}\\
&=\frac{\mathfrak{c}'_M\left(\forall x(R(x)\rightarrow B(x)\right)}{\lim\limits_{n\to\infty}\mathfrak{c}'_M\left(\bigwedge\limits_{i=1}^{n}(R(a_i)\rightarrow B(a_i))\right)}\\
&=\frac{\mathfrak{c}'_M\left(\forall x(R(x)\rightarrow B(x)\right)}{\mathfrak{c}'_M\left(\forall x(R(x)\rightarrow B(x)\right)} \qquad \text{[Lemma 3]}\\
&=1
\end{aligned}$$

So far we know $\mathfrak{c}'_M$ can confirm “all ravens are black”. On the other hand, as for Carnap’s λ-continuum, let’s see why $\mathfrak{c}_\lambda\left(\forall x(R(x)\rightarrow B(x)\right) > 0$ fails.

$$\begin{aligned}
&\mathfrak{c}_\lambda\left(\forall x(R(x)\rightarrow B(x)\right)\\
&=\lim_{n\to\infty}\sum_{\bigwedge\limits_{i=1}^{n}Q_{h_i}(a_i)\models\bigvee\limits_{j=1}^{2^n}\left(\bigwedge\limits_{i=1}^{n}(\neg R(a_i)/B(a_i))\right)_j}\mathfrak{c}_\lambda\left(\bigwedge_{i=1}^{n}Q_{h_i}(a_i)\right)\\
&=\lim_{n\to\infty}\sum_{\bigwedge\limits_{i=1}^{n}Q_{h_i}(a_i)\models\bigvee\limits_{j=1}^{2^n}\left(\bigwedge\limits_{i=1}^{n}(\neg R(a_i)/B(a_i))\right)_j}\prod_{i=0}^{n-1}\mathfrak{c}_\lambda\left(Q_{h_{i+1}}(a_{i+1}) \,\middle|\, \bigwedge_{j=1}^{i}Q_{h_j}(a_j)\right)\\
&=\lim_{n\to\infty}\sum_{\bigwedge\limits_{i=1}^{n}Q_{h_i}(a_i)\models\bigvee\limits_{j=1}^{2^n}\left(\bigwedge\limits_{i=1}^{n}(\neg R(a_i)/B(a_i))\right)_j}\prod_{i=0}^{n-1}\frac{i_{h_{i+1}}+\lambda\gamma_{h_{i+1}}}{i+\lambda}
\end{aligned}$$

$$\leq \lim_{n\to\infty} \prod_{i=0}^{n-1} \frac{i+\lambda(1-\min\limits_{1\leq t\leq r}\gamma_t)}{i+\lambda}$$
$$=0$$

The last step follows from

$$\prod_{n\geq 1} a_n = 0 \iff \sum_{n\geq 1}(1-a_n)=\infty \quad \text{for} \quad \forall n:\ 0<a_n\leq 1$$

The reason that $\mathfrak{c}_\lambda$ fails to confirm universal generalization is that the speed of convergence is too slow.

$$\sum_{i=0}^{\infty} \frac{\lambda\cdot\min\limits_{1\leq t\leq r}\gamma_t}{i+\lambda}=\infty$$

It seems that it is easy to rectify. If we use

$$\mathfrak{c}_\lambda^\delta\left(Q_j(a_{n+1})\;\middle|\;\bigwedge_{i=1}^{n} Q_{h_i}(a_i)\right) := \frac{\dfrac{n_j^{1+\delta}+\lambda\gamma_j}{n^{1+\delta}+\lambda}}{\sum\limits_{1\leq k\leq r}\dfrac{n_k^{1+\delta}+\lambda\gamma_k}{n^{1+\delta}+\lambda}} = \frac{n_j^{1+\delta}+\lambda\gamma_j}{\sum\limits_{1\leq k\leq r} n_k^{1+\delta}+\lambda}$$

rather than $\mathfrak{c}_\lambda$, then the convergence is guaranteed. However, although $\mathfrak{c}'_\lambda$ agrees with the principle (Ex), yet it violates the postulate (SP). If the δ is small enough, it can even agree with the frequency interpretation.

Actually, there are several approaches to confirm universal generalizations, for example, [2] and [10]. Hintikka's methods is very complicated. It depends on how many Q-predicates are non-empty. Zabell's methods is much simpler. He modifies the sufficientness postulate (SP) a little bit, which makes it seem a little bit ad hoc. Neither of their methods agree with the sufficientness postulate (SP). It seems that the difficulty focuses on the sufficientness postulate, however, sufficientness postulate is not the crux of the problem. Carnap's methods, as well as other similar methods, misses something very important for the "causality". The temporal order should not be neglected by any inductive logic. $\mathfrak{c}'_M$ tries to characterize causality with computable functions, as a byproduct, the proposition "all ravens are black" gets confirmed.

4 The Prediction of some specific pattern with Random Sampling

The Prediction of some specific pattern According to theorem (3), given complete records of past history, we can predict the next states of our universe. However, we can't keep all of the past information in memory in practice. Usually we only focus on some specific pattern of the universe. For example, when we talk about whether ravens are black we do not care about whether it will rain tomorrow or whether dog barks. Can we predict whether some pattern will last in the next state without recording the complete information of the past history? Theorem (4) assures us that we can concentrate only on some specific pattern and predict well.

For any formula $\phi(x)$, we write $\phi(\vec{a}) \equiv \bigwedge_{i=1}^{n} \phi(a_i/x)$ without confusion, and $a_{t:n} := a_t a_{t+1} \ldots a_{n-1} a_n$, and $a_{<n} := a_1 \ldots a_{n-1}$. Hence, $\phi(a_{1:t})$ means $\bigwedge_{i=1}^{t} \phi(a_i/x)$.

Lemma 4 For probability measure p and q, we have

$$\sum_x (p(x) - q(x))^2 \leq 2D(p\|q)$$

where $D(p\|q) := \sum_x p(x) \ln \frac{p(x)}{q(x)}$.

Proof Let $f(p) := -p \ln p$, then we have $f'(p) = -\ln p - 1$ and $f''(p) = -\frac{1}{p}$.

According to Taylor's theorem, there exists some r between p and q such that

$$f(p) \leq f(q) + \frac{f'(q)(p-q)}{1!} + \frac{f''(r)(p-q)^2}{2!}$$

insert f, f', f'' to the above formula, we have,

$$(p-q)^2 \leq 2r \left(p \ln \frac{p}{q} + q - p \right)$$

therefore,

$$\begin{aligned}
\sum_x (p(x) - q(x))^2 &\leq 2 \sum_x r(x) \left(p(x) \ln \frac{p(x)}{q(x)} + q(x) - p(x) \right) \\
&\leq 2 \sum_x \left(p(x) \ln \frac{p(x)}{q(x)} + q(x) - p(x) \right) \\
&= 2D(p\|q)
\end{aligned}$$

□

Lemma 5 For $\nu \in \mathcal{M}_U$, and for any sentence ϕ,

$$\mathfrak{c}'_M(\phi) \overset{+}{\geq} 2^{-K(\mu)} \mu(\phi)$$

Proof

$$\mathfrak{c}'_M(\phi) = \sum_{\Theta \models \phi} \mathfrak{c}'_M(\Theta)$$

$$\geq \sum_{\Theta \models \phi} \mathfrak{c}_M(\Theta)$$

$$\stackrel{+}{\geq} \sum_{\Theta \models \phi} 2^{-K(\mu)} \mu(\Theta) \qquad \text{[Lemma2]}$$

$$= 2^{-K(\mu)} \mu(\phi)$$

□

Theorem 4 (Convergence Theorem) For $\nu \in \mathcal{M}_U$, and for any formula ϕ,

$$\sum_{t=1}^{\infty} \sum_{\phi(a_{<t})} \mu(\phi(a_{<t})) \sum_{\phi(a_{t:k})} \left(\mathfrak{c}'_M \left(\phi(a_{t:k}) \,\middle|\, \phi(a_{<t}) \right) - \mu \left(\phi(a_{t:k}) \,\middle|\, \phi(a_{<t}) \right) \right)^2 < \infty$$

Proof

$$\sum_{t=1}^{\infty} \sum_{\phi(a_{<t})} \mu(\phi(a_{<t})) \sum_{\phi(a_{t:k})} \left(\mathfrak{c}'_M \left(\phi(a_{t:k}) \,\middle|\, \phi(a_{<t}) \right) - \mu \left(\phi(a_{t:k}) \,\middle|\, \phi(a_{<t}) \right) \right)^2$$

$$\stackrel{(a)}{\leq} 2 \sum_{t=1}^{\infty} \sum_{\phi(a_{<t})} \mu(\phi(a_{<t})) \sum_{\phi(a_{t:k})} \mu \left(\phi(a_{t:k}) \,\middle|\, \phi(a_{<t}) \right) \ln \frac{\mu \left(\phi(a_{t:k}) \,\middle|\, \phi(a_{<t}) \right)}{\mathfrak{c}'_M \left(\phi(a_{t:k}) \,\middle|\, \phi(a_{<t}) \right)}$$

$$= 2 \sum_{t=1}^{\infty} \sum_{\phi(a_{1:k})} \mu(\phi(a_{1:k})) \sum_{m=t}^{k} \ln \frac{\mu \left(\phi(a_{1:m}) \,\middle|\, \phi(a_{<m}) \right)}{\mathfrak{c}'_M \left(\phi(a_{1:m}) \,\middle|\, \phi(a_{<m}) \right)}$$

$$= 2 \sum_{t=1}^{\infty} \sum_{m=t}^{k} \sum_{\phi(a_{<m})} \mu(\phi(a_{<m})) \left(\sum_{\phi(a_{m:k})} \mu \left(\phi(a_{m:k}) \,\middle|\, \phi(a_{<m}) \right) \right) \ln \frac{\mu \left(\phi(a_{1:m}) \,\middle|\, \phi(a_{<m}) \right)}{\mathfrak{c}'_M \left(\phi(a_{1:m}) \,\middle|\, \phi(a_{<m}) \right)}$$

$$\leq 2(k-t) \sum_{m=1}^{\infty} \sum_{\phi(a_{<m})} \mu(\phi(a_{<m})) \ln \frac{\mu \left(\phi(a_{1:m}) \,\middle|\, \phi(a_{<m}) \right)}{\mathfrak{c}'_M \left(\phi(a_{1:m}) \,\middle|\, \phi(a_{<m}) \right)}$$

$$= 2(k-t) \lim_{n \to \infty} \sum_{m=1}^{n} \sum_{\phi(a_{<m})} \left(\sum_{\phi(a_{m:n})} \mu(\phi(a_{1:n})) \right) \ln \frac{\mu \left(\phi(a_{1:m}) \,\middle|\, \phi(a_{<m}) \right)}{\mathfrak{c}'_M \left(\phi(a_{1:m}) \,\middle|\, \phi(a_{<m}) \right)}$$

$$= 2(k-t) \lim_{n \to \infty} \sum_{m=1}^{n} \sum_{\phi(a_{1:n})} \mu(\phi(a_{1:n})) \ln \frac{\mu \left(\phi(a_{1:m}) \,\middle|\, \phi(a_{<m}) \right)}{\mathfrak{c}'_M \left(\phi(a_{1:m}) \,\middle|\, \phi(a_{<m}) \right)}$$

$$
\begin{aligned}
&= 2(k-t)\lim_{n\to\infty}\sum_{\phi(a_{1:n})}\mu\left(\phi(a_{1:n})\right)\sum_{m=1}^{n}\ln\frac{\mu\left(\phi(a_{1:m})\,\middle|\,\phi(a_{<m})\right)}{\mathfrak{c}'_M\left(\phi(a_{1:m})\,\middle|\,\phi(a_{<m})\right)}\\
&= 2(k-t)\lim_{n\to\infty}\sum_{\phi(a_{1:n})}\mu\left(\phi(a_{1:n})\right)\ln\prod_{m=1}^{n}\frac{\mu\left(\phi(a_{1:m})\,\middle|\,\phi(a_{<m})\right)}{\mathfrak{c}'_M\left(\phi(a_{1:m})\,\middle|\,\phi(a_{<m})\right)}\\
&= 2(k-t)\lim_{n\to\infty}\sum_{\phi(a_{1:n})}\mu\left(\phi(a_{1:n})\right)\ln\frac{\mu\left(\phi(a_{1:n})\right)}{\mathfrak{c}'_M\left(\phi(a_{1:n})\right)}\\
&\overset{+}{\leq} 2(k-t)K(\mu)\ln 2\\
&< \infty
\end{aligned}
$$

where $\overset{(a)}{\leq}$ follows from Lemma (4), and the last inequality $\overset{+}{\leq}$ follows from Lemma (5). □

It follows from the above theorem that

$$
\mathfrak{c}'_M\left(\phi(a_{1:k})\,\middle|\,\phi(a_{<t})\right)\xrightarrow{t\to\infty}\mu\left(\phi(a_{1:k})\,\middle|\,\phi(a_{<t})\right)
$$

As long as the pattern we are focusing on will last in the true environment μ, $\mathfrak{c}'_M$ can help us build our belief as we gain more experience.

Prediction with Random Sampling In practice, even for some specific pattern ϕ, we can't check all of the relevant information in past history. When we want to predict whether it will hold in the next states, we just do some random sampling to study it.

Assume that the real world $\bigwedge_{i=1}^{\infty} Q_{h_i}(a_i)$ is computable and deterministic. We want to predict whether $\phi(x) \equiv R(x) \to B(x)$ will hold in the next state. Then we do random sampling in the following strong sense. We select all of the positions to check whether ϕ holds with a sampling function $t : \mathbb{N} \to \mathbb{N}$ such that $\forall i(t_i < t_{i+1})$ and $w_{1:\infty}$ is martin-löf random ([5]), where $w_i = [\![\exists k(t_k = i)]\!]$. If all of the samples we select make ϕ true, then, should we believe that ϕ will be true in the next state—even if ϕ may be false in many states that we have not checked? Yes! Although we can't give some concrete bounds as in theorem (4), our belief will converge in the limits.

Theorem 5 If the sampling function t follows the above convention, then

$$
\mathfrak{c}'_M\left(\forall x\phi(x)\,\middle|\,\bigwedge_{i=1}^{n}\phi(a_{t_i})\right)\xrightarrow{n\to\infty} 1
$$

Proof We prove by two cases.

Case 1. If there are only finite positions that make ϕ false in total, namely,

$$|\{i : Q_{h_i}(a_i) \models \neg\phi(a_i)\}| < \infty$$

then all these cases can be coded with finite Kolmogorov complexity. According to the Weak instantaneous bounds in [4], it is easy to prove that

$$\mathfrak{c}_M\left(\neg\phi(a_{t_n}) \,\middle|\, \bigwedge_{i=1}^{n-1} \phi(a_{t_i})\right) \stackrel{+}{\leq} 2^{-K(t_n)} \xrightarrow{n\to\infty} 0$$

then it follows that

$$\mathfrak{c}'_M\left(\phi(a_{t_n}) \,\middle|\, \bigwedge_{i=1}^{n-1} \phi(a_{t_i})\right) \xrightarrow{n\to\infty} 1$$

Case 2. If ϕ is false infinitely often, $|\{i : Q_{h_i}(a_i) \models \neg\phi(a_i)\}| = \infty$, we can define some random test as follows.

Since the real world $\bigwedge\limits_{i=1}^{\infty} Q_{h_i}(a_i)$ is computable by assumption, we can generate all of the positions that ϕ is false, and define the following subsequence $\zeta_{1:\infty}$ of $w_{1:\infty}$.

$$\zeta_n = w_m \text{ where } m := \mu x\left(|\{i \leq x : Q_{h_i}(a_i) \models \neg\phi(a_i)\}| = n\right).$$

So we have the subsequence $\zeta_{1:\infty} = 0^\infty$ of $w_{1:\infty}$, which contradicts with the assumption of the Martin-löf randomness of $w_{1:\infty}$. □

5 Conclusion and Future Work

In this paper I have introduced Solomonoff prior into inductive logic, and analyzed the advantages of Solomonoff prior over Carnap's λ-continuum. Causality is neglected by Carnap and is characterized by Solomonoff with computable functions. As a byproduct, the proposition "all ravens are black" can be confirmed in a natural way. In the new "inductive logic", if we can do random sampling in a right way, we can prove the convergence theorem for some specific pattern, and make good predictions without record of the complete information of the history.

However, there are lots of further work that can improve the results of this paper.

- In this paper we do random sampling in a very strong sense. In the future work we should prove the convergence with some weak random sampling.
- The expressive power of the monadic first order logic is very limited, we will study how to define $\mathfrak{c}'_M$ when relation symbols are added into the language.

References

[1] R. Carnap, 1973, "Notes on probability and induction", *Synthese*, **25(3/4)**: 269–298.

[2] J. Hintikka, 1966, "A two-dimensional continuum of inductive methods", in J. Hintikka and P. Suppes (eds.), *Aspects of Inductive Logic*, pp. 113–132, Amsterdam: North-Holland.

[3] M. Hutter, 2005, *Universal Artificial Intelligence: Sequential Decisions Based on Algorithmic Probability*, Berlin: Springer.

[4] M. Hutter, 2007, "On universal prediction and Bayesian confirmation", *Theoretical Computer Science*, **384(1)**: 33–48.

[5] M. Li and P. M. B. Vitányi, 2008, *An Introduction to Kolmogorov Complexity and Its Applications, 3rd Edition*, Berlin: Springer.

[6] H. Putnam, 1963, "'Degree of confirmation' and inductive logic", in P. A. Schilpp (ed.), *Aspects of Inductive Logic*, pp. 761–783, La Salle: The Open Court Publishing Cooperation.

[7] H. Putnam, 1963, "Probability and confirmation", *The Voice of America, Forum Philosophy of Science, 10*, pp. 293–304, U.S. Information Agency.

[8] S. Rathmanner and M. Hutter, 2011, "A philosophical treatise of universal induction", *Entropy*, **13(6)**: 1076–1136.

[9] R. J. Solomonoff, 1978, "Complexity-based induction systems: Comparisons and convergence theorems", *IEEE Transactions on Information Theory*, **IT-24**: 422–432.

[10] S. L. Zabell, 1997, "Confirming universal generalizations", *Erkenntnis*, **45(2/3)**: 267–283.

为什么归纳逻辑需要所罗门诺夫先验？

李熙
北京大学 哲学系
xli@pku.edu.cn

摘　要

二十世纪五十年代，卡尔纳普发展了归纳逻辑，他把概率看作一种证据对假设对“确证度”；二十世纪六十年代，所罗门诺夫用通用归纳方法进行预测。为了增强归纳逻辑的归纳预测能力以及扩展所罗门诺夫通用归纳方法的表达力，本文整合二者。本文首先将所罗门诺夫先验概率的思想引入归纳逻辑中，在这个框架下，证明一阶逻辑版本的所罗门诺夫完全性定理，然后比较二者的优略。在卡尔纳普的归纳逻辑中，不管正面证据有多少，对像“所有乌鸦都是黑的”这种全称句的支持度最终都为零，而在用所罗门诺夫先验改造的归纳逻辑中，可以证明，在任何可计算的世界中，“所有乌鸦都是黑的”可以得到确证，只要在那些世界上真的所有乌鸦都是黑的。在所罗门诺夫的模型中，要证明完全性定理需要记录所有的过去信息，在修改后的归纳逻辑中，我们可以只关注某种具体的模式而忽略其它无关信息并证明类似的收敛定理。我们甚至可以不用记录所有的相关信息而采用随机抽样的方法建立合理的信念。