Inductive Inference of Σ^0_1- vs. Σ^0_2-Definitions for Computable Functions*

John Case and **Mandayam Suraj**
Computer and Info. Sciences Department
University of Delaware, Newark, DE
19716-2586 (USA)

*Slides are a click from URL =

http://www.cis.udel.edu/~case/colt.html

Standard Model of Inductive Inference Machines M

Depiction At Time t:

$$d_t \quad \xleftarrow{\text{Out}} \quad \boxed{M} \quad \xleftarrow{\text{In}} \quad (t, f(t))$$

- $\mathbf{M}$ is a scientist/machine for inductive inference.*
- $\mathbf{M}$, fed data sequence $(0, f(0)), (1, f(1)), (2, f(2)), \ldots$ from the graph of a (computable) function f, outputs a succession of **definitions**, $d_0, d_1, d_2, \ldots, d_t, \ldots$.
- One hopes that the sequence $d_0, d_1, d_2, \ldots, d_t, \ldots$ ends in some useful definition(s) for (close variants of) f.
- One can think of a data point, $(t, f(t))$, from (the graph of) f as representing, for some associated real-world phenomenon F, (experiment # t on F, outcome # $f(t)$ on F).

*In a computable universe even a collection of human scientists **is** a machine. In a universe with computable probability distributions for its behavior (QM?), the statistically **expected** behavior will still be computable.

Successful Endings For $d_0, d_1, d_2, \ldots, d_t, \ \ldots$?

We consider two of many possible:

- **Limiting Syntactic Convergence Of** $d_0, d_1, d_2, \ldots, d_t, \ \ldots$ means (by definition), for some t, $d_t = d_{t+1} = d_{t+2} = \cdots$. In this case we write $\mathbf{M}(f) \stackrel{\text{def}}{=} d_t$.

- A set of computable functions $\mathcal{S}$ is $\in \mathbf{Ex}$ $\stackrel{\text{def}}{\Leftrightarrow}$ **some** $\mathbf{M}$ on the graph of any $f \in \mathcal{S}$ outputs a limiting syntactically convergent sequence of standard Σ^0_1-definitions such that $\mathbf{M}(f)$ **defines** f.[*]

- A set of computable functions $\mathcal{S}$ is $\in \mathbf{LimEx}$ $\stackrel{\text{def}}{\Leftrightarrow}$ **some** $\mathbf{M}$ on the graph of any $f \in \mathcal{S}$ outputs a limiting syntactically convergent sequence of standard Σ^0_2-definitions such that $\mathbf{M}(f)$ **defines** f.[†]

[*] Σ^0_1-definitions are trivially intercompilable with ordinary computer programs. '$\mathbf{Ex}$' stands for **explanatory**.

[†] Total functions f having a Σ^0_2-definition are characterized as those expressible as $\lim_{t\to\infty} g(\cdot, t)$, for some computable function g.

Why Care Re Σ^0_2-Definitions For Computable Functions?

- $(\mathbf{LimEx} - \mathbf{Ex}) \neq \emptyset$.*
- But, what **use** (beyond for limiting-computability) are Σ^0_2-definitions of computable functions?
  - In physics it is easier to acquire the **shape** of a planetary orbit (ellipse) than it is to acquire a **program** for the position of the planet as a **function** of time.
  - Somewhat analogously, when one wants to prove global properties about the overall **shape** of computable function, e.g., **monotonicity**, in **some** cases, this **can** be done from a suitable Σ^0_2-definition of a computable function, but **not** from any Σ^0_1-definitions of that computable function.†

*[Case, Jain, Sharma, **IJFCS**, **3**, 1992].

†[Case, Jain, Sharma, **IJFCS**, **3**, 1992; Baliga, Case, Jain, Suraj, **JSL**, **59**, 1994].

Some Notation And Terminology

- If E is an informal sentence, we write

$$\ll \text{E} \gg$$

  to denote a **natural** translation of E into first order arithmetic.

- For each $i \in \{1, 2\}$, if q is a Σ_i^0-definition of some (partial) computable function ψ, then we let

$$\varphi_q^i = \text{(by definition) } \psi.$$

- '$=^\infty$' stands for equality of partial functions for **in**finitely many arguments.

- ψ is **monotone**

$$\overset{\text{def}}{\Leftrightarrow} (\forall x, y \mid x < y)[\psi(x)\downarrow < \psi(y)\downarrow].$$

Main Result And Philosophy Of Science

Simplified Main Theorem Suppose $\mathbf{T}$ is a **computably** axiomatizable first order theory which extends Peano Arithmetic ($\mathbf{PA}$) and which does **not** prove sentences (as below) that are **false** in the standard model. Then, there is a class of **monotone**, **computable** functions $\mathcal{C}$ such that

1. $\mathcal{C} \notin \mathbf{Ex}$,[*]
2. $(\forall f \in \mathcal{C})(\forall p \mid \varphi_p^1 =^\infty f)$
   $[\mathbf{T} \not\vdash \ll \varphi_p^1$ is monotone $\gg]$, and
3. There exists a machine $\mathbf{M}$ witnessing that $\mathcal{C} \in \mathbf{LimEx}$, and, for every $f \in \mathcal{C}$, for every e such that $\mathbf{M}(f) = e$,
   (a) $\mathbf{PA} \vdash \ll \varphi_e^2$ is monotone $\gg$,
   (b) $(\forall x, y)[\mathbf{T} \not\vdash \ll \varphi_e^2(x) = y \gg]$,[†] and
   (c) $\mathbf{PA} \vdash \ll \varphi_e^2$ is computable $\gg$.

[*] We also have a variant of the present theorem in which clause (1) is replaced by $\mathcal{C} \in \mathbf{Ex}$, and the other clauses remain the same.

[†] By clause (3b), one cannot predict, from the e's and $\mathbf{T}$, for the sake of Popper's Refutability Criterion for science, any data points at all in the graphs of the φ_e^2's!

Some Open Problems

1. Is the Theorem true for Σ^0_2- vs. Σ^0_3-Definitions, etc. whilst keeping the class of functions $\mathcal{C}$ **computable**?

2. Characterize or generalize the properties such as totality, monotonicity, etc. for which the Theorem is true.

3. Can we add to the Theorem something like:
   'For all machines $\mathbf{M}'$ witnessing that $\mathcal{C} \in \mathbf{LimEx}$, certain interesting subcases of clause (3) hold, but with $\mathbf{M}$ replaced by $\mathbf{M}'$' ?

4. Is the Theorem true if we replace **PA** by Heyting Arithmetic (**HA**) in clause (3a) only?*

*Of course in clause (3c) we can**not** replace **PA** by **HA**.