ON THE EQUIVALENCE OF INCOMPRESSIBILITY AND INCOMPLETENESS IN MACHINE LEARNING

Aidan Rocke
aidanrocke@gmail.com

February 20, 2021

Human reason has this peculiar fate that in one species of its knowledge it is burdened by questions which, as prescribed by the very nature of reason itself, it is not able to ignore, but which, as transcending all its powers, it is also not able to answer.-Immanuel Kant

1 Motivation

Motivated by the question of what problems are suitably addressed using methods in machine learning, I decided to formulate this general problem using Algorithmic Information Theory. My reason for this formulation is that the Kolmogorov Complexity $K(\cdot)$ appears to be a suitable measure of both incompressibility and epistemic uncertainty. I specifically derived two observer dependence theorems that appear to have subtle epistemological implications.

These theorems imply that randomness is not observer independent, that there is a fundamental relationship between incompressibility and inextricable epistemic uncertainty(i.e. incompleteness) in machine learning, and that in some settings the bounds on epistemic uncertainty may be more important than bounded rationality.

2 Bounds on epistemic uncertainty via expected Kolmogorov Complexity

Given the family of probabilistic models P_M, the Minimum Description Length of a dataset X of N samples from a discrete probability distribution P_X relative to the optimal model $\Omega \in P_M$ is given by the *First Observer Dependence Theorem*:

$$\mathbb{E}[K(X)] = H(X|\Omega) + H(\Omega) \tag{1}$$

where $H(\Omega)$ is the inextricable epistemic uncertainty of Ω concerning its own operation and $H(X|\Omega)$ is the inextricable epistemic uncertainty of Ω relative to P_X, i.e. a measure of the complexity of P_X relative to Ω.

In (1) I used the fact that Ω is a probabilistic program so it makes sense to compute the expected Kolmogorov Complexity, as well as the fact that the expected Kolmogorov Complexity of a random variable equals the Minimum Description Length of that variable [1]. I also implicitly assumed that ergodic assumptions are satisfied, which is the case in the regime of repeatable scientific experiments. For clarity, $\mathbb{E}[K(X)]$ is simultaneously a measure of *incompressibility*, *algorithmic randomness* and *incompleteness* where incompleteness is understood as a measure of inextricable epistemic uncertainty. The reason for this three-way correspondence is given in the next section.

Furthermore, from (1) we may deduce the *Second Observer Dependence Theorem* which gives upper and lower bounds on epistemic uncertainty:

$$H(\Omega) \leq \mathbb{E}[K(X)] \leq N \cdot H(P_X) \tag{2}$$

where the upper-bound results from an application of Shannon's source coding theorem. Now, the general question concerns the non-trivial epistemological implications of (2).

3 Analysis

If Ω is identified with what physicists call an observer i.e. a system that collects data X and tests hypotheses concerning the probabilistic structure of X in order to discover P_X then we find that relative to this observer, incompressibility, algorithmic randomness and incompleteness(as defined) are all equivalent to $\mathbb{E}[K(X)]$. We also find that the inextricable epistemic uncertainty of Ω relative to P_X is bounded by:

$$\mathbb{E}[K(X)] \geq H(\Omega) \tag{1}$$

This suggests that perceived randomness is not observer independent. More generally it suggests that the scientific method is not observer independent and it provides an information-theoretic formulation of Planck's claim that:

Science can't solve the ultimate mystery of nature. And that is because, in the last analysis, we ourselves are a part of the mystery that we are trying to solve.-Planck

Indeed, a closer inspection of the difference between $H(\Omega)$ and $H(X|\Omega)$ shows that there are two fundamentally different sources of perceived randomness i.e. epistemic uncertainty. Given an observer Ω, $H(X|\Omega)$ is a relative measure of system complexity. Financial markets and Brownian motion appear random because these systems have complex behaviours that remain unexplained by existing models. In the setting of complex systems, there is hope that a sufficiently powerful machine learning model may allow us to significantly reduce $H(X|\Omega)$. On the other hand, the epistemic uncertainty of Ω concerning itself is exactly $H(\Omega)$ i.e. the Minimum Description Length of Ω relative to a universal computer.

It follows that the structure underlying the fundamental axioms embedded in Ω is beyond the knowledge of Ω. In particular, since Ω implicitly contains a theory of computation and all such theories require arithmetic an optimal encoding of arithmetic is beyond the knowledge of Ω independently of the amount of data and computational resources available to Ω.

4 Discussion

My analysis implicitly assumes the Physical Church-Turing thesis, which many theoretical computer scientists and theoretical physicists find plausible. The implications of this thesis are consistent with the 'unreasonable effectiveness of mathematics' in the natural sciences [4].

While I have colleagues in the machine learning community that would like to use machine learning to investigate the Riemann Hypothesis, I believe that a closer inspection of (3) would provide a very useful analysis of the epistemic limits of such an enterprise. In fact, based on the insights I gained from the observer dependence theorems I recently managed to derive the prime number theorem using information-theoretic arguments [5].

I believe a closer inspection of the inequality (2) might also explain why the most fundamental theories in physics, Quantum Field Theory in particular, appear to describe algorithmically random phenomena.

References

[1] Aidan Rocke. An ergodic approach to Occam's razor. 2021.

[2] Olivier Rioul. This is IT: A Primer on Shannon's Entropy and Information. Séminaire Poincaré. 2018.

[3] Marcus Hutter. Algorithmic information theory. Scholarpedia. 2007.

[4] Eugene Wigner. The Unreasonable Effectiveness of Mathematics in the Natural Sciences. 1960.

[5] Aidan Rocke (https://mathoverflow.net/users/56328/aidan-rocke), information-theoretic derivation of the prime number theorem, URL (version: 2021-02-20): https://mathoverflow.net/q/384109

[6] Peter Grünwald and Paul Vitányi. Shannon Information and Kolmogorov Complexity. 2010.