Simplicity, probability and Ockham's razor: An impossibility result

Eric Johannesson
Stockholm University

New Ideas in Mathematical Philosophy, ENS
November 3, 2017

1 Introduction

(1) Among two *mutually exclusive* hypotheses that explain the evidence equally well, one should have a higher degree of belief in the simpler one.

(2) Among two *equally strong* hypotheses that explain the evidence equally well, one should have a higher degree of belief in the simpler one.

(3) Among two *comparable* hypotheses that explain the evidence equally well, one should have a higher degree of belief in the simpler one.

(4) For any two comparable hypotheses H and H' and evidence E such that $P(E|H) = P(E|H')$, we have $P(H|E) < P(H'|E)$ iff $K(H') < K(H)$.

With $E = H \vee H'$, this entails that

(5) For any two comparable hypotheses H and H', we have $P(H) < P(H')$ iff $K(H') < K(H)$.

2 Definitions and results

Definition 2.1 (Complexity)**.** If L is a first order language, let Ω_L be the class of all L-models. We say that an L-sentence *defines* a class of of L-models iff the sentence is true precisely in all members of that class, and that $\Delta_L \subseteq \mathcal{P}(\Omega_L)$ is the class of all classes of L-models defined by some L-sentence. Observe that Δ_L is a countably infinite Boolean algebra with top element Ω_L and bottom element $\emptyset$. We will take Δ_L to be our space of hypotheses. The complexity $K : \Delta_L \to \mathbb{N}$ of each hypothesis is then defined as the length of the shortest L-sentence defining that class.

Definition 2.2 (Respecting Ockham's razor)**.** Relative to a language L and a comparability relation $\Gamma \subseteq \Delta_L \times \Delta_L$, a probability function $P : \Delta_L \to [0, 1]$ *respects Ockham's razor* iff, for any $\langle A, B \rangle \in \Gamma$, $P(A) < P(B)$ iff $K(B) < K(A)$.

Lemma 2.1. *Let L be a first order language and let $\Sigma \subseteq \Delta_L$ be generated by a fragment of L with only a finite non-logical vocabulary. If Σ is infinite, then, for any $n \in \mathbb{N}$, there's $A \in \Sigma$ such that $n < K(A)$. In other words, if Σ is infinite, there's $A \in \Sigma$ of arbitrarily high complexity.*

Theorem 2.1. *Relative to a language with at least one binary predicate and a comparability relation that contain all pairs of mutually exclusive hypotheses, there's no probability function respecting Ockham's razor.*

Proof. Let Γ be such that $\langle A, B\rangle \in \Gamma$ if $A \cap B = \emptyset$. Let R be a binary predicate of L, and let $A \in \Delta_L$ be the class of models satisfying the sentence $\forall x Rxx$. Assume, towards contradiction, that P is a probability function respecting Ockham's razor relative to L and Γ. Clearly, $K(A) = 5$ and $K(\bar{A}) = 6$. Hence, $P(\bar{A}) < P(A)$. For each, $n \in \mathbb{N}$, let φ_n be the sentence $\forall x Rxx \wedge \exists x_1 ... \exists x_{n+1} \bigwedge_{1 \leq i < j \leq n+1} \neg R x_i x_j$, and let $A_n \in \Delta_L$ be the class of models satisfying it. Clearly, for any $i, j \in \mathbb{N}$ such that $i < j$, there's a model satisfying φ_i but not φ_j. Hence, $A_0, A_1, A_2, ...$ forms an infinite sequence of distinct members of A. By Lemma 2.1, it follows that there are subsets of A of arbitrarily high complexity. In particular, there's $B \subseteq A$ such that $100 < K(B)$. And since $K(B) \leq K(A - B) + K(A) + 3$, we have $92 < K(A - B)$. Now, let C be the class of models satisfying $\forall x \neg Rxx$. Since $K(C) \leq 6$ and $K(\bar{A} - C) \leq K(\bar{A}) + K(C) + 3$, we have $K(\bar{A} - C) \leq 15$. Hence $K(C) < K(B)$ and $K(\bar{A} - C) < K(A - B)$. Since $B \cap C = \emptyset$ and $(A - B) \cap (\bar{A} - C) = \emptyset$, we have $P(B) < P(C)$ and $P(A - B) < P(\bar{A} - C)$. But this contradicts the fact that $P(\bar{A}) < P(A)$. $\square$

Definition 2.3 (Closure)**.** A comparability relation $\Gamma \subseteq \Delta_L \times \Delta_L$ is *closed* iff $\langle A, B \rangle \in \Gamma$ for any $A, B \in \Delta_L$ such that

1. there are probability functions P, P' and P'' such that $P(A) < P(B)$, $P'(A) = P'(B)$ and $P''(A) > P''(B)$, and

2. either of the following obtains:

   (a) For every probability function P respecting Ockham's razor with respect to Γ, $P(A) < P(B)$.

   (b) For every probability function P respecting Ockham's razor with respect to Γ, $P(A) = P(B)$.

   (c) For every probability function P respecting Ockham's razor with respect to Γ, $P(A) > P(B)$.

Lemma 2.2. *Assume that L contains the language of arithmetic, and let $A \in \Delta_L$ be a class of models containing a model of Robinson arithmetic. Then there's $B \in \Delta_L$ of arbitrarily high complexity such that $B \subseteq A$.*

Theorem 2.2. *Relative to a language containing arithmetic and a closed comparability relation containing at least one pair of hypotheses of different complexity such that neither entails the other and both are consistent with but not entailed by Robinson arithmetic, there's no probability function respecting Ockham's razor.*

Proof. Assume that Γ is closed and let $\langle A, B\rangle \in \Gamma$ be a pair of hypotheses such that $K(A) \neq K(B)$, neither entails the other and both are consistent with, but not entailed by, Robinson arithmetic. By symmetry of Γ, we can assume without loss of generality that $K(A) < K(B)$. Let C be the class of models satisfying the axioms of Robinson arithmetic, and let $D = \bar{A} \cap C$. By assumption, $D \neq \emptyset$. Now, either $D \cap B \neq \emptyset$ or $D \cap \bar{B} \neq \emptyset$. Let E be either of those two which is non-empty. By Lemma 2.2, there's $E' \subseteq E$ of arbitrarily high complexity. Since $K(E') \leq K(A \cup E') + K(A) + 4$, that means there's $A' = A \cup E'$ of arbitrarily high complexity. In particular, there's $A' = A \cup E'$ such that $B \not\subseteq A'$, $A' \not\subseteq B$ and $K(B) < K(A')$. It follows that there are probability functions P, P' and P'' such that $P(A) < P(B)$, $P'(B) < P'(A)$ and $P''(A) = P''(B)$. Assume, towards contradiction, that P is a probability function respecting Ockham's razor with respect to Γ. Then $P(B) < P(A)$. Hence, for any P' respecting Ockham's razor with respect to Γ, we have $P'(B) < P'(A')$. Since Γ is closed, we get $\langle A', B\rangle \in \Gamma$. But since $K(B) < K(A')$, that means $P(A') < P(B)$, yielding a contradiction. $\square$

3 Conclusion

Our main result is that the following five assumptions are inconsistent with the laws of probability:

1. A scientific language is first order and contains arithmetic.

2. Any serious empirical hypothesis formulated in such a language is consistent with, but not entailed by, some sufficiently strong theory of first order arithmetic (e.g. Peano arithmetic).

3. The best measure of simplicity of an hypothesis is the length of the shortest formulation of it in some scientific language.

4. Ockham's razor applies to at least two serious empirical hypotheses, neither of which entails the other, and one of which is simpler than the other.

5. Ockham's razor doesn't determine the relation between the subjective probabilities of two hypotheses without taking the simplicity of those two hypotheses into account

References

Goodman, N. (1961). Safety, strength, simplicity. *Philosophy of Science*, 28(2):150–151.

Kolmogorov, A. (1968). Logical basis for information theory and probability theory. *IEEE Transactions on Information Theory*, 14(5):662–664.

Popper, K. (2002). *The Logic of Scientific Discovery*. Classics Series. Routledge.

Sober, E. (2015). *Ockham's Razors: A User's Manual*. Cambridge University Press.

Solomonoff, R. J. (1964a). A formal theory of inductive inference. part i. *Information and control*, 7(1):1–22.

Solomonoff, R. J. (1964b). A formal theory of inductive inference. part ii. *Information and control*, 7(2):224–254.

Vitányi, P. and Li, M. (2002). Simplicity, information, kolmogorov complexity and prediction. In Zellner, A., Keuzenkamp, H. A., and McAleer, M., editors, *Simplicity, Inference and Modelling*, pages 135–155. Cambridge University Press.