From Bayesian Epistemology to Inductive Logic

Jon Williamson

University of Kent CENTRE FOR REASONING

www.kent.ac.uk/reasoning

19 May 2011
Formal Epistemology Workshop
University of Southern California

Contents

1 Bayesian Epistemology as Semantics for Inductive Logic

Bayesian Epistemology: A Primer

? How strongly should one believe proposition θ?

The betting interpretation of degrees of belief motivates three norms:

Probability. Degrees of belief should satisfy the axioms of probability.

- $\mathcal{L}_n$ is a propositional language on elementary propositions $A_1, \ldots, A_n$.
- Ω_n is the set of *atomic states* of $\mathcal{L}_n$, i.e., propositions ω_n of the form $\pm A_1 \wedge \cdots \wedge \pm A_n$.

P1. $P(\omega_n) \geq 0$ for each $\omega_n \in \Omega_n$,

P2. $P(\tau) = 1$ for some tautology $\tau \in S\mathcal{L}$,

P3. $P(\theta) = \sum_{\omega_n \models \theta} P(\omega_n)$ for each $\theta \in S\mathcal{L}$.

✓ Dutch book argument: minimises worst-case expected loss / avoids sure loss.

Calibration. Degrees of belief should be set to physical probability where known.

EG The Principal Principle.

- $\mathbb{P}^*$ is the set of probability functions that satisfy known constraints on chances.
- $\langle\rangle$ is the convex hull operator.

C. $P \in \langle \mathbb{P}^* \rangle$.

✓ Minimises worst-case expected loss / long run loss.

Equivocation. One should not adopt extreme degrees of belief unless forced to by the Probability Norm or the Calibration Norm.

IE One's degrees of belief should equivocate sufficiently between the basic possibilities that one can express (i.e., between the atomic states of $\mathcal{L}_n$).

- *Equivocator* function $P_=(\omega_n) = 1/|\Omega_n|$ for each $\omega_n \in \Omega_n$.
- KL-divergence $d(P, Q) = \sum_{\omega_n \in \Omega_n} P(\omega_n) \log P(\omega_n)/Q(\omega_n)$.

E. P should be sufficiently close to $P_=$.

✓ Minimises worst-case expected loss, where the default loss function is logarithmic.

NB What counts as 'sufficiently close' can depend on pragmatic considerations.

Bayesians disagree as to whether to endorse Calibration and Equivocation.

Predicate Languages

- $\mathcal{L}$ is a first-order predicate language without equality,
  - with finitely many predicate symbols,
  - with a constant symbol $t_1, t_2, \ldots$, for each element of the domain.
- For $n \geq 1$, let $\mathcal{L}_n$ be the finite predicate language involving only constants $t_1, \ldots, t_n$.
- Let $A_1, \ldots, A_{r_n}$ be the atomic propositions expressible in $\mathcal{L}_n$.
- An *atomic n-state* ω_n is an atomic state $\pm A_1 \wedge \cdots \wedge \pm A_{r_n}$ of $\mathcal{L}_n$.

The **Calibration** norm remains the same.

Probability. Degrees of belief should satisfy the axioms of probability.

PP1. $P(\omega_n) \geq 0$ for each $\omega_n \in \Omega_n$ and each n,

PP2. $P(\tau) = 1$ for some tautology $\tau \in S\mathcal{L}$,

PP3. $P(\theta) = \sum_{\omega_n \models \theta} P(\omega_n)$ for each quantifier-free proposition θ, for any n large enough that $\mathcal{L}_n$ contains all the atomic propositions occurring in θ, and

PP4. $P(\exists x \theta(x)) = \sup_m P\left(\bigvee_{i=1}^m \theta(t_i)\right)$.

NB A probability function is determined by its values on the atomic n-states.

Equivocation. Degrees of belief should be equivocate sufficiently between the basic possibilities that one can express.

- Equivocator $P_=(\omega_n) = |\Omega_n|$ for all n and ω_n.
- *n-divergence* $d_n(P,Q) = \sum_{\omega_n \in \Omega_n} P(\omega_n) \log P(\omega_n)/Q(\omega_n)$.
- P is *closer* to R than Q if there is some N such that for all $n \geq N$, $d_n(P,R) < d_n(Q,R)$.

E. P should be sufficiently close to $P_=$.

Inductive Logic

Consider entailment relationships of the form:

$$\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \approx \psi^Y.$$

- $\varphi_1, \ldots, \varphi_k, \psi$ are propositions of some given logical language.
- $X_1, \ldots, X_k, Y$ denote inductive qualities that attach to these propositions.

EG Their certainty, plausibility, reliability, weight of evidence, or probability.

Normally some given semantics:

- will say something like 'the entailment relationship holds if all interpretations that satisfy the premisses $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k}$ also satisfy the conclusion ψ^Y',
- will say what an interpretation is,
- will say what it is to satisfy the premisses and satisfy the conclusion.

In a *probabilistic logic*, or *progic*:

- $X_1, \ldots, X_k, Y$ are sets of probabilities,
- interpretations are probability functions.

EG The *standard semantics* says that P satisfies θ^Z if and only if $P(\theta) \in Z$.

EG Other semantics: probabilistic argumentation, evidential probability, classical statistical inference, Bayesian statistical inference and Bayesian epistemology (Haenni et al., 2011, Part I).

Bayesian epistemology as semantics:

- Construe the language of $\varphi_1, \ldots, \varphi_k, \psi$ as the language of an agent.
- Construe the premisses $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k}$ as the agent's evidence of physical probabilities.
  - They say that $P^*(\varphi_1) \in X_1, \ldots, P^*(\varphi_k) \in X_k$.
- Interpretations are belief functions, i.e., probability functions.
- By the Calibration Norm a belief function P satisfies the premisses iff it lies in the convex hull of all probability functions that satisfy the relevant evidential constraints.

  IE Iff $P \in \langle \mathbb{P}^* \rangle$.
- If the premisses are consistent we can simply take $\mathbb{P}^* = \{P : P(\varphi_1) \in X_1, \ldots, P(\varphi_k) \in X_k\}$.
- Otherwise we cannot identify $\mathbb{P}^* = \emptyset$.
  - Inconsistent premisses tell us not that there is no chance function but rather that there is something wrong with the premisses.
  - Apply some consistency maintenance procedure.
- Interpret the conclusion ψ^Y as an assertion about rational degree of belief: $P(\varphi) \in Y$.
- The norms dictate that a belief function that satisfies the premisses also satisfies the conclusion iff $P(\varphi) \in Y$ for all those $P \in \langle \varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \rangle$ that are sufficiently equivocal.

- **?** What counts as ‘sufficiently’ equivocal?
  - There may be no contextual information in the logical setting.
- Let $\mathbb{E} = \langle \mathbb{P}^* \rangle$ denote the set of probability functions that satisfy constraints imposed by the agent’s evidence.
- Let $\downarrow\mathbb{E}$ denote the functions in $\mathbb{E}$ that are maximally equivocal (this set may be empty).
- Let $\Downarrow\mathbb{E}$ denote the set of functions in $\mathbb{E}$ that are sufficiently equivocal.

Arguably,

E1: $\Downarrow\mathbb{E} \neq \emptyset$. An agent is always entitled to hold some beliefs.

E2: $\Downarrow\mathbb{E} \subseteq \mathbb{E}$. Need to be calibrated with evidence.

E3: If $Q \in \Downarrow\mathbb{E}$ and $R \in \mathbb{E}$ is more equivocal than Q then $R \in \Downarrow\mathbb{E}$.

E4: If $\downarrow\mathbb{E} \neq \emptyset$ then $\Downarrow\mathbb{E} = \downarrow\mathbb{E}$. If it is possible to be maximally equivocal then one should be.

E5: $\Downarrow\Downarrow\mathbb{E} = \Downarrow\mathbb{E}$.

In the absence of contextual information the only option is to set:

$$\Downarrow\mathbb{E} = \begin{cases} \downarrow\mathbb{E} & : \quad \downarrow\mathbb{E} \neq \emptyset \\ \mathbb{E} & : \quad \text{otherwise} \end{cases} .$$

2 Critiques of Inductive Logic

Language Dependence

- **?** To what extent do inferences in this inductive logic depend on the underlying language?

If one can formulate an inference in more than one language then the two formulations will agree as to whether the premisses entail the conclusion.

Theorem 2.1 *Given predicate languages $\mathcal{L}^1$ and $\mathcal{L}^2$, suppose that $\varphi_1, \ldots, \varphi_n, \psi$ are propositions of both $\mathcal{L}^1$ and $\mathcal{L}^2$. Let $\mid\!\approx^1$ be the entailment relation with respect to $\mathcal{L}^1$ and $\mid\!\approx^2$ be the entailment relation with respect to $\mathcal{L}^2$. Then $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \mid\!\approx^1 \psi^Y$ if and only if $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \mid\!\approx^2 \psi^Y$.*

- ∴ There is normally no need to spell out the underlying language.

However, there is another sense in which inferences *do* depend on the underlying language:

EG Suppose
  - $\mathcal{L}^1$ has two unary predicates, *Green* and *Blue*.
  - $\mathcal{L}^2$ just has one unary predicate, *Grue*, which is synonymous with *Green or Blue*.

Then we have that

$$\mid\!\approx^1 Green(t_1) \vee Blue(t_1)^{3/4}$$

but

$$\mid\!\approx^2 Grue(t_1)^{1/2}.$$

- **?** Which is the correct inference?

The Bayesian would say that both are correct:

- If your language were $\mathcal{L}_1$ then you ought to believe that t_1 is Green or Blue to degree 3/4.
- If your language were $\mathcal{L}_2$ you ought to believe that t_1 is Grue to degree 1/2.
- ✓ Just as degrees of belief should depend on explicit evidence because that evidence tells us about the world, so too should degrees of belief depend on language because language tells us about the world.
  - Evidence tells us facts about the world.
    - Two evidence bases can be compared wrt strength and accuracy.
  - Language tells us about how the world can be carved up.
    - Two languages can be compared wrt how well they carve up the world.

But the Bayesian would also say that both inferences are lacking:

- There is an important piece of information that has not been taken into account:
  - That Grue is synonymous with Green or Blue.

∴ Need a language $\mathcal{L}^3$ in which one can express the proposition

$$\forall x, Grue(x) \leftrightarrow (Green(x) \vee Blue(x)).$$

Then one can formulate the inference

$$\forall x, Grue(x) \leftrightarrow (Green(x) \vee Blue(x)) \approx^3 Grue(t_1)^{3/4}.$$

But we also find that

$$\forall x, Grue(x) \leftrightarrow (Green(x) \vee Blue(x)) \approx^3 Green(t_1) \vee Blue(t_1)^{3/4},$$

so there is no inconsistency.

? Is this sort of dependence on language problematic?

Arguably not, given this triviality result:

- A a *synonymy map* between predicate languages $\mathcal{L}$ and $\mathcal{L}'$ is a consistent, countable set of propositions of the form $\theta_i \leftrightarrow \theta'_i$ where the θ_i are propositions of $\mathcal{L}$ and the θ'_i are propositions of $\mathcal{L}'$.

Theorem 2.2 *If an entailment relation $\approx$ of probabilistic logic with underlying predicate language $\mathcal{L}$ is invariant under all synonymy maps between $\mathcal{L}$ and $\mathcal{L}'$, for all $\mathcal{L}'$, then,*

1. $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \approx \psi^Y$ *if and only if* $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \approx \neg\psi^Y$,
2. $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \approx \psi^Y$ *implies that* $\{0, 1\} \subseteq Y$.

In sum,

- Theorem 2.1 shows that inferences are independent of the underlying language.
- However, they are not invariant under arbitrary synonymy maps.
- Theorem 2.2 shows that one cannot demand this stronger invariance condition without trivialising inductive logic.

The Principle of Indifference

? Does this logic fall to paradoxes of the Principle of Indifference.

> The principle of indifference asserts that if there is no *known* reason for predicating of our subject one rather than another of several alternatives, then relatively to such knowledge the assertions of each of these alternatives have an *equal* probability. These *equal* probabilities must be assigned to each of several arguments, if there is an absence of positive ground for assigning *unequal* ones. (Keynes, 1921, p. 45)

Keynes restricted the Principle to *indivisible* alternatives:

> In short, the principle of indifference is not applicable to a pair of alternatives, if we know that either of them is capable of being further split up into a pair of possible but incompatible alternatives of the same form as the original pair. (Keynes, 1921, p. 66)

- On a propositional language, the set Ω_n of atomic states represents the partition of indivisible alternatives.
- On a predicate language there is no partition of indivisible alternatives:
  - For every set Ω_n of atomic n-states, there is another set Ω_i for $i > n$ that splits up the original alternatives.

But one can formulate the Principle of Indifference in a way that applies equally to the infinite and the finite case:

POI. If atomic n-states ω_n^* and $\omega_n^\dagger$ are treated symmetrically by the premisses then,

$$\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \approx {\omega_n^*}^Y \text{ iff } \varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \approx {\omega_n^\dagger}^Y.$$

Here we can say that ω_n^* and $\omega_n^\dagger$ are *treated symmetrically by the premisses* just in case:

- for any probability function P satisfying the premisses, there is another function satisfying the premisses which swaps the probabilities of ω_n^* and $\omega_n^\dagger$ but which otherwise agrees with P as far as possible.

Interestingly:

Theorem 2.3 *The Bayesian entailment relation $\approx$ satisfies POI.*

? Is satisfying the Principle of Indifference a problem?

× There is no inconsistency: probabilities are attached to a language $\mathcal{L}$ in a consistent way.

✓ But paradoxes might still arise when one changes the conceptualisation of a particular problem to an equivalent but different conceptualisation in which the partition of indivisible alternatives is different.

- IE If one changes the language and at the same time asserting an equivalence between certain propositions of the new language and of the old.
  - IE By introducing a synonymy map.
  - EG The *Grue* example above.
  - × But Bayesians can argue that dbs should display this behaviour.
  - × Any demand that inductive logic should be immune to this sort of behaviour is untenable since there is no non-trivial inductive logic that satisfies such a demand (Theorem 2.2).

In sum, POI holds, but arguably not with problematic consequences.

Learning from Experience

Any inductive logic that satisfies POI has the property that

$$|\!\!\!\approx \psi^Y \text{ iff } P_=(\psi) \in Y$$

IE If there are no premisses, the logic is determined by the equivocator function.

- This property has been criticised:
  - It apparently fails to capture the phenomenon of learning from experience.
  - ✓ The equivocator yields probability $\frac{1}{2}$ that the 101st observed raven will be found to be black,

    $$|\!\!\!\approx Br_{101}{}^{.5}.$$

  - × But it also yields probability $\frac{1}{2}$ that the 101st observed raven will be found to be black, conditional on the first 100 ravens having been found to be black,

    $$|\!\!\!\approx Br_{101}|Br_1 \wedge \cdots \wedge Br_{100}{}^{.5}.$$

In fact, this apparent inability to capture learning from experience is based on a simple misinterpretation.

- $Br_1 \wedge \cdots \wedge Br_{100}$ appearing in the conclusion $Br_{101}|Br_1 \wedge \cdots \wedge Br_{100}{}^{.5}$ should not be interpreted as evidence or experience.
- It is the premisses that are intended to reflect evidence.

∴ The situation has been misrepresented.

- According to the Bayesian semantics presented here, if 100 ravens are observed and found to be black, this needs to be translated into constraints on physical probability, of the form $P^*(Br_i) \in X$.
- It is statistical theory that tells us how to do this:
  - Granting that the physical probs are iid, statistical theory yields claims of the form:

    $$P^*(P^*(Br_i) \geq 1-\delta) = 1-\epsilon, \text{ for } i > 100.$$

  - Given some threshold $1-\epsilon_0$ of acceptance, an agent can then choose δ_0 such that $P^*(P^*(Br_i) \geq 1-\delta_0) = 1-\epsilon_0$, i.e., such that the claim that $P^*(Br_i) \geq 1-\delta_0$ reaches the threshold of acceptance.
  - Then the question under consideration is better represented as

    $$Br_1{}^1, \ldots, Br_{100}{}^1, Br_{101}{}^{[1-\delta_0,1]}, Br_{102}{}^{[1-\delta_0,1]}, \ldots |\!\!\!\approx Br_{101}{}^?.$$

Our inductive logic will choose the maximally equivocal point in the interval $[1-\delta_0, 1]$:

$$Br_1{}^1, \ldots, Br_{100}{}^1, Br_{101}{}^{[1-\delta_0,1]}, Br_{102}{}^{[1-\delta_0,1]}, \ldots |\!\!\!\approx Br_{101}{}^{1-\delta_0},$$

which is clearly a much more reasonable value than $\frac{1}{2}$.

NB A conclusion of the form $Br_{101}|Br_1 \wedge \cdots \wedge Br_{100}{}^Y$ is simply a formal abbreviation of the claim that $P(Br_1 \wedge \cdots \wedge Br_{100} \wedge Br_{101})/P(Br_1 \wedge \cdots \wedge Br_{100}) \in Y$ for every $P \in \Downarrow\mathbb{E}$.

NB In general, under an objective Bayesian interpretation, conditional probabilities are not always interpretable as conditional beliefs (Williamson, 2011).

Universal Hypotheses

POI also implies that many universal hypotheses are given zero probability,

$$\varphi_1^{X_1}, \dots, \varphi_k^{X_k} \approx \forall x\theta(x)^0.$$

× This can be counterintuitive.

EG Finding the first 100 observed ravens to be black offers no support to the conclusion that all ravens are black:

$$Br_1 \wedge \cdots \wedge Br_{100}, Br_{101}{}^{[1-\delta_0,1]}, Br_{102}{}^{[1-\delta_0,1]}, \dots \approx \forall x Bx^0.$$

- In general, *no generalisations in, no generalisations out*:
  - If premisses are to raise the probability of a universally quantified proposition away from zero, then those premisses must themselves involve quantifiers:

Theorem 2.4 *Suppose that, for $\theta(x)$ quantifier-free, $\approx \forall x\theta(x)^0$ but $\varphi_1^{X_1}, \dots, \varphi_k^{X_k} \approx \forall x\theta(x)^Y$ where $\inf Y > 0$. Then $\varphi_1, \dots, \varphi_k$ are not all quantifier-free.*

? Isn't this a problem for applications to the sciences, which appear to routinely invoke universal generalisations?

Perhaps not:

- Arguably, universal generalisations are not so central to science.
  - In Carnap's time, largely under the influence of the logical empiricists, scientific theories were understood as collections of universal generalisations.
    - Augmented by statements specifying boundary conditions, bridge laws etc.
  - ∴ Inductive logics that gave universal generalisations probability zero were taken to be refuted by scientific practice.
  - But in the 1980s and 1990s this view of laws was found to be untenable.
    - The ubiquity of ceteris paribus laws and pragmatic laws became recognised.
  - More recently, the Hempelian DN account of explanation, which saw scientific explanations as deductions from universal generalisations, has been replaced by a mechanistic view of explanation.
  - Science is increasingly understood as a body of mechanisms, not generalisations.
  - ∴ The relevance of Theorem 2.4 to science is less obvious now than it would have appeared a few decades ago.

Also, there are a variety of attitudes one can take towards universal generalisations.

- Bayesian epistemology distinguishes what is *believed* and what is *granted*.
- Given what is already granted, BE provides rational norms for degrees of belief:
  - In terms of inductive logic, it tells us how strongly one should believe a conclusion proposition having granted some premisses.
- Moreover, different norms cover granting and believing.
  - Grounds for *granting* include coherence, simplicity, strength, accuracy, technical convenience, unifying power etc.
  - But propositions should only be *believed* to the extent warranted by their Bayesian probability relative to what is granted.
- So a generalisation may have probability zero yet be ripe for granting.
- ∴ The Bayesian can argue that:
  - One should remain sceptical about universal hypotheses that have probability 0.
  - Yet one can go on to grant those same hypotheses for other reasons.

In sum,

- Universal generalisations appear to play less of a role in science than previously thought.
- The key question may be whether they should be *granted*, rather than believed.

Conclusion

The Bayesian semantics:

- $\varphi_1^{X_1}, \dots, \varphi_k^{X_k} \approx \psi^Y$ iff an agent with evidence represented by the premisses ought to believe ψ to some degree in Y.
- This semantics is independent of language.
  - But not invariant under synonymy maps, a requirement that leads to triviality.
- The Principle of Indifference is satisfied.
  - But no inconsistency or paradox arises.
- The logic captures the phenomenon of learning from experience.
- It advocates a sceptical attitude to universal generalisations.
  - But this need not set it at odds with science.

References

Haenni, R., Romeijn, J.-W., Wheeler, G., and Williamson, J. (2011). *Probabilistic logics and probabilistic networks*. Synthese Library. Springer.

Keynes, J. M. (1921). *A treatise on probability*. Macmillan (1948), London.

Williamson, J. (2011). An objective Bayesian account of confirmation. In Dieks, D., Gonzalez, W. J., Hartmann, S., Uebel, T., and Weber, M., editors, *Explanation, Prediction, and Confirmation. New Trends and Old Ones Reconsidered*, pages 53–81. Springer, Dordrecht.