Do probability and causality exist?

Jon Williamson
Department of Philosophy & Centre for Reasoning, University of Kent

10 February 2009
Oxford series on The Nature and Scope of Information

1 Existence and Objectivity

Bruno de Finetti:

> My thesis, paradoxically, and a little provocatively, but nonetheless genuinely, is simply this:
>
> PROBABILITY DOES NOT EXIST.
>
> The abandonment of superstitious beliefs about the existence of Phlogiston, the Cosmic Ether, Absolute Space and Time, . . . , or Fairies and Witches, was an essential step along the road to scientific thinking. Probability, too, if regarded as something endowed with some kind of objective existence, is no less a misleading misconception, an illusory attempt to exteriorize or materialize our true probabilistic beliefs.
>
> In investigating the reasonableness of our own modes of thought and behaviour under uncertainty, all we require, and all that we are reasonably entitled to, is consistency among these beliefs, and their reasonable relation to any kind of objective data ('relevant' in as much as subjectively deemed to be so). This is Probability Theory. (de Finetti, 1970, p.x)

- ▸ Probabilities are rational degrees of belief.
- • Subjective in two senses: (i) non-physical ('does not exist'); (ii) a matter of arbitrary choice.

Two distinctions:

Physical/Epistemic: Do probabilities exist independently of (actual or ideal) epistemic states?

Objective/Subjective: If two agents disagree as to the probabilities, can both be right?

De Finetti's position was epistemic and subjective.

- ✓ An epistemic view nicely handles a wide variety of probabilistic claims.
- × De Finetti's position seems rather too subjective:
  - If your evidence says nothing about a proposition are you rational to fully believe it?
- ▸ Can one have an epistemic view that isn't so radically subjective?

Contents

1 Existence and Objectivity **2**

2 Objective Bayesianism **5**

2.1 Probability 7
2.2 Calibration 10
2.3 Equivocation 11
2.4 Updating 17
2.5 Language Relativity 19
2.6 Norm Relativity 20

3 Epistemic Causality **23**

4 Ultimate Belief Analyses **28**

2 Objective Bayesianism

Question. To what extent should an agent with evidence $\mathcal{E}$ believe a proposition a expressible in her language $\mathcal{L}$?

Answer. The strengths of her beliefs should satisfy certain norms:

Probability: Strength of belief should be representable by probability $P : S\mathcal{L} \longrightarrow [0,1]$

- $P(a_1) + P(a_2) + \cdots = 1$ if the a_i form a partition.
- Define $P(a|b) \stackrel{\text{df}}{=} P(a \wedge b)/P(b)$.

Calibration: Degrees of belief should be set to physical probabilities, where available

- $\mathcal{E}$ = {it rains the day after 70% of days like this}; $P(\text{it will rain tomorrow}) = 0.7$
- $\mathcal{E}$ = {cancer recurs in at least 60% of patients like this}; $P(\text{this patient's cancer will recur}) \in [0.6, 1]$

Equivocation: Degrees of belief should be otherwise be maximally equivocal

- $\mathcal{E} = \emptyset$; $P(a) = 1/2$ (for elementary a),
- $\mathcal{E}$ = {cancer recurs in at least 60% of patients like this}; $P(\text{this patient's cancer will recur}) = 0.6$.

Evidence. = everything the agent takes for granted in her current operating context.

- e.g., observations, theory, background knowledge and assumptions.
- medical consultant: education, medical training, uncontroversial studies in the medical literature, observations of the patient's symptoms.
- philosophical sceptic: can't take much for granted at all.
- OBE:
  - given $\mathcal{E}$, $\mathcal{L}$, how strongly should the agent believe α?
  - *not*: should the agent take the components of $\mathcal{E}$ for granted in the first place?

Language.

- Picks out the propositions that are the objects of the agent's beliefs.
- May vary according to context.
- May change.
- May constitute evidence.

We shall assume that $\mathcal{L}$ is closed under propositional connectives $\neg, \wedge, \vee, \rightarrow, \leftrightarrow$.

2.1 Probability

Strength of beliefs should be probabilities $P_{\mathcal{E}} : S\mathcal{L} \longrightarrow [0,1]$ (or $P^{\mathcal{L}}_{\mathcal{E}}$):

- $P_{\mathcal{E}}(a_1) + P_{\mathcal{E}}(a_2) + \cdots = 1$ where the a_i form a partition,

Why? Two kinds of argument:

Derivation: Cox's theorem. If

- logically equivalent propositions have the same degree of belief,
- $P(a|b) = 1$ if b logically implies a,
- $P(a \wedge b)$ is a continuous, strictly increasing function of $P(a|b)$ and $P(b)$,
- $P(\neg a)$ is a decreasing function of $P(a)$,
- the values of P are sufficiently dense in the unit interval,
- ▸ then P is representable by a probability function (Paris, 1994, p. 24).

Interpretation: Dutch book argument.

- Interpret $P(a)$ as a *betting quotient*
  - pay $P(a)S$ with return S if a is true (unknown $S \in \mathbb{R}$).
- ▸ to avoid the possibility of certain loss, P must be a probability function.

Derivation versus Interpretation

Derivation: Derive result from assumptions: $\models \alpha \to r$.

- Clearly α is at least as strong as r.
- ▸ α can't be less controversial than r.
- ▸ A derivation can't be used to convince an opponent of r.
- Opponent: $\models \neg r \to \neg \alpha$.
  - ▸ Since I reject r I should not endorse α.
  - – If α seems plausible then that is just artifice:
    - ∗ that α is dubious can be shown by its consequences (e.g., r).

Interpretation: explicate meaning, then show that R follows from plausible α'.

- The interpretation can do the bulk of the work.
- α' can be quite uncontroversial (or even tautologous).
- ▸ An opponent of r would have to
  - – admit to talking about something else, or,
  - – deny plausible α'.

▸ Dutch book better motivation of the Probability norm than Cox's theorem.

Objectivity

- The Dutch book argument leads to countable additivity:
  - $P_{\mathcal{E}}(a_1) + P_{\mathcal{E}}(a_2) + \cdots = 1$ where countably infinite a_i form a partition.
- But countable additivity precludes uniqueness of $P_{\mathcal{E}}$:
  - $\mathcal{E} = \varnothing$
  - $P_{\mathcal{E}}(a_1) + P_{\mathcal{E}}(a_2) + \cdots = 1$.
  - ▸ $P_{\mathcal{E}}(a_m) > P_{\mathcal{E}}(a_n)$ for some m and n.
  - the skewing is arbitrary, a matter of subjective choice.
  - ▸ $P_{\mathcal{E}}$ is not uniquely determined

▸ No version of Bayesian epistemology can be fully objective.

▸ We can have strong norms, but no *logical interpretation* (Keynes, Carnap).

2.2 Calibration

Degrees of belief should be compatible with evidence.
Two types of constraint:

Quantitative: If $\mathcal{E}$ implies that physical probability $P^* \in \mathbb{P}^* \neq \emptyset$, then $P_{\mathcal{E}} \in \langle \mathbb{P}^* \rangle$.

- If $\mathbb{P}^* = \emptyset$ then some kind of consistency maintenance procedure is required.

Qualitative: Evidence of causal, semantic, logical, mereological, mathematical structure imposes constraints.

- e.g., if an agent finds out about a new variable that is not a cause of her others, she should not change her beliefs about the others: $P_{\mathcal{E}}^{\mathcal{L}'} \restriction \mathcal{L} = P_{\mathcal{E}}^{\mathcal{L}}$.

▸ $P_{\mathcal{E}} \in \mathbb{E} = \langle \mathbb{P}^* \rangle \cap \mathbb{S}$

Why?

Quantitative: Degrees of belief are interpreted as betting quotients—for our bets to pay off in the long term they need to latch onto physical probabilities as far as possible.

Qualitative: Degrees of belief should respect relevance relations between variables.

- Otherwise the agent will make poor predictions in the long run.

2.3 Equivocation

Degrees of belief should otherwise be maximally equivocal

Maximum Entropy Principle (maxent): $P_{\mathcal{E}} \in \{P \in \mathbb{E} : P \text{ maximises } H(P)\}$

$$H(P) = -\sum_{\omega} P(\omega) \log P(\omega)$$

Why? Argument by Derivation

Paris and Vencovská (1990): an *inference process* is a function which selects a probability function that satisfies $\mathcal{E}$. Desiderata:

Equivalence: given two logically equivalent knowledge bases the inference process should select the same probability function.

Continuity: Small changes in $\mathcal{E}$ should lead to small changes in $P_{\mathcal{E}}$.

Irrelevant Information: Adding information about unrelated variables should not change degrees of belief.

Open-Mindedness: Only give probability 0 or 1 if forced to by $\mathcal{E}$.

Renaming: $P_{\mathcal{E}}$ should be independent of the ordering of the ω.

Obstinacy: If current degrees of belief are compatible with new evidence then they should not change.

Relativisation: Degrees of belief conditional on b should not depend on evidence that is conditional on $\neg b$.

Independence: Given probabilities of b and a and c conditional on b, $P_{\mathcal{E}}$ should yield a and c independent conditional on b.

- The only inference process that satisfies these desiderata is maxent.

But:

- × This is an argument by derivation—won't convince an opponent of maxent.
- × It assumes that we require an inference process, i.e., unique $P_{\mathcal{E}}$.
  - we know this is unattainable in general.

Why? Argument by Interpretation

Degrees of belief should otherwise be maximally equivocal

- i.e., as close as possible to an *equivocator* function $P_=$.

$$P_=(\omega) = P_=(a_1 \cdots a_n) = \frac{1}{2^n}$$

$$d(P, Q) = \sum_{\omega} P(\omega) \log \frac{P(\omega)}{Q(\omega)}$$

On a finite domain, distance from the equivocator

$$d(P, P_=) = \sum_{\omega} P(\omega) \log 2^n P(\omega)$$

is minimised when entropy

$$H(P) = -\sum_{\omega} P(\omega) \log P(\omega)$$

is maximised.

- Why $P_=$?
  - finitely many variables—$P_=$ is intuitively equivocal,
  - countably many variables—can be extended,
  - uncountable domain—depends on evidence (& may not be unique).
- Why d?
  - natural choice in information theory and information geometry.
  - Information Theory:
    - $d(P, Q)$ is interpreted as the amount of information in P that is not in Q.
  - Information Geometry:
    - minimising distance is a kind of projection operation.
  - (Infinitely many variables: distance minimiser may not be unique.)
- Why equivocate at all? **Caution**:
  - Extreme degrees of belief trigger risky actions.
  - One should only take on such risk to the extent forced by evidence.
  - ▸ if evidence allows lee-way choose less extreme degrees of belief.

The story so far

Degrees of belief are used as a basis for action, so our justifications are pragmatic:

Probability: Avoid certain loss.

Calibration: Maximise long-term gain.

Equivocation: Minimise unwarranted risk

2.4 Updating

If $\mathcal{E}$ changes to $\mathcal{E}'$, then $P_{\mathcal{E}}$ changes to $P_{\mathcal{E}'}$.

- Foundational: update by determining degrees of belief afresh
- Degrees of belief track the agent's epistemic background
- Since $\mathcal{E}$ strongly constrains degrees of belief, changes to $\mathcal{E}$ strongly constrain changes to her degrees of belief.

Subjective Bayesian Updating

Problem: Probability + Calibration is very weak

- $\mathcal{E}$: cancer recurs in at least 60% of patients like this,
- one could initially adopt $P(a) = 0.6$ and then change to $P'(a) = 1$ for no reason

▸ Subjective Bayesians impose a further constraint on updating:

Conditionalisation: On learning e, set $P'(a) = P(a|e)$.

- Conservative rather than Foundational: new degrees of belief are tied to old.

✓ Strong constraint: if e is in the language and $P(e) > 0$ then P' is uniquely determined.

✓ Intuitively compelling.

× Requires all evidence to be expressible in prior language $\mathcal{L}$.

× Requires new evidence to be consistent with old.

× Requires new evidence to only impose the constraint $P(e) = 1$.

× Requires prior degrees of belief conditional on any eventuality.

× Requires being eternally true to prior beliefs—no re-evaluation.

Where maxent and conditionalisation disagree, maxent should be preferred (Williamson, 2009).

2.5 Language Relativity

Objection: $P_{\mathcal{E}}$ depends on the agent's language $\mathcal{L}$:

- if $\mathcal{L} = \{$ sun, rain, snow$\}$ then $P_{\emptyset}(rain) = 1/3$.
- if $\mathcal{L} = \{$ sun, rain, snow, hail$\}$ then $P_{\emptyset}(rain) = 1/4$.

Is this a problem for objective Bayesianism? **No**:

- An agent's degrees of belief should depend on her information.
- An agent's language contains implicit information about the world.
  - e.g., $\mathcal{L} = \{$ sun, rain, powder-snow, wet-snow, sleet, dry-snow, icy-snow,$\ldots\}$.

▸ a change in language is a change in information.

▸ one should expect P to depend on language as well as explicit evidence $\mathcal{E}$.

2.6 Norm Relativity

What is the status of the norms of objective Bayesianism?

- are they voluntary or is there some compulsion?

Subjectivist: for Conditionalisation and against Equivocation

- the above arguments go part-way towards meeting these concerns.

Radical Subjectivist: rational norms express individual inclinations or value judgements and cannot be imposed on others.

- norms of OB justified by avoiding sure loss, inferential success, avoiding unwarranted risk.
- the radical subjectivist may accept the justifications but disavow these goals
  - ▸ objective and subjective Bayesianism are just two stances, neither right nor wrong.

A possible response:

- degrees of belief are used as a basis for decision, inference and action
- the ends or values connected with objective Bayesianism come part-and-parcel with these uses.
  - Good decision, inference and action requires that the norms be followed.

× Radical Subjectivist: when I say ‘degrees of belief’ I do not buy into these values; I do not care whether or not I make good decisions, inferences and actions on the basis of my ‘degrees of belief’; I talk about them in isolation from their consequences.

  × Fine, but then the radical subjectivist is talking about something else
    - what we might call ‘degrees of idle-belief’,
    - not the degrees of belief that do unavoidably ground his actions.
  - The norms are compulsory for degrees of belief
    - not for degrees of idle-belief.

- Analogy: boat-building is subject to a *floating* norm.
  - Disavowing this norm is to disavow boat-building.

The story so far

- OBE can be characterised by 3 norms:
  - Probability, Calibration and Equivocation.
- These 3 norms are justifiable:
  - avoid sure loss, maximise long-term gain, avoid unwarranted risk.
  - No need for arguments by derivation.
- One can defend OBE against 3 kinds of relativism:
  - subjectivism, language relativity, radical subjectivism.

3 Epistemic Causality

De Finetti again:

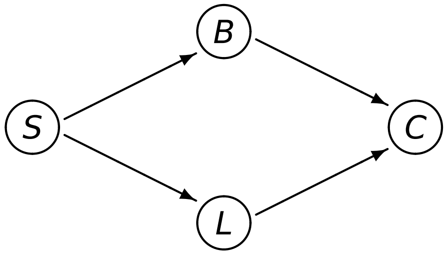
> Rigid laws are formulated and accepted by our minds for the same reasons that lead us to formulate and accept any judgment of probability whatever; the only difference consists in the very high probability that we attribute, in the case of rigid laws, to their exact agreement with experimental facts. The probability is so high that we can call it "practical absolute certainty," or, simply "certainty," understanding all the while the qualification that is essential from the philosophical and logical point of view.
>
> The notion of "cause" thus depends on the notion of probability, and it follows also from the same subjective source as do all judgments of probability: this explanation seems to constitute the true logical translation of the conception of "cause" advanced by David Hume, which I consider the highest peak that has been reached by philosophy. (de Finetti, 1937, pp. 114–115)

The epistemic theory of causality

- characterises causality in terms of the beliefs of an agent.
  - causality is *epistemic* rather than *physical*.
- But, as with OBE, causal beliefs are heavily constrained by evidence: objectivity.

Causal Beliefs

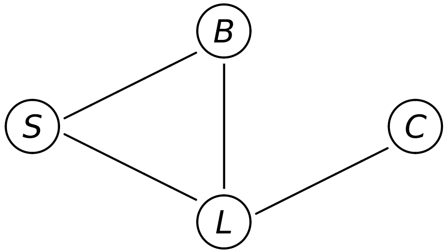
An agent's causal beliefs are like a map—they help her reason about the world.
They help her with inference (prediction, diagnosis, control) and explanation.



Causal beliefs are useful because

- typically cause and effect are probabilistically dependent,
- typically intervening to change a cause changes its effects,
- ▸ typically **strategic dependencies** accompany causal connections: cause and effect are probabilistically dependent when intervening to fix a cause and controlling for the effect's other direct causes,
- typically causal chains reflect underlying mechanisms.

Analogy. Link two towns if one can travel between them in two hours.



- Nodes are physical, links are just representation,
- links are objective,
- physical mechanisms explain particular instances,
- an agent’s beliefs depend on her evidence.

Epistemic Causality in a Nutshell

Question. What causal beliefs should an agent with evidence $\mathcal{E}$ hold?

Answer. Her causal beliefs should satisfy certain norms:

Acyclicity: Causal beliefs should be representable by a directed acyclic graph $C_{\mathcal{E}}$.

Calibration: Causal beliefs should be compatible with evidence $\mathcal{E}$:

- strategic dependencies should be explained where possible;
- causal beliefs should be explained by mechanisms where possible, e.g.:
  - if A occurs after B then $A \not\longrightarrow B$ in $C_{\mathcal{E}}$,
  - if any physical link between A and B goes via C then $A - C - B$.

Equivocation: Causal beliefs should otherwise be maximally equivocal:

- No causal beliefs that are not warranted by $\mathcal{E}$.

- ▸ Causal beliefs should account for $\mathcal{E}$ but no more.

Constructing a Causal Belief Graph

Construct a **minimal** graph $C_{\mathcal{E}}$ that
- accounts for all strategic dependencies,
- satisfies other constraints imposed by mechanistic evidence.

Special Case: no non-causal dependencies. If every strategic dependency is represented by an arrow in $C_{\mathcal{E}}$ then:

Causal Markov Condition: each variable is probabilistically independent of its non-effects, conditional on its direct causes.

- The set of minimal causal belief graphs that account for strategic dependencies and any $\not\longrightarrow$ constraints imposed by $\mathcal{E}$ is the set of minimal graphs satisfying the Causal Markov Condition and the $\not\longrightarrow$ constraints.
- We can use standard algorithms to build a causal belief graph.

Objectivity: on average all but the directions of two arrows will be fully determined (Gillispie and Perlman, 2002).

4 Ultimate Belief Analyses

OB and EC give accounts of probability and causality relativised to evidence.

- Intuitively probability and causality are more objective still:
  - Chance is relative only to time,
  - Causality is not relative at all.

Ultimate belief analysis of probability. $P^*_t(a) = x$ iff an agent with complete evidence $\mathcal{E}^*_t$ at t ought to believe a to degree x. I.e., $P^*_t = P_{\mathcal{E}^*_t}$.

- Complete evidence: $\mathcal{E}^*_t$ includes all information about the pattern of instantiation of fundamental properties and relations by particular things, up to time t.

Ultimate belief analysis of causality. $A \longrightarrow^* B$ iff an agent with ultimate evidence $\mathcal{E}^*$ ought to hold causal belief $A \longrightarrow B$. I.e., $C^* = C_{\mathcal{E}^*}$.

- Ultimate evidence: $\mathcal{E}^*$ includes all information about the pattern of instantiation of fundamental properties and relations by particular things at all times.

N.B. probability and causality are *not* fundamental relations here.

Circularity

Aren't causal beliefs beliefs about causality?

- if so, then it is circular to analyse causality in terms of causal beliefs (Price, 1998).

× here causal beliefs are a type of belief, not beliefs about causality

  - similarly probabilistic beliefs are degrees of belief, not beliefs about probability.

$$P_{\mathcal{E}}(A|B) = x \in [0,1]$$

$$C_{\mathcal{E}}(A|B) = x \in \{-, \emptyset, +\}$$

Conclusion

Neither probability nor causality exist!

- They are epistemic notions, relative to evidence or information.
- But they are not subjective—they are highly constrained.
- Evidence can be eliminated to yield unrelativised versions.

Further Reading

http://www.kent.ac.uk/secl/philosophy/jw.htm

References

de Finetti, B. (1937). Foresight. its logical laws, its subjective sources. In Kyburg, H. E. and Smokler, H. E., editors, *Studies in subjective probability*, pages 53–118. Robert E. Krieger Publishing Company, Huntington, New York, second (1980) edition.

de Finetti, B. (1970). *Theory of probability*. Wiley (1974).

Gillispie, S. B. and Perlman, M. D. (2002). The size distribution for Markov equivalence classes of acyclic digraph models. *Artificial Intelligence*, 141:137–155.

Paris, J. B. (1994). *The uncertain reasoner's companion*. Cambridge University Press, Cambridge.

Paris, J. B. and Vencovská, A. (1990). A note on the inevitability of maximum entropy. *International Journal of Approximate Reasoning*, 4:181–223.

Price, H. (1998). Two paths to pragmatism II. In Casati, R. and Tappolet, C., editors, *European Review of Philosophy*, volume 3, pages 109–147.

Williamson, J. (2009). Objective Bayesianism, Bayesian conditionalisation and voluntarism. *Synthese*, in press.