

Formal Epistemology: Lecture Notes

Horacio Arló-Costa
Carnegie Mellon University
`hcosta@andrew.cmu.edu`

Bayesian Epistemology

‘Radical probabilism doesn’t insist that probabilities be based on certainties; it can be probabilities all the way down, to the roots.’ [...] Radical probabilism adds the ‘non-foundational’ thought that there is no bedrock of certainty underlying our probability judgments. Richard Jeffrey.

One of the central ideas of radical probabilism is that any notion of traditional epistemology, if legitimate, should be derivable probabilistically. And if most of these epistemological notions turn not to be derivable in this way, the worst for them. *Eliminativism* of propositional attitudes is consistent with this view.

What I hope for is some reconciliation of the diverse intuitions of Bayesians and traditionalists, within a rather liberal probabilism. The old we might call defensive epistemology, for it concentrates on justification, warrant for, and defense of one's beliefs.

The whole burden of rationality has shifted from justification of our opinion to the rationality of change of opinion.

Van Fraassen: *Laws and Symmetry*. Oxford: Oxford University Press, 1989.

This does not mean that we have a general opinion to the effect that what people find themselves believing at the outset is universally likely to be true. It means rather that rationality cannot require the impossible. We believe that our beliefs are true, and our opinions reliable. We would be irrational if we did not normally have this attitude toward our own opinion. As soon as we stop believing that a hitherto held belief is not true, we must renounce it – on pain of inconsistency!

Laws and Symmetry. Oxford: Oxford University Press, 1989.

Personal or subjective probability entered epistemology as a cure for certain perceived inadequacies in the traditional notion of belief. But there are severe strains in the relationship between probability and belief. They seem too intimately related to exist as separate but equal; yet if either is taken as the more basic, the other may suffer. [...] I would like to propose a single unified account, which takes conditional personal probability as basic.

Fine-grained opinion, probability and the logic of full belief, JPL 24, 349.

There is a third aspect of opinion, besides belief and subjective grading, namely supposition. Much of our opinion can be elicited only by asking us to suppose something, which we may or may not believe. Here supposition will be a central ingredient of an unified probabilistic picture.

Fine-grained opinion, probability and the ,logic of full belief, JPL 24, 349.

Conditional probability: two traditions

One tradition is represented by Dubins' principle of Conditional Coherence: For all pairs of events A and B such that $A \cap B \neq \emptyset$:

(1) $Q(\cdot) = P(\cdot|A)$ is a finitely additive probability.

(2) $Q(A) = 1$, and

(3) $Q(\cdot|B) = Q[B](\cdot) = P(\cdot|A \cap B)$

When $P(A \cap B) > 0$, Conditional Coherence captures some aspects of De Finetti's idea of conditional probability *given an event*, rather than *given a σ -field*.

Conditional Coherence does not capture, nevertheless, important aspects of De Finetti's ideas about primitive conditional probabilities.

De Finetti's views

In almost all circumstances, and at all times, we find ourselves in a state of uncertainty. Uncertainty in every sense. [...] It would therefore seem natural that the customary modes of thinking, reasoning and deciding should hinge explicitly and systematically on the factor uncertainty as the conceptually pre-eminent and determinative element.

The opposite happens however: there is no lack of expressions referring to uncertainty, but it seems that these expressions, by and large, are no more than verbal padding. The solid, serious, effective and essential part of arguments, on the other hand, would be the nucleus that can be brought within the language of certainty - of what is certainly true or certainly false. It is in this ambit that our faculty of reasoning is exercised, habitually, intuitively and often unconsciously.

Thinking of a subset of truths as given (knowing, for instance, that certain facts are true, certain quantities have given values, or values between certain limits, certain shapes, bodies or graphs of given phenomena enjoy certain properties, and so on), we will be able to ascertain which conclusions, among those of interests, will turn to be - on the basis of the data - either certain (certainly true), or impossible (certainly false), or else possible.

Using a visual image, which at a later stage could be taken as an actual representation, we could say that the logic of certainty reveals to us a space in which the range of possibilities is seen in outline, whereas the logic of the probable will fill in this blank outline by considering a mass distributed upon it

The well-known Kolmogorovian alternative to the former view operates as follows. Let $\langle \Omega, \mathcal{B}, P \rangle$ be a measure space where Ω is a set of points, \mathcal{B} a σ -field of sets of subsets of Ω , with points w . Then when $P(A) > 0$, $A \in \mathcal{B}$, the conditional probability over \mathcal{B} given A is defined by: $P(.|A) = P(. \cap A) / P(A)$.

Of course, this does not provide guidance when $P(A) = 0$. For that the received view implements the following strategy. Let \mathcal{A} be a sub- σ -field of \mathcal{B} . Then $P(.|\mathcal{A})$ is a *regular conditional distribution* [rcd] of \mathcal{B} , given \mathcal{A} provided that:

(4) For each $w \in \Omega$, $P(\cdot|\mathcal{A})(w)$ is a probability on \mathcal{B} .

(5) For each $B \in \mathcal{B}$, $P(B|\mathcal{A})(\cdot)$ is an \mathcal{A} -measurable function.

(6) For each $A \in \mathcal{A}$, $P(A \cap B) = \int_A P(B|\mathcal{A})(w) dP(w)$

Blackwell and Dubins discuss conditions of *propriety* for rdc's. An rcd $P(\cdot|\mathcal{A})(w)$ on \mathcal{B} given \mathcal{A} , is proper at w , if $P(\cdot|\mathcal{A})(w) = 1$, whenever $w \in A \in \mathcal{A}$. $P(\cdot|\mathcal{A})(w)$ is improper otherwise.

Recent research has shown that when \mathcal{B} is countably generated, almost surely with respect to P , the rcd's on \mathcal{B} given \mathcal{A} are maximally improper [?]. This is so in two senses. On the one hand the set of points where propriety fails has measure 1 under P . On the other hand we have that $P(a(w)|\mathcal{A})(w) = 0$, when propriety requires that $P(a(w)|\mathcal{A})(w) = 1$.

It seems that failures of propriety conspire against any reasonable epistemological understanding of probability of the type commonly used in various branches of mathematical economics, philosophy and computer science.

To be sure finitely additive probability obeying Conditional Coherence is not free from foundational problems, but, by clause 2 of Conditional Coherence, each coherent finitely additive probability is proper. In addition Dubins shows that each unconditional finitely additive probability carries a full set of coherent conditional probabilities.

I shall start here with Conditional Coherence and I shall add the axiom of Countable Additivity only to restricted applications where the domain Ω , when infinite, is at most countable. Then I shall define qualitative belief from conditional probability.

(I) for any fixed A , the function $P(X|A)$ as a function of X is either a (finitely additive) probability measure, or has constant value 1.

(II) $P(B \cap C|A) = P(B|A)P(C|B \cap A)$ for all A, B, C in F .

The probability (*simpliciter*) of A , $pr(A)$, is $P(A|U)$. If $P(X|A)$ is a probability measure as a function of X , then A is *normal* and otherwise *abnormal*. Conditioning with abnormal events puts the agent in a state of incoherence represented by the function with constant value 1. Thus A is normal iff $P(\emptyset|A) = 0$. van Fraassen shows in [?] that supersets of normal sets are normal and that subsets of abnormal sets are abnormal. Assuming that the whole space is normal, abnormal sets have measure 0, though the converse need not hold (why?). In the following we

shall confine ourselves to the case where the whole space U is normal.

Probability Cores

A *core* as a set K which is normal and satisfies the *strong superiority condition* (SSC) i.e. if A is a nonempty subset of K and B is disjoint from K , then $P(B|A \cup B) = 0$ (and so $P(A|A \cup B) = 1$). Thus any non-empty subset of K is more “believable” than any set disjoint from K .

Main properties of cores: All non-empty subsets of a core are normal (Finesse). Cores are nested.

Lemma 1 (*Descending Chains*). *The chain of belief cores induced by a non-coreless 2-place function P cannot contain an infinitely descending chain of cores.*

Proof Assume by contradiction that there is a 2-place P , such that the chain of belief cores induced by it contains a core B_0 and a countable chain of cores $B_0, B_1, B_2, \dots, B_n, \dots$, such that B_0 is the outermost belief core of this subsystem of cores for P . Consider in addition a set of points x_j , with j in \mathbb{N} , such that for every index i and j in \mathbb{N} , $x_i \in B_i$, and $x_i \notin B_j$, if $j > i$.

For every index m , we have (by axiom A2 of belief cores) that

$$P(\{x_{m+1}\} \mid \{x_{m+1}\} + \{x_m\}) = 1$$

By axiom A4 the 1-place function $P(\dots \mid \{x_{m+1}\} + \{x_m\})$ is a probability function obeying finite additivity. Therefore:

$$(1) P(\{x_m\} \mid \{x_{m+1}\} + \{x_m\}) = 0$$

Now notice that the Multiplication Axiom guarantees that for all propositions A, B, C in the sigma-field F

$$(E) P(A \mid A \cup B \cup C) = 0 \text{ if } P(A \mid A \cup B) = 0.$$

To see that E is true it is enough to consider the following instance of the Multiplication Axiom:

$$P((A \cup B) \cap A \mid A \cup B \cup C) = P(A \cup B \mid A \cup B \cup C) \cdot P(A \mid (A \cup B) \cap (A \cup B \cup C))$$

Call C the countable set of points $\{x_0, x_1, \dots, x_m, \dots\}$. Therefore (1) and (E) guarantee that for every x_m in U:

$$(1') P(\{x_m\} \mid C) = 0$$

Since $C \subseteq B_0$, A4 implies that $P(\dots \mid C)$ is a 1-place probability function. By the Multiplication Axiom and the fact that $P(\dots \mid C)$ is *countably additive* we have the desired contradiction: $1 = P(C \mid C) = \sum P(x_i \mid C) = 0$.

□