KAIL Ch. 10 – Evidential Probability

Brian T. Miller

November 7, 2016

10.1 – Vague Probability

Goal of the chapter: to formulate a theory of 'evidential probability' that can account for the fact that even the propositions we treat as evidence are uncertain.

Three standard interpretations of probability:[1]

[1] NB that each of these are identity claims – they seek to explain what probabilities *are*, not just to make some weaker observation about probabililties.

Frequency: probabilities are actual or infinite-sequence frequencies

Problem: non-repeatable events (Obama winning the '08 election, a coin that lands heads in a particular instance) have probabilities of 1 or 0, but often that's not what the evidence supports

NB: observed frequencies might provide *good evidence* about actual or counterfactual frequencies, but the frequentist's claim is one of *identity*, which is much stronger.

Propensity: probabilities are properties of objects, like dispositions [TW doesn't consider this interpretation]

Problem: Propensities of non-repeatable events needn't have probabilities of 1 or 0, but as with frequency view there's no clear connection between evidential support and propensities (subject to the same NB above)

Degree of belief: (also: subjective probability, credence) a subjective mental state of the agent measuring degree of confidence

Problem: the fact that people become more confident in h upon learning that p doesn't imply that e is evidence for h – maybe people are just being irrational

Proposed solution: probabilities are the credences of perfectly rational agents, not potentially irrational agents like us.

Note: much of formal epistemology is dedicated to spelling out the norms governing credences. Those norms tend to be unrealistic for normal people (e.g. they tend to imply that logical omniscience is a norm of rationality), so it's quite common for formal epistemologists to retreat to this proposed solution: the norms in question apply only to ideal agents, and our obligations are to approximate those norms to the degree possible (this last bit isn't well understood).

Problem with the proposal: q is a highly non-obivous logical truth, so non-obvious that we have strong evidence that no one will have a rational high credence in q.[2]. But since q is a logical truth, a perfectly rational agent would have a credence of 1 in q*, and so

$$P(q \text{ \& no one will have a high credence in } q)$$
$$= P(\text{no one will have a high credence in } q).$$

Since *no one will have a high credence in q* is part of the evidence, that credence will be very high.

[2] perhaps we know that no one has ever had a high rational credence in the past, and we know that the world is about to end in a nuclear holocaust

Two reasons this is problematic:

1. It's Moore-paradoxical, so no ideally rational agent would ever believe it, so a perfectly rational agent would never believe it.[3] Presumably she'd know that someone has great evidence for q because she herself has great credence in q, so she'd have a very *low* credence in *q&no one has great credence in q*.

2. But it's perfectly rational for us (rationally imperfect as we are) to be highly confidence on the basis our evidence that no one will ever have high credence in p, so the evidential probability of ϕ on evidence e is *not* the credence that a rational agent would have in ϕ given evidence e

[3] Note that agents ideally rational in a strictly *formal* sense might believe all the Moore-paradoxical propositions they please – whatever the norm ruling out belief in Moore-paradoxical propositions turns out to be, it's not a formal one.

This shows that the probability of *q & no one will have a high credence in q* on our evidence is not the same as an ideal agent's credence in that proposition would be. To generalize: the probability of a hypothesis on our evidence — evidential probabilities — are not the credences of rational agents.

*Background: *probabilism* is the thesis that one norm of coherence for belief is probabilistic coherence. Why accept that thesis?

A Dutch Book is a series of bets s.t. no matter how things turn out, you lose money. If you're habitually inclined to make such bets then you're a money pump: people can keep making these bets with you and you'll continually pay out more than you bring in. There's something pathological about being a money pump: it's a sign of practical irrationality. A Dutch Book *Argument* aims to show that probabilistically incoherent beliefs are irrational because[4] they turn you into a money pump.

[4] given further plausible assumptions from decision theory about how beliefs translate and values into betting behavior

Example:

I'm 70% confident that the coin will land heads and 70% confident that it will land tails. My beliefs are incoherent: I now think that there's a 140% chance that it will land either heads or tails.

If I'm 70% confident that the coin will land heads, then I should be willing to pay $7 for a bet that pays $10 if it really does come up heads and nothing if it comes up tails: in that case I think that there's a 70% chance that I'll win $10 and a 30% chance I'll win $0, and ($10 x .7) + ($0 x .3) = $7.

For parallel reasons I'll be willing to pay $7 for a bet that pays $10 if it comes up tails and nothing if it comes up tails.

But that means that I'll be willing to pay $14 to take both bets, even though that's guaranteed to lose me $4 overall. So I'm a money pump.

So, decision theory requires that credences be probabilistically coherent.

So what *are* evidential probabilities?

> We should resist demands for an operational definition; such demands are as damaging in the philosophy of science as they are in science itself. To require mathematicians to give a precise definition of 'set' would be to abolish set theory. Sometimes the best policy is to go ahead and theorize with a vague but powerful notion. (211)

and

> Consider an analogy. The concept of possibility is vague and cannot be defined syntactically. But that does not show that it is spurious. In fact, it is indispensable. Moreover, we know some sharp structural constraints on it: for example, that a disjunction is possible if and only if at least one of its disjuncts is possible. The present suggestion is that probability is in the same boat as possibility, and not too much the worse for that. (211)

10.2 – Uncertain Evidence

In order to set up his formal system for expressing evidential probabilities, TW first criticizes rivals.

Subjective Bayesians think that new evidence is incorporated into one's credence function by a process of *conditionalization:*

BCOND $P_{new}(h) = P_{old}(h \mid e) = P_{old}(h\&e)/P_{old}(e)$

Remember P is a credence function: a function from propositions to numbers between 0 and 1 representing the agent's degree of confidence in those propositions.

Because functions aren't allowed to produce two different values for a single argument, if one's credence in h is going to be different at t_1 and t_2, the agent must have different credence functions at those different times. BCOND relates the credence functions that one has a different times: P_{old} is the credence function that one has before obtaining evidence e and P_{new} is the credence function that one has after obtaining e.

First criticism of BCOND: suppose that all credence revisions proceed via BCOND. In that case, once one conditionalizes on e, $P_{new}(e) = P_{new}(e \mid e) = 1$. The problem is that once a proposition has credence of 1, it's impossible to reduce that credence in light of any future evidence: if $P(e) = 1$ then for any future evidence f, $P(e \mid f) = P(e\&f)/P(f) = P(f/f) = 1$. Call this the *Invincibility of Evidence* objection.

But if conditionalizing on e requires having the highest possible confidence in it, and if it's impossible to reduce that credence on any future evidence, then my confidence in e is indefeasible, which seems wrong.

Second criticism of BCOND: in response to the first puzzle, some Bayesians retreat to a conception of evidence as phenomenal states on the idea that we can't be mistaken about those. But this move has trouble capturing the intersubjectivity of evidence as required in the sciences.

Bayesians can (sort-of) avoid this problem if the propositions updated upon are about the world, but not if they're about mental states. But propositions about the world are defeasible, while evidence propositions are not. Potential solution: generalize BCOND into JCOND:

JCOND $P_{new}(h) = \sum_i P_{old}(h \mid e_i) P_{new}(e_i)$

NB: BCOND is a special case of JCOND: the case in which there are only two propositions in the input partition, one of them weighted to 1 and the other to 0. Usually this will be a proposition and its negation, e.g. e and $\neg e$. It is common to simply fail to mention the partition elements weighted to 0, in which case one ends up updating on e alone; that's precisely what Classical Bayesians – those committed to BCOND – do.

Jeffrey conditionalization allows the credence in evidence propositions to change without going all the way to 1, so problem averted.[5]

TW's criticism of JCOND: it offers no account of where the input partitions come from. This is unlike BCOND, on which the inputs

[5] But if a credence *does* happen to go all the way to 1, it's stuck there, just like with BCOND.

are propositions, which can in turn be identified with one's propositional evidence, which can in turn be supplemented with an auxiliary theory of evidence such as TW's E=K hypothesis.

The invincibility of evidence objection comes from two claims:

PROPOSITIONALITY The evidential probability of a proposition is its probability conditional on the evidence propositions.

MONOTONICITY Once a proposition has evidential probability 1, it keeps it thereafter.

PROPOSITIONALITY entails that evidence propositions have probability 1, and MONOTONICITY ensures that they keep that probability.

JCOND rejects PROPOSITIONALITY but keeps MONOTONICITY

TW's suggestion: reject MONOTONICITY and keep PROPOSITIONALITY

Why reject MONOTONICITY? It implies an evidential asymmetry: we can gain new certainties but never lose them. But that's false: it's possible to forget some of your evidence, in which case it's probability should be reduced

TW's suggestion:

ECOND $P_\alpha(h) = P(h \mid e_\alpha) = P(h\&e_\alpha)/P_{old}(e_\alpha)$

P (with no subscript) is the One True Probability Function, representing the probability that each proposition should have in the absence of all evidence, i.e. it's *initial plausibility.*[6]

P_α is the probability function representing the credence function one ought to have in a case α, in which one has total body of evidence e_α.

P_α is equivalent to what would happen if you start out with P and then update upon e_α using BCOND or JCOND[7].

TW's objection to BCOND and JCOND is that they impose implausible constraints upon how the probability of propositions change over time: they prohibit future reductions in propositions that now have a credence of 1. On these rules, certainty is cumulative: once gained it can never be lost. The problem is exacerbated for BCOND because it requires that all evidence propositions be certain, i.e. have a credence of 1.

[6] NB: Bayesians have something similar: the credence function that one accepts in the absence of any evidence. Generally Bayesians accept many different 'starting credence functions' as epistemically permissible – the details are contentious (we can talk more about this if you'd like), but they all agree that rationally permissible starting credence functions must be probabilistically coherent (i.e. they must be *probability functions.* But the Bayesian's initial credence function is making claims about beliefs, or credences. What are evidential probabilities in the absence of evidence? Are we back to frequencies or propensities?

[7] e_α determines a partition, so no problem updating via JCOND

NB that ECOND avoids the problem by imposing no constraints upon how the probabilities of propositions can change over time. If at t_1 some proposition e is part of the evidence set e_α, then $P(e\&e_\alpha)/P_{old}(e_\alpha) = P(e_\alpha)/P_{old}(e_\alpha) = 1$, so $P(e \mid e_\alpha) = 1$, so $P_\alpha(e) = 1$. But if at t_2 my evidence β does not include e, then there's no constraint upon the probability of e at t_2. That's because the rule imposes no constraints at all upon what evidence one possesses at a given moment — that's determined outside of the model — and since we're told nothing about the nature of P, as long as e is not in β, the model imposes no constraints upon $P_\beta(e)$.

But, if no evidence is lost between t_1 and t_2[8] then $P_\beta(h) = P_\alpha(h \mid f)$, where f is just the conjunction of all of the propositions that are in β and not in α. When this happens the transition from P_α to P_β proceeds exactly as it would according to BCOND, which is itself a special case of ECOND.[9]

[8] and hence that the evidence in α is a proper subset of the evidence in β

[9] BCOND is a special case of ECOND in these situations because when the change in evidence is cumulative, they give the exact same prescription for updating probabilities; BCOND thinks all changes of evidence is cumulative, so it thinks that all updates are of this sort, but ECOND allows non-cumulative evidence changes, so it can handle a wider range of cases.

10.3 – Evidence and Knowledge

TW thinks that without an account of the nature of propositional evidence, formal epistemology is 'empty'.

TW thinks that your evidence consists in all the propositions that you know: E=K

He also thinks that subjective Bayesians think that your evidence consists in all the propositions that you *believe*: E=B

Problem with E=B: you could manipulate your evidence by manipulating your beliefs (assuming that's possible), thereby manipulating what it's rational for you to believe

BTM: TW presents this as a problem for subjective Bayesianism, but his view has a similar problem. Assume E=K. K entails B, so if some piece of evidence is inconvenient then by refusing to believe[10] it would ensure that I don't know it, so it's not part of my evidence. Of course this only works in one direction: by choosing to believe something I don't thereby come to know it, so although I could lose evidence in this way, I can't gain it, so the problem is worse for E=B than for E=K.

[10] this objection depends on belief voluntarism, but so does TW's. NB that both my version and TW's are evidence of a broader phenomenon of people manipulating their evidence in order to manipulate what their evidence supports. This is possible on a broad range of theories of evidence, and it's not entirely clear what to say about it, other than that it clearly involves some sort of irrationality.

But there's a deeper problem with this objection. Subjective Bayesians needn't accept E=B. Bayesianism imposes certain coherence norms on beliefs, but it doesn't imply that there aren't other norms of belief. It may be that one should conditionalize upon only rational beliefs, and in that case TW's objection fails (assuming that I can't generally be rational in believing p when I will myself to believe it).

End BTM.

The rest of the section rehashes previous material about how it's possible to follow a rule even it you're not always in a position to be able to do so, and hence it's not a problem that we're not always in a position to conform our beliefs in ECOND + E=K due to the fact that we're not always in a position to know what we know.

10.4 – Epistemic Accessibility

Point of this section is to utilize epistemic modal logic to develop a formal framework for combining ECOND and E=K. This logic has the following features:

- rather than cases, which are centered on specific agents, propositions are true at worlds, which are not
  - worlds are equivalent to mutually exclusive and jointly exhaustive sets of propositions
  - worlds needn't actually be possible: we allow worlds in which water is composed of XYZ (*water*, not twin-water)[11]
- a priori probability function P specifies the evidential probability that each world has conditional on no evidence.
  - The sum of the probabilities of all worlds is 1.
  - The probability of any specific proposition q is the sum of the probabilities of all of the worlds in which q is true.
  - If q is true in all possible worlds its probability is 1.
- Worlds are related by an accessibility relation R. A world w' is accessible to S from w iff every proposition that S knows in w is true at w'. In other words, R connects this world with all of the worlds that are consistent with the evidence that I have here.
  - Accessibility relations have directions: just because wRw', it doesn't follow that $w'Rw$. We can impose that (symmetry) condition if we want, but that's a further step (see below).
  - if some proposition p is consistent with what I know, then there exists a world accessible to me at which p is true. This provides the semantics for $\Diamond$: $\Diamond p$ is true at w iff there exists a world w' s.t. p is true at w' and wRw' (we allow that w might be identical to w')

[11] TW says that the worlds needn't be *metaphysically* possible. Though he says nothing about *logical* impossible worlds, he's going to need to include them if he want's to allow that a person would not be logically omniscient. Let's keep an eye on this issue as the section unfolds.

  - p follows from what I know iff p is true in every world accessible from w. This provides the semantics for $\Box$: $\Box p$ is true at w iff for any world w', if wRw' then p is true at w'.
  - Since knowledge is factive, each world sees (i.e. is accessible to) itself; R is *reflexive*.

- With this framework in place, we can combine ECOND and E=K
  - take e_w to be S's evidence in w
  - by E=K, e_w consists in all the propositions that S knows at w
  - by the factivity of knowledge, e_w is true at w
  - since e_w consists in all the propositions that S knows in w, e_w is true in all the worlds accessible to w (by our account of the accessibility relation)
  - P_w is the function specifying S's evidential probability for every proposition in w.
  - by ECOND, $P_w(\cdot) = P(\cdot \mid e_w)$. Intuitively, this says that the evidential probability of p for a person with evidence e_w is equal to the weighted sum the probabilities of p in each of the worlds consistent with evidence.

BTM: Interesting consequence of what's been said. TW has committed to the claim that: p follows from e_w iff p is true in every world consistent with e_w. In that case the evidential probability of p at w is 1. But, he also says that 'one need not know that which follows from what one knows.' (225) So on this account, although all known propositions have evidential probability 1, not all propositions with evidential probability 1 are known.

That's not an objection, just an observation. Remember that unlike the subjective Bayesian, TW's probabilities are not mental states like belief or (maybe) knowledge. But this leaves open the question of what rational demands the account imposes on agents: if p has an evidential probability of 1 on my evidence, what should my attitude be toward p? What are the rational norms governing evidential probabilities? The straightforward answer would be to say that if p has an evidential probability of 1 then S should believe it, but in that case it looks like rationality requires logical omniscience, a requirement that TW rejects: 'The account will not assume any general principle about knowledge, except that a proposition is true in any world in which it is known. In particular, it will not assume logical omniscience; if p and q are true in exactly the same worlds, one may know p and not know q.' (224)[12]

There's a second way of interpreting what's going on here. I've been assuming that the 'follows from' in the semantics of $\Box$ and

[12] I'm not sure what to make of the bit after the semi-colon: do all permissible failures of logical omniscience have that form? Am I still required to know all of the logical truths that don't share that form? NB that if I know one logical truth then one might think that all questions of logical omniscience are of this form.

$\Diamond$ is equivalent to 'is a logical consequence of'. If that's right then once e_w is determined, many facts about the worlds consistent with e_w are determined as well. In particular, if p really is a logical consequence of e_w then the worlds consistent with e_w must be p-worlds. But TW might instead be appealing to facts about the worlds in order to define the expression 'follows from'. In that case p could be a logical consequence of e_w, but some of the worlds consistent with e_w could be $\neg p$ worlds (remember that the worlds don't actually have to be *possible* worlds). In that case, what TW has told us about what 'follows from' what implies that p does not in fact follow from e_w, in spite of their logical relationship. But if some of the worlds consistent with e_w are $\neg p$ worlds, then $P_w(p) \leq 1$, so the puzzle of why probability 1 propositions needn't be known (or believed) does not arise.

I find TW in KAIL 10 to be very unclear on this point, but in his response to Kaplan he's much more explicit: the connection between evidential probabilities and credences is pretty weak. In particular, p's having a probability of 1 conditional on my evidence *does not* require that I have a credence of 1 in p. Of course this leaves open the question of what evidential probabilities mean for credences...

[End BTM]

In this framework, constraints on R are constraints upon knowledge: Kp implies $\Box p$. What else can be said about R?

TW has told us that R is reflexive. This ensures that in our logic $\Box p \rightarrow p$ (this is the 'M' axiom). Suppose that's false, and that in w I know that p. That implies $\Box p$. A failure of reflexivity means that p would be false in w, i.e. that in w I know something that's false. But knowledge is factive, so that's impossible. So R must be reflexive.

He also tells us that R can't be transitive. Transitivity ensures that $\Box p \rightarrow \Box\Box p$ (this is the '4' axiom). This is essentially the KK principle (actually something slightly weaker, but presumably TW thinks that's false too).

He also tentatively rejects the symmetry[13] of R, which ensure that $p \rightarrow \Box\Diamond p$ (this is the 'B' axiom). This says that if p is true then in every world w' consistent with my evidence, there is at least one world w'' s.t $w'Rw''$ and p is true at w''. Problem case:

> Let x be a world in which one has ordinary perceptual knowledge that the ball taken from the bag is black. In some world w, the

[13] There are lots of possible constraints upon R, and modal logics are individuated by which of them are imposed. Presumably TW singles out M, 4, and B because a logic that accepts these three is called S5, which the most familiar modal logic for most philosophers. The modal logic that TW is utilizing here is just called M (after the single axiom that it accepts).

ball taken from the bag is red, but freak lighting conditions cause it to look black, and everything which one knows is consistent with the hypothesis that one is in x. Thus x is accessible from w, because every proposition which one knows in w is true in x; but w is not accessible from x, because the proposition that the ball taken from the bag is black, which one knows in x, is false in w. Let p be the proposition that the ball taken from the bag is red. In w, p is true, but that p is consistent with what one knows does not follow from what one knows, for what one knows is consistent with the hypothesis that one knows $\neg p$.