Some Recent Fallacies of Approximation in Bayesian Confirmation Theory

Branden Fitelson
University of California–Berkeley

- Several recent Bayesian discussions make use of "approximation"
  - Earman on the Quantitative Old Evidence Problem
  - Vranas on Quantitative Approaches to the Ravens Paradox
  - Dorling's Quantitative Approach to Duhem–Quine
  - Strevens's Quantitative Approach to Duhem–Quine
  - [There are also examples not involving confirmation: *E.g.*, Carlstrom & Hill's "Triviality Proof" – time permitting]
- Each of these discussions is illicit [or enthymematic] in some way
- The arguments can be salvaged, but they may be less compelling
- My main purpose today is to urge *caution* concerning $\approx$ and Pr
- Coda: *The Very Idea* of "Quantitiative Confirmation"

Earman on the Quantitative Old Evidence Problem I

- One possible way out of the problem of old evidence is to simply refuse to assign probability 1 to any contingent proposition (thus, to give up on learning by strict conditionalization – in favor, say, of the Jeffrey style).
- This avoids the *qualitative* old evidence problem, as Earman explains:

> The original problem of old evidence would vanish for Bayesian personalists for whom $\Pr(E) \neq 1$, with Pr interpreted as personal degree of belief. ... However, denying that $\Pr(E) = 1$ only serves to trade one version of the old-evidence problem for another. Perhaps it was not certain in November 1915 that the true value of the anomalous advance was roughly 43" of arc per century, but most members of the scientific community were pretty darn sure, e.g., $Pr(E) = .999$. Assuming that Einstein's theory does entail E, we find that the confirmatory power $C(T, E)$ of E is $\Pr(T) \times .001/.999$, which is less than .001002. This is counterintuitive, since, to repeat, we want to say that the perihelion phenomenon did (and does) lend strong support to Einstein's theory.

Earman on the Quantitative Old Evidence Problem II

- Earman is using the difference measure $d(H, E) = \Pr(H \mid E) - \Pr(H)$ to measure the degree to which E confirms H, which has this property:

(1) If $H \vDash E$ and $\Pr(E) \approx 1$, then $\mathfrak{c}(H, E) \approx 0$.

- This "approximation argument" of Earman's has two potential flaws.
  - It is measure-sensitive. (1) holds for d and the log-ratio measure $r(H, E) = \log\left[\frac{\Pr(H \mid E)}{\Pr(H)}\right]$, but not for the log-likelihood ratio measure l or the Joyce-Christensen measure s (contrary to Earman's claim):
  $$l(H, E) = \log\left[\frac{\Pr(E \mid H)}{\Pr(E \mid {\sim}H)}\right],\ s(H, E) = \Pr(H \mid E) - \Pr(H \mid {\sim}E)$$
  - The argument presupposes that E is *deductive evidence* ($H \vDash E$).
- The latter is not so bad, since $H \vDash E$ can be relaxed for both d and r.[a]
- The former is important, since the facts about l and s are rather different.

[a] More rigor: $\forall\, \epsilon \in (0, \frac{1}{10})$, if $\Pr(E) \geqslant 1 - \epsilon$, then $d(H, E) \leqslant \epsilon$, and $r(H, E) \leqslant \log(1 + 2\epsilon)$.

Earman on the Quantitative Old Evidence Problem III

- For l, the salient, *deductive* "approximation theorem" is the following.

  If $H \vDash E$, $\Pr(E) \approx 1$, *and* $\Pr(H)$ is *not large*, then $l(H, E) \approx 0$.[a]
- So, a similar argument works for l (non-deductive case is subtle!). What about s? Interestingly, s avoids all "approximation theorems" *so far*. But,

  If $\Pr(E) \approx 1$, *and* $\Pr(H) \approx 0$, then $l(H, E) \approx 0$, and $s(H, E) \approx 0$.[b]
- Note: the following "normalized" version of l avoids *even this* result:
$$l_N(H, E) = \frac{1}{\Pr({\sim}E)} \cdot \log\left[\frac{\Pr(E \mid H)}{\Pr(E \mid {\sim}H)}\right]$$
- But, like s, l_N violates the following *desideratum*, and so is inadequate:
$$\Pr(H \mid E_1) \geqslant \Pr(H \mid E_2) \Rightarrow \mathfrak{c}(H, E_1) \geqslant \mathfrak{c}(H, E_2).$$
- So, there still seems to be something right about Earman's claim (*maybe* – see my Coda!), but more care is needed to establish his conclusion.

[a] *E.g.*, $\forall\, \epsilon \in (0, \frac{1}{10})$, $H \vDash E$, $\Pr(H) \leqslant \frac{1}{2} - \epsilon$ and $\Pr(E) \geqslant 1 - \epsilon \Rightarrow l(H, E) \leqslant \log(1 + 2\epsilon)$.
[b] *E.g.*, $\forall\, \epsilon \in (0, \frac{1}{10})$, $\Pr(H) \leqslant \epsilon$ and $\Pr(E) \geqslant 1 - \epsilon \Rightarrow l(H, E) \leqslant \log(1 + 2\epsilon)$ and $s(H, E) \leqslant 2\epsilon$.

Vranas on Quantitative Approaches to the Ravens Paradox I

- An object a is to sampled at random from the universe. Let Ra be the claim that a is a raven, Ba be the claim that a is black, and H be the hypothesis that all ravens are black. Now, consider the following probabilistic conditions:

(*i*) $\Pr(\sim Ba) \gg \Pr(Ra) \approx 0 \quad [i.e., \Pr(Ra)/\Pr(\sim Ba) \approx 0]$

(*ii*) $\Pr(Ra \mid H) = \Pr(Ra)$

(*iii*) $\Pr(\sim Ba \mid H) = \Pr(\sim Ba) \quad [\Dashv\vDash \Pr(Ba \mid H) = \Pr(Ba)]$

- It is well-known that these conditions are *sufficient* for the following:

(*iv*) $\Pr(H \mid Ra \,\&\, Ba) \gg \Pr(H \mid \sim Ra \,\&\, \sim Ba) > 0$, and

(*v*) $\Pr(H \mid \sim Ra \,\&\, \sim Ba) \approx 0$

- Vranas points out that, while (*i*) may seem plausible, (*ii*) and (*iii*) are far less clear, and not much in the way of justification has been given for them. He suspects some "conditional principle of indifference" (CPI) underlies them.
- Vranas argues that (*iii*) is not only *sufficient* [given (*i*) and (*ii*)] for (*iv*), but that (*iii*) is "approximately necessary" for (*v*), given (*i*). Let's take a closer look.

Vranas on Quantitative Approaches to the Ravens Paradox II

- There is a scope ambiguity in Vranas' argument. He argues that *independence* assumptions [like (*iii*)] are ill-motivated, since they rest on some (CPI). He then claims to show that (*iii*) is "approximately necessary" for (*v*), given (*i*).
- But, what he *proves* is actually (something like) the following theorem:

$$\text{Given } (i),\ \Pr(H \mid \sim Ra \,\&\, \sim Ba) \approx 0 \Rightarrow \Pr(\sim Ba \mid H) \approx \Pr(\sim Ba).$$

- So, he has *not* shown that (*iii*) is "approximately necessary" [given (*i*)] for (*v*). What he has shown is that "approximately (*iii*)" [call this (*iii*)*, for short] is *necessary* for (*v*), given (*i*). But, what, exactly *is* "approximately (*iii*)" [(*iii*)*]?
- It would be a mistake to think of (*iii*)* as anything like *independence*. It is, in fact, far weaker than this. Note that all of the following entailments hold:

$$\Pr(\sim Ba \mid H) = \Pr(\sim Ba) \Dashv\vDash \Pr(H \mid \sim Ba) = \Pr(H) \Dashv\vDash$$
$$\Pr(\sim Ba \mid H) = \Pr(\sim Ba \mid \sim H) \Dashv\vDash \Pr(H \mid \sim Ba) = \Pr(H \mid Ba)$$

Vranas on Quantitative Approaches to the Ravens Paradox III

- But, on the other hand, all of the following *non*-entailments hold:

$$\Pr(\sim Ba \mid H) \approx \Pr(\sim Ba) \not\Dashv\not\vDash \Pr(H \mid \sim Ba) \approx \Pr(H) \not\Dashv\not\vDash$$
$$\Pr(\sim Ba \mid H) \approx \Pr(\sim Ba \mid \sim H) \not\Dashv\not\vDash \Pr(H \mid \sim Ba) \approx \Pr(H \mid Ba)$$

- "Approximate independence" is stronger than $\Pr(\sim Ba \mid H) \approx \Pr(\sim Ba)$. It entails several other approximate equalities, which do *not* follow from (*iii*)*.
- Vranas' main argument against assuming *independence* in this sort of context is that it rests on some (CPI), the unrestricted application of which will lead to inconsistent sets of constraints. This is because *independence* assumptions are actually quite strong, and they need to be imposed with care (and warrant).
- But, Vranas does not (and could not) show any such thing about (*iii*)*, since it is a far weaker constraint than the locution "independence is approximately necessary" would imply. There are many (weak) varieties of "approximate independence", but there is only one (rather strong) notion of *independence*.
- Again, the moral is that using "approximation theorems" can be misleading.

Dorling's Quantitative Approach to Duhem–Quine I

- Most of the details of the Duhem–Quine problem (or Dorling's approach to it) are unimportant for our purposes today. What's important is that Dorling makes use at one point of something similar to the following assumption:

(†) If one knows $\Pr(E)$ and $\Pr(E \mid H)$ (with perfect precision), one can approximate $\Pr(H \mid E)$ *via* a sufficiently accurate approximation of $\Pr(H)$.

- In Dorling's case, he approximates $\Pr(H) \doteq 1 - \sum_i \Pr(H_i)$, where $\{H_i\}$ are the "not-highly-implausible alternatives" to H (in some partition logical of $\sim H$).
- The confirmation-theoretic version of this sort of assumption is as follows:

(‡) If one knows $\Pr(E)$ and $\Pr(E \mid H)$ (with perfect precision), one can approximate $\mathfrak{c}(H, E)$ *via* a sufficiently accurate approximation of $\Pr(H)$.

- The former principle is false, and the latter is true only for inadequate measures of confirmation. Basically, this is because highly implausible alternatives to H can only be ignored under certain circumstances.

Dorling's Quantitative Approach to Duhem–Quine II

- Let $\Pr^*(H)$ be an approximation of $\Pr(H)$ such that $|\Pr(H) - \Pr^*(H)| \leqslant \epsilon$. How would we use $\Pr^*(H)$, $\Pr(E)$, and $\Pr(E \mid H)$ to approximate $\Pr(H \mid E)$?
- Naively, we could try something akin to Bayes's Theorem, and approximate:

$$\Pr^*(H \mid E) \doteq \frac{\Pr(E \mid H) \cdot \Pr^*(H)}{\Pr(E)}$$

- $\Pr^*(H \mid E)$ can be quite different from $\Pr(H \mid E)$. In fact, no matter how you calculate $\Pr(H \mid E)$ here, it won't always lead to a good approximation.
- More rigorously, we can have two probability distributions Pr and Pr* that agree perfectly on $\Pr(E)$ and $\Pr(E \mid H)$, and "almost perfectly" on $\Pr(H)$, but disagree wildly on $\Pr(H \mid E)$. Thomason and I discuss concrete examples.
- In fact, there are algorithms for generating pairs Pr and Pr* (as functions of ϵ) such that $\Pr(E) = \Pr^*(E)$, $\Pr(E \mid H) = \Pr^*(E \mid H)$, $|\Pr(H) - \Pr^*(H)| \leqslant \epsilon$, but $|\Pr(H \mid E) - \Pr^*(H \mid E)| \geqslant 1 - \epsilon$. In this sense, (†) is *very* false.

Dorling's Quantitative Approach to Duhem–Quine III

- An Explanation: while the catch-all likelihoods $\Pr(E \mid {\sim}H)$ and $\Pr^*(E \mid {\sim}H)$ cannot differ by that much (that *is* a theorem), the likelihood *ratios* defined in terms of them $\frac{\Pr(E \mid H)}{\Pr(E \mid {\sim}H)}$ and $\frac{\Pr^*(E \mid H)}{\Pr^*(E \mid {\sim}H)}$ *can* (and *do* in the examples we give).
- Roughly, this happens when: (a) H itself has a very low likelihood, (b) there is at least one very improbable alternative H' that has a very high likelihood, and (c) H' is so improbable that it is not included in the approximation $\Pr^*(H)$.
- This explains both why the assumption about posteriors (†) is false (since likelihood ratios are important in determining them), and why the assumption about degrees of confirmation (‡) is false, for *adequate* $\mathfrak{c}$-measures, like l.
- There are confirmation measures that satisfy (‡). For instance, the likelihood *difference* (advocated by Nozick, and others): $\Pr(E \mid H) - \Pr(E \mid {\sim}H)$.
- But, the likelihood difference is inadequate, since it violates the *desideratum*:

$$\Pr(H \mid E_1) \geqslant \Pr(H \mid E_2) \Rightarrow \mathfrak{c}(H, E_1) \geqslant \mathfrak{c}(H, E_2).$$

Strevens's Quantitative Approach to Duhem–Quine

- Again, the details of Strevens's approach to Duhem–Quine are not important (see my forthcoming *BJPS* discussion with Waterman for more criticisms).
- What's important, for our purposes, is that Strevens presupposes that the following inferences involving various "approximations" are cogent.
  - ① $H \,\&\, A \vDash {\sim}E$ and $\Pr(E \mid H \,\&\, {\sim}A) \approx \Pr(E \mid {\sim}(H \,\&\, A))$ $\Rightarrow \Pr(H \mid {\sim}(H \,\&\, A)) \approx \Pr(H \mid E)$.[a]
  - ② $H \,\&\, A \vDash {\sim}E$ and $\Pr(H \mid {\sim}(H \,\&\, A)) \approx \Pr(H \mid E)$ $\Rightarrow \mathfrak{c}(H, {\sim}(H \,\&\, A)) \approx \mathfrak{c}(H, E)$.[b]
- But, possibly: $H \,\&\, A \vDash {\sim}E$ and $|\Pr(E \mid H \,\&\, {\sim}A) - \Pr(E \mid {\sim}(H \,\&\, A))| \leqslant \epsilon$, while $|\Pr(H \mid {\sim}(H \& A)) - \Pr(H \mid E)| \geqslant 1 - \epsilon$, for any $\epsilon > 0$. Similarly for ②.
- Strevens's notation misleads. *E.g.*, "$\Pr(H \mid E) = \Pr(H \mid {\sim}(H \,\&\, A)) + \delta$" makes it sound as if "small δs" can be ignored. But, nonlinearity reigns here!

[a] Near-①: $H \& A \vDash {\sim}E$ and $\Pr(E \mid H \& {\sim}A) = \Pr(E \mid {\sim}(H \& A)) \Rightarrow \Pr(H \mid {\sim}(H \& A)) = \Pr(H \mid E)$.

[b] Near-②: Our *desideratum* $\Pr(H \mid E_1) \geqslant \Pr(H \mid E_2) \Rightarrow \mathfrak{c}(H, E_1) \geqslant \mathfrak{c}(H, E_2)$ does imply that $\Pr(H \mid {\sim}(H \,\&\, A)) = \Pr(H \mid E) \Rightarrow \mathfrak{c}(H, {\sim}(H \,\&\, A)) = \mathfrak{c}(H, E)$. So, again, there's a theorem "nearby".

Another Contemporary Approximation Fallacy (first noted by Alan Hájek)

- Carlstrom and Hill's "triviality proof" for the (CCCP) $\Pr(p \to q) = \Pr(q \mid p)$:
  - Let p and q be contingent and logically independent, and assume there are worlds at which $p \to q$, p, and q are all true. Let w_1 be such a world.
  - Assume $\to$ is not a truth function, and let w_2 and w_3 be worlds that agree on the truth values of p and $p \,\&\, q$, s.t. $p \to q$ is false at w_2 but true at w_3.
  - Assume Pr divides almost all probability roughly equally between w_1 and w_2, and Pr′ does the same w.r.t w_1 and w_3. Then, $\Pr(p \to q) \approx \frac{1}{2}$, and $\Pr'(p \to q) \approx 1$. Thence, $\Pr(p \to q) \not\approx \Pr'(p \to q)$. But, we also have:
    1. $\Pr(p \,\&\, q) \approx \Pr'(p \,\&\, q)$, and
    2. $\Pr(p) \approx \Pr'(p)$, with both of these positive.
    3. "∴" $\Pr(q|p) \approx \Pr'(q|p)$, assuming the ratio definition of CP.

    So, P and P' cannot both be CCCP-functions for $\to$, which seems absurd.
- "∴" is fallacious: (1) & (2) $\not\Rightarrow$ (3) $\dfrac{\Pr(q \,\&\, q)}{\Pr(p)} \approx \dfrac{\Pr'(q \,\&\, q)}{\Pr'(p)}$, for usual reasons.

Another Interesting Aspect of "Approximation" in Pr-Theory

- It is well-known (Simpson's paradox) that conditional independences can be made to disappear on all cells of a suitable partition of the conditioning variable. *E.g.*, it is possible that: $E \perp\!\!\!\perp H \mid \top$, but $E \not\perp\!\!\!\perp H \mid X$, and $E \not\perp\!\!\!\perp H \mid {\sim}X$.
- It is not widely discussed that independences can be made to disappear – not only by conditionalizing – but also by *perturbing* a probability function.
- In fact, given any probability model $\mathcal{M} = \langle \Omega, \mathscr{F}, \Pr \rangle$, there exists another $\mathcal{M}' = \langle \Omega, \mathscr{F}, \Pr' \rangle$ such that Pr and Pr′ are arbitrarily close, and none of the independence relations imposed by Pr are respected by Pr′ (*modulo* logic).
- What do I mean by "arbitrarily close"? Interestingly, it doesn't much matter. Use whatever measure of divergence between probability distributions you like, and pick any ϵ for your threshold of "arbitrarily close". This won't help.
- What grounds confidence in $\perp\!\!\!\perp$ judgments, when there are *indistinguishable* distributions according to which all of our independence judgments are false?

Coda: *The Very Idea* of "Quantitative Confirmation" I

- We've seen a lot of claims of the form $\mathfrak{c}(H, E)$ is "small" or "large" or "approximately the same as (some quantity)". Are these claims even *meaningful*? Is there even such a thing as *quantitative* confirmation?
- Analogy: I might say to you that the temperature of an object a [$T(a)$] is "small". But, what does this mean? On one temperature scale, it might be "small", but on another it might be "large". Which is right? Is there a FOTM?
- Still, we must agree on which objects are warmer than which [$T(a) \geqslant T(b)$]. That is, our scales must agree on *ordinal comparisons* (else they won't both be *temperature* scales!). Shouldn't we say the same about "measures" $\mathfrak{c}(H, E)$?
- I think we *should*: measures that are ordinally equivalent should be considered equivalent *full stop*. People seem to think that cardinal confirmation claims make sense. But aren't they really implicitly *ordinal*?
- And, do we really need "small" or "large" claims in confirmation theory anyway? What work are they doing? Are they worth the theoretical overhead?

Coda: *The Very Idea* of "Quantitative Confirmation" II

- What do I mean by "theoretical overhead"? Recall that only certain measures of confirmation will support certain ordinal, comparative confirmation claims.
- As a result, arguments are needed to favor some measures over others, for various *comparative* purposes. That's already some "theoretical overhead".
- If we also insist that certain "small" or "large" or "approximately the same as" claims are correct (in certain contexts), then we need *even stronger* arguments to rule-out *even more* candidate measures. That's *even more* "overhead".
- In the case of ordinal structure, we have very simple, intuitive desiderata that get us down to a very small number of ordinal equivalence classes of measures. But, in the case of "cardinal structure", this is not so (seems to me).
- In my opinion, *even if* cardinal confirmation claims could be given a precise meaning (relative to $\mathfrak{c}$s, somehow "finely" individuated), there just aren't any compelling desiderata that can adjudicate between ordinally equivalent $\mathfrak{c}$s.
- Absent such desiderata, I think much of this cardinal $\mathfrak{c}$-talk is poppycock.

Coda: *The Very Idea* of "Quantitative Confirmation" III

- Richard Royall has recently claimed that there is a universal calibration or standard for "strong confirmation": a likelihood-ratio of at least eight.
- This strikes me as implausible. Why think there is a single meaning of "strong confirmation" that does not depend on any contextual or pragmatic factors?
- To be fair, I do think there are *some* such universal truths about "strong confirmation", but they are *a priori* and determined by deductive logic alone.
- *E.g.*, if $E \vDash H$, then E provides "strong evidence" for H (this is an *infinite* LR!). But, this is a *consequence* of the following universal *ordinal* principle:

(**) If $E \vDash H$, then $\mathfrak{c}(H, E) \geqslant \mathfrak{c}(H', E')$.

- The log-likelihood ratio measure l (and ordinal equivalents!) is the only $\mathfrak{c}$ we've seen today that satisfies (**). d, r, and s violate it. And, *of all relevance measures I have even seen, only l satisfies both (**) and our desideratum*:

$$\Pr(H \mid E_1) \geqslant \Pr(H \mid E_2) \Rightarrow \mathfrak{c}(H, E_1) \geqslant \mathfrak{c}(H, E_2).$$