Probability Notes I

PHIL 735 Week 6
Matt Kotzen and Jim Pryor

October 11, 2023

1 Graded Belief

- So far, we've been focusing on *binary* (or perhaps ternary, if suspension is a distinct attitude) conceptions of belief
- For most of the remainder of the course, we'll be focusing on *graded* conceptions of belief
- Why might we want a graded conception of belief?
- And what is the relation between graded belief and binary belief?
  - Can graded belief be reduced to binary belief in propositions that somehow involve probability or chance?
  - Can binary belief be reduced to graded belief above a particular threshold? If so, is that threshold fixed, or does it vary based on context? If the latter, what causes the threshold to move?
  - Might both notions of belief be important, perhaps for different purposes, and not (fully) reducible to the other? If so, are there metaphysical and/or normative connections between the two types of belief? What's going on with someone who (allegedly) binary-believes p, and yet also has a graded-belief in p of .1?

2 An Abridged History of Confirmation Theory

- Confirmation theory in its modern form (beginning with Ayer and Hempel) began as an attempt to formally explicate the notion of "support" between statements
  - From the beginning, there was a hope that we could do this purely syntactically, just as we can for deductive inference

- Hypothetico-deductivism
  - Motivated by the thought that a hypothesis is (to some degree) supported by the data that it entails
  - HD Confirmation: E supports H iff there are true "auxiliary hypotheses" $A_1, A_2, \ldots, A_n$ such that (a) $A_1 \wedge A_2 \wedge \cdots \wedge A_n$ does not entail E, but (b) $H \wedge A_1 \wedge A_2 \wedge \cdots \wedge A_n$ entails E (but not $\neg E$).
  - But, as Duhem had already pointed out, auxiliary hypotheses that figure in confirmation relations are, like the hypothesis under test, fallible conjectures based on inconclusive evidence
    - This led Quine in "Two Dogmas" to insist that confirmation is holistic—i.e., that evidence never confirms or disconfirms any hypothesis in isolation
  - One limitation of the H-D approach: since H-D confirmation is entailment-based, H-D confirmation doesn't seem to apply to statistical hypotheses like "The coin is strongly heads-biased" or "The coin is heads-biased to degree .9" — no pattern of heads and tails is entailed by either of these hypotheses.
  - But, we're going to see later that the core idea of H-D confirmation survives in modern Bayesian approaches
- Hempel
  - Hempel's four "conditions":
    - (1) Entailment Condition: If E entails H, then E confirms H
    - (2) Special Consequence Condition: If E confirms H, and H entails H', then E also confirms H'
    - (3) Special Consistency Condition: If E confirms H, and H is incompatible with H', then E does not confirm H'
    - (4) Converse Consequence Condition: If E confirms H, and H is entailed by H', then E also confirms H'
  - But, as Hempel realized, any confirmation relation satisfying (1)–(4) will hold between every pair of propositions, which trivializes the confirmation relation
    - Actually, (1) and (4) are enough for the triviality result. Consider arbitrary p and q. By (1), p confirms $p \vee q$. By (4), p confirms q.
  - Hoping to preserve as much of (1)–(4) as possible, Hempel's solution was to restrict (4) to cases where H is obtained from H' by instantiation. (He held onto 1–3.)
  - In the preface to the 2nd edition of his *Logical Foundations of Probability*, Carnap accused Hempel of conflating an "absolute" notion of confirmation, on which H is highly supported given E, with an "incremental" notion of confirmation, on which E increases the evidential support for H

    - * There's a reasonable debate to be had about whether Carnap's charge against Hempel was legitimate
      * Regardless, (1)–(3) are plausible for absolute confirmation, but (4) is not
        · Example?
      * With regard to incremental confirmation:
        · (1) fails when H is already certain
        · (2)–(4) all fail too. Examples?
  - Hempel also endorsed a famous condition that is reasonably plausible for incremental confirmation, but totally implausible for absolute confirmation
    * Nicod's criterion: Universal generalizations of the form "All F's are G" are confirmed by any statement of the form "x is both F and G"
    * Hempel (and nearly everyone else except Quine) also accepted the Equivalence Condition (EC): If H is logically equivalent to H', and E confirms H, then E also confirms H'
      · Actually, EC is plausible for both absolute and incremental confirmation
      · Notice too that EC is a special case of both (2) and (4).
    * But, Nicod's criterion and EC give rise to the Ravens Paradox

- Carnap's system
  - Suppose we have just one predicate, R. And suppose we have three individuals: a, b, and c.
  - Then, we have 8 "state descriptions"
    * $Ra \wedge Rb \wedge Rc$
    * $Ra \wedge Rb \wedge \neg Rc$
    * $Ra \wedge \neg Rb \wedge Rc$
    * $Ra \wedge \neg Rb \wedge \neg Rc$
    * $\neg Ra \wedge Rb \wedge Rc$
    * $\neg Ra \wedge Rb \wedge \neg Rc$
    * $\neg Ra \wedge \neg Rb \wedge Rc$
    * $\neg Ra \wedge \neg Rb \wedge \neg Rc$
  - First Idea ($c^{\dagger}$): assign equal probability to every state description, such that they sum to 1. When one state description gets ruled out, reassign the probability associated with that state description to every "live" state description, equally.
    * On this approach, learning that one or more of the individuals satisfies predicate R can never confirm the hypothesis that others will too.
    * For example, we start off with each state description having probability $1/8$. Rb and Rc each have probability $1/2$.

    - Now we learn Ra. That rules out state descriptions 5–8, and the probabilities for state descriptions 1–4 go up from 1/8 to 1/4. But Rb and Rc still both have probability 1/2.
- Second Idea (c^*):
    - Say that two state descriptions S1 and S2 are "permutations" of each other if one can be obtained from the other by relabeling the individuals
    - So, for example, $\neg Ra \wedge Rb \wedge Rc$ and $Ra \wedge \neg Rb \wedge Rc$ are permutations of each other
    - A structure description is a disjunction of state descriptions, each of which is a permutation of the others. (Here, there are four structure descriptions, corresponding to the number of individuals satisfying R: 0, 1, 2, or 3.)
    - Now, assign equal probability to each of the structure descriptions. Each state description belonging to a particular structure description is given an equal share of the probability assigned to that structure description.
    - When one state description gets ruled out, reassign the probability associated with that state description to every "live" state description, proportionally.
    - On this approach, learning that one or more individuals satisfies predicate R *can* confirm the hypothesis that others do too.
    - For example, using the example from above, here are the structure descriptions:
        - "3": $(Ra \wedge Rb \wedge Rc)$
        - "2": $(Ra \wedge Rb \wedge \neg Rc) \vee (Ra \wedge \neg Rb \wedge Rc) \vee (\neg Ra \wedge Rb \wedge Rc)$
        - "1": $(Ra \wedge \neg Rb \wedge \neg Rc) \vee (\neg Ra \wedge Rb \wedge \neg Rc) \vee (\neg Ra \wedge \neg Rb \wedge Rc)$
        - "0": $(\neg Ra \wedge \neg Rb \wedge \neg Rc)$
    - We start off with each structure description having probability 1/4; hence, the three state descriptions in "2" and "1" get 1/12 each
    - As before, Rb and Rc both have probability 1/2
    - Now we learn Ra. That rules out state descriptions 5–8, which collectively had probability 1/2. So, state descriptions 1–4 all double in probability. $(Ra \wedge Rb \wedge Rc)$ goes up from 1/4 to 1/2. $(Ra \wedge Rb \wedge \neg Rc)$ goes up from 1/12 to 1/6. $(Ra \wedge \neg Rb \wedge Rc)$ goes up from 1/12 to 1/6. $(Ra \wedge \neg Rb \wedge \neg Rc)$ goes up from 1/12 to 1/6.
    - So, the probability of Rb goes up from 1/2 to 1/2+1/6=2/3. Ditto for Rc.
- In later work, Carnap introduced systems (the λ systems) in which different predicates could be more or less sensitive to evidence.

3 The Probability Axioms

- We start with a set of mutually exclusive and jointly exhaustive possibilities, which is designated Ω.
- Then, we define an "σ-algebra" or "σ-field" F on Ω: i.e., F is a nonempty set of subsets of Ω, which is closed under complement, countable unions, and countable intersections. Think of the members of F as propositions that get assigned a probability.
- Finally, we define a probability function p from F to the [0,1] interval of $\mathbb{R}$ that obeys the following three axioms, for all propositions A, $B \in F$:
  - Non-negativity: $p(A) \geq 0$
  - Normality: $p(\Omega) = 1$
  - Countable additivity: For any countable set $\{A, B, C, \dots\} \subseteq F$ such that $A \cap B \cap C \cap \dots = \emptyset$, $p(A \cup B) = p(A) + p(B) + p(c) + \dots$
- Several noteworthy consequences of these axioms are listed on Titelbaum p. 34
- Probabilism is the view that our credences are rationally constrained to be probabilities — i.e., to obey the formal treatment given above

4 Conditional Probability

- A conditional probability is the probability of some proposition conditional on (i.e., supposing) some proposition's truth; the conditional probability of A, conditional on (or "given") B, is written as $p(A|B)$.
- Conditional probabilities are standardly introduced by way of a definition: $p(A|B) = p(A \wedge B)/p(B)$, where $p(B) \neq 0$ (though not everyone regards this is a definition).
- When we talk about conditionalization in more detail next week, we'll see that the Rule of Conditionalization is the rule that, when an agent learns exactly B, they ought to "update" all of their credences so that their new credence in X is whatever their old conditional credence in X conditional on B was, for each X in F.
- Some theorems of note:
  - Law of Total Probability (Titelbaum p. 59)
  - Bayes's Theorem (Titelbaum p. 61 and p. 64)
  - Residue of Hypothetico-deductivism:
    - If A entails B, then $p(A|B) \geq p(A)$
  - $p(A|B) > p(A)$ iff $p(B|A) > p(B)$

    * $p(A|B) > p(A)$ iff $p(B|A) > p(B|\neg A)$

- Some useful definitions:
    * A and B are **probabilistically independent** iff $p(A|B) = p(A)$, iff $p(A|B) = p(A|\neg B)$, iff $p(B|A) = p(B)$, iff $p(B|A) = p(B|\neg A)$, iff $p(A \wedge B) = p(A) \times p(B)$
    * C **screens off** A from B iff
        · $p(A|B) \neq p(A)$
        · $p(A|B \wedge C) = p(A|C)$
        · $p(A|B \wedge \neg C) = p(A|\neg C)$