A PRIMER ON BAYESIAN EPISTEMOLOGY

NILANJAN DAS

UCL

1. Motivation for Probabilistic Epistemology

- Traditional epistemology assumes a particular model of belief:
  - Belief is **binary** ("on/off"): you either believe p or you don't.
  - Rational beliefs obey the constraints of **classical logic**.

For instance, one cannot rationally believe both p and $\neg p$. Nor can one rationally believe p, $p \to q$ and $\neg q$.

- The traditional model faces problems:
  - Belief and disbelief are too coarse-grained. We need to model levels of uncertainty too.
  - Classical logical constraints are too strong.

    Preface Paradox. I rationally believe, of each sentence in my history book, that it is true. If my beliefs should obey classical logic, I should also believe that the conjunction of all the sentences is true. But I know that books almost always contain errors; so I should believe that the conjunction of all the sentences is false! Something has gone wrong.

- The probabilistic model of belief overcomes these problems:
  - Belief is **graded** ("comes in degrees"): you have a particular degree of belief, or "credence," in p.
  - Rational credences obey the constraints of **probability theory**, rather than classical logic.

    Preface Paradox resolved. You can have high credence, of each sentence in your history book, that it is true. But your credence that all the sentences are true may be quite low.

This is not to say that logic plays *no* role in constraining rational credences. More on this in §2.1.

2. Bayesianism

- Bayesianism is one of the more popular approaches to probabilistic epistemology. It is characterized by two normative claims: Probabilism and Conditionalization.

2.1 Probabilism

- To state the first claim, we'll need to introduce some concepts in the mathematical theory of probability. The most fundamen-

tal is that of a **probability space**. A probability space is a triple $\langle \Omega, \mathcal{F}, P \rangle$, where:

- Ω is a set of possible worlds.
- $\mathcal{F}$ is a set of subsets of Ω. More specifically, $\mathcal{F}$ is a **σ-algebra** on Ω.

  Definition: $\mathcal{F}$ is a σ-algebra on Ω iff
  1. only subsets of Ω are members of $\mathcal{F}$;
  2. Ω is a member of $\mathcal{F}$ (i.e. $\Omega \in \mathcal{F}$);
  3. for any subset X of Ω, if $X \in \mathcal{F}$, then $\overline{X} \in \mathcal{F}$; and,
  4. for any subsets X and Y of Ω, if $X \in \mathcal{F}$ and $Y \in \mathcal{F}$, then $X \cup Y \in \mathcal{F}$.

There is controversy over how to interpret Ω. Some wish to interpret it as the set of all logically possible worlds. Others the set of all metaphysically possible worlds. Yet others the set of all worlds compatible with the agent's evidence.

- P is a **probability function** taking members of $\mathcal{F}$ to real numbers in the interval [0, 1].

  Definition: P is a probability function from $\mathcal{F}$ to [0,1] iff
  1. for any set $X \in \mathcal{F}$, $P(X) \geq 0$;
  2. $P(\Omega) = 1$; and,
  3. for any disjoint sets $X, Y \in \mathcal{F}$, $P(X \cup Y) = P(X) + P(Y)$.

- Conditions 1-3 in the definition of a probability function state the standard axioms of probability theory, due to Kolmogorov (1950). The first is called **Non-negativity**, the second **Normalization**, and the third **Finite Additivity**.

Some Bayesians wish to replace Finite Additivity with **Countable Additivity**, which says that for any countable sequence of pairwise disjoint sets $X_1, X_2, ..., X_n \in \mathcal{F}$,

$$P(X_1 \cup ... \cup X_n) = \sum_{i=1}^{\infty} P(X_i).$$

- We can now state the first Bayesian claim:

  Probabilism
  The credences of a rational agent form a probability function.

Credences that form a probability function are (probabilistically) **coherent**.

- Probabilism has various implications. For example,
  - Suppose I'm 80% confident that it will rain tomorrow. Then, according to Probabilism, I better be 20% confident that it will not rain tomorrow, on pain of irrationality.

This follows from Non-negativity and Additivity.

  - Suppose I know that the train arrived either at noon, 1, or 2, and I'm equally confident in each possibility. Then, according to Probabilism, I better be 2/3 confident that it arrived at either noon or 1, on pain of irrationality.

This follows from Additivity.

  - Let p be some complex logical truth. According to Probabilism, I'm rationally required to assign maximal credence 1 to p.

This follows from Normalization. It is an example of one way in which logic constrains rational constraints.

2.2 Conditionalization

- The second Bayesian claim requires the further concept of **conditional probability**. The conditional probability of X, conditional on Y, is written '$P(X|Y)$' and defined as follows:

  Ratio
  $P(X|Y) = \frac{P(X \cap Y)}{P(Y)}$, where $P(Y) > 0$.

- Ratio entails three useful theorems:

  Bayes' Theorem
  $P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)}$, where $P(X), P(Y) > 0$.

  The Law of Total Probability
  $P(X) = \sum_n P(X|Y_n)P(Y_n)$, where the Y_ns partition Ω.

  The Multiplication Rule
  $P(X \cap Y) = P(X)P(Y)$, where X and Y are independent.

Definition: a set Δ of subsets of Ω is a **partition** of Ω iff (i) the members of Δ are non-empty and pairwise disjoint, and (ii) the union of all the members of Δ is Ω itself. The Law of Total Probability follows straightforwardly from this definition and Ratio.

Definition: X and Y are (probabilistically) **independent** iff (i) $P(X|Y) = P(X)$ and (ii) $P(Y|X) = P(Y)$. The Multiplication Rule follows straightforwardly from this definition and Ratio.

- Whereas Probabilism places a **synchronic** constraint on how one's credences ought to be distributed *at a particular time*, the second Bayesian claim places a **diachronic** constraint on how one's credences ought to change *across time*, in light of new information.

- The claim is that, upon learning Y, one's **posterior** (i.e. "new") unconditional credence in X should equal one's **prior** (i.e. "old") conditional credence in X, conditional on Y. Letting P represent one's prior credences and P' represent one's posterior credences, the claim requires that one's credences satisfy the following equation, on pain of irrationality:

  Conditionalization
  $P'(X) = P(X|Y)$.

- Conditionalization has various implications, some of which are exhibited in the Exercises.

3. Arguments for Bayesianism

Many of the arguments in this section are discussed in Earman (1992) and Howson and Urbach (1993).

- *Dutch book arguments*
  A **Dutch book** is a series of bets which, when taken together, guarantee a loss. According to the synchronic Dutch book argument

Traditional DBAs assume a tight psychological or constitutive connection between an agent's betting behaviour and her credences.

DBAs were originally proposed by Ramsey (1926) and de Finetti (1937). Teller (1973) and Lewis (1999) were the first to offer diachronic DBAs for Conditionalization. For an introductory discussion of DBAs, see Hajek (2008).

(DBA), one's credences recommend accepting a Dutch Book iff they are incoherent. A diachronic DBA is used to defend Conditionalization in the same manner.

- *Representation theorem arguments*
  Representation theorem arguments (RTAs) take the following form: If one is rational, then one's preferences obey certain constraints, like transitivity. One's preferences obey these constraints iff one can be represented as having a set of utilities and coherent credences. If one can be represented thus, then one *has* coherent credences. Therefore, if one is rational, one has coherent credences.

To be **transitive**, one's preferences must be such that, for any gambles x, y and z, if one prefers x to y and y to z, then one prefers x to z as well.

The basic idea of the RTAs is again due to Ramsey. The traditional RTA assumes a tight metaphysical connection between the credences one can be represented as having and the credences that one actually has.

- *Accuracy- or utility-based arguments*
  Credences are more accurate as they get closer to the truth. For instance, if p is true, then a credence of $2/3$ in p is more accurate than a credence of $1/2$. Utility functions that measure the overall accuracy of a credal state are called **scoring rules**. Accuracy- or utility-based arguments proceed by showing either that (a) incoherent credences are dominated by coherent credences, accuracy-wise, or (b) rational agents can be understood as "maximizing epistemic utility" iff their credences are coherent. A similar argument available for Conditionalization.

For accuracy- or utility- based arguments for Probabilism, see Joyce (1998, 2009 and Leitgeb and Pettigrew (2010a, 2010b), and Pettigrew (2016). For corresponding arguments for Conditionalization, see Greaves and Wallace (2006) and Leitgeb and Pettigrew (2010a, 2010b). For discussion for arguments for Conditionalization more generally, see Pettigrew (2020).

4. Further Constraints

- The constraints placed by Bayesianism on rational credence are rather minimal.

  The Problem of the Priors. Conditionalization requires one to align one's current credences with one's prior credences in a particular way. But it places no constraint on what those prior credences must be. Probabilism places *some* constraint. For instance, if my prior credence in Y is $2/3$, then my prior credence in $\neg Y$ must be $1/3$. But one could respect Probabilism just as well by assigning credence $1/3$ to Y and $2/3$ to $\neg Y$. Of course, if Y is a tautology or contradiction, then Probabilism will give a determinate answer as to what my prior credence in Y must be. But this is cold comfort, since the vast majority of propositions are contingent. So the question is this: Apart from the standard Bayesian constraints, are there any further rational constraints on prior credence?

- The problem of the priors divides Bayesians into two camps.
  - According to **Subjective Bayesianism**, the answer is no: there are no further *rational* constraints on prior credence beyond Probabilism and Conditionalization. Priors are determined by non-rational factors (evolution, socialization, etc.).

Objection to Subjective Bayesianism: Consider two agents who have been exposed to exactly the same evidence throughout their cognitive lives. Suppose each agent has responded rationally to her evidence at every step along the way. Could these agents end up with vastly different credences? Subjective Bayesianism *appears* to say yes, but the intuitions of Objective Bayesians say no.
Subjective Bayesian reply: There are "convergence" theorems (due to Gaifman and Snir (1982)) which show that even the credences of agents with vastly different priors eventually converge, given a sufficiently long series of observations, via conditionalization on the same evidence from initially coherent priors. Some Objective Bayesians question the significance of such theorems. For discussion, see Earman (1992).

  - According to **Objective Bayesianism**, the answer is yes: there are further rational constraints on credence.

- What are the further constraints recommended by Objective Bayesians? They fall roughly into two categories:
  - *Deference Constraints*
    * Constraints in this category require one to defer, in one's credal assignments, to an "expert function". Such constraints typically assume the same general form,

      $$P(X|Exp(X) = \alpha) = \alpha,$$

      differing only in how they interpret the expert function Exp.

    * The **Principal Principle**, for instance, interprets $Exp(X)$ as the *objective chance* of X, thereby requiring one to match one's credences to the objective chances.

For classic discussions of the Principal Principle, see Lewis (1980, 1994).

    * The **Reflection Principle**, on the other hand, interprets $Exp(X)$ as one's own *future rational credence* in X, thereby requiring one to match one's current credences to those of one's future self.

Van Fraassen (1984) introduces the Reflection Principle. For an attempt to fix some of the problems that arise for this principle, see Briggs (2009).

  - *Equivocation Constraints*
    * Constraints in this category require one to adopt the most "ambivalent" or "uninformative" distribution of credences over those propositions not ruled out by one's evidence. One particularly widely-discussed constraint in this category is:

      The Principle of Indifference
      If one has no more reason to suppose that X is true than one has to suppose that Y is true, and vice versa, then X and Y deserve equal credence.

This formulation of the Principle of Indifference is borrowed from White (2010). For objections to the Principle of Indifference, see Keynes (1929) and van Fraassen (1989).

5. Arguments against Conditionalization

A. Fission

- *Fission*. Conditionalization assumes that the credences of a person at any time should cohere with the credences of the *same person* at an earlier time. So, it cannot handle cases of fission.

  > [C]onsider a fission case involving double teletransportation. One person (call her 'Pre') enters the teletransporter in New York. Her body is scanned, and at the moment her body is vaporized, two different molecule-for-molecule duplicates of her are created, one in Los Angeles and the other in San Francisco. Call them 'Lefty' and

> 'Righty', respectively. Lefty and Righty are qualitatively just like Pre in all physical and mental respects. (Hedden 2015, p. 456)

Hedden's claim: in this case, we can know what the credences of Lefty and Righty should be without deciding which of them is identical to Pre.

- *Response*. Why can't we just restrict Conditionalization to non-branching cases? That doesn't take away its explanatory power.

B. Internalism

- *Internalism*. What an agent is rational to be believe at a time depends solely on her mental states at that time.
- *An Example*.

> *Two Roads to Shangri La.* There are two paths to Shangri La, the Path by the Mountains, and the Path by the Sea. A fair coin will be tossed by the guardians to determine which path you will take: if heads you go by the Mountains, if tails you go by the Sea. If you go by the Mountains, nothing strange will happen: while traveling you will see the glorious Mountains, and even after you enter Shangri La, you will forever retain your memories of that Magnificent Journey. If you go by the Sea, you will revel in the Beauty of the Misty Ocean. But, just as you enter Shangri La, your memory of this Beauteous Journey will be erased and be replaced by [an apparent] memory of the Journey by the Mountains. (Hedden 2015, p. 456)

Here, even if you take the path by the mountains, your credence that the coin landed heads should 0.5. This is in tension with Conditionalization.

- *Response*. First of all, it's not immediately obvious that you don't remember that you came by the mountains here. Second, this is not a problem for *Conservatism*, but only for Conditionalization. Finally, it also seems possible to treat this as a case of *rational* evidence loss.

Hedden says: "Note the internalist intuition here: that what you ought to believe depends on what your evidence is, and your evidence supervenes on your present mental states. Your evidence includes your present apparent memory of the Mountains, but not your past visual experiences of the Mountains (which entail that you went by the Mountains)."

We'll look at an argument given by Lasonen-Aarnio (2010) that calls into question the possibility of knowledge defeat.

A defender of *Conservatism* could say that you do get some counterevidence here, namely that you are now entering Shangri-La, and therefore that there is 0.5 chance your memories are unreliable.

C. Epistemic Impartiality

- *The Principle of Epistemic Impartiality*. The considerations determining which beliefs it would be epistemically rational for an agent to adopt do not give special status to any of the agent's present opinions on the basis of their belonging to the agent (Christensen 2000, pp. 363-4)
- *An Example*.

> *The Belief Downloader*. Suppose that I have a serious lay interest in fish, and have a fairly extensive body of beliefs about them. At a party, I meet a professional ichthyologist. Although I of course believe that she shares the vast majority of my beliefs about fish, I know that she can probably set me straight about some ichthyological matters. Especially, she might be conditional credences that are more accurate than mine. However, I don't want to trouble her by asking a lot of work-related questions. Fortunately, I have a belief-downloader, which works as follows: If I turn it on, it scans both of our brains, until it finds some ichthyological proposition about which we disagree. It then replaces my belief with that of the ichthyologist, and turns itself off. (Ibid., p. 360)

Here, it seems that I have very strong epistemic reason to use the belief downloader.

- *The Problem*. Prima facie, the *Principle of Epistemic Impartiality* seems incompatible with Conditionalization, since it requires us to hold fixed some of our conditional credences as we are updating on new evidence.

6. Arguments against Reflection

- *Counterexamples.* Reflection is also subject to numerous counterexamples.

> *Future Irrationality*. The drug LSQ makes people believe to degree .99 that they can fly by flapping their arms. At t0, you become certain that you will take LSQ before t1. You deduce that at t1, you will place credence .99 in the proposition (call it F) that you can fly.
>
> *Memory Loss*. At t0, you are eating a dinner of spaghetti and meatballs. You expect to forget this by t1, but you'll remember that t0 was your dinner time. You'll also remember that you eat spaghetti for dinner 10 percent of the time.

For other such counterexamples, see Briggs (2009).

- *Problems for Modified Reflection Principles*. We may qualify the principle as follows:

> **Modified Reflection**. For any two times t_1 and t_2, if p_1 is the agent's rational credence function at t_1 and $[p_2(H) = r]$ is proposition that she will rationally assign a credence of r to a proposition H at t_2 without losing any evidence, then
>
> $p_1(H|[p_2(H) = r]) = r$ (provided $p_1([p_2(H) = r]) > 0$).

- *Two Problems*. Hedden mentions two problems for thi principle.
  - These reflection principles are insufficiently general in virtue of being future-directed.
  - These reflection is insufficiently general in virtue of being about the beliefs you believe *you* will later have.

The basic idea: these principles are arbitrary insofar as they attach special weight to beliefs you'll have in the future.

7. Time-Slice Rationality

- *Time-Slice Epistemology*. Hedden understands time-slice epistemology as the conjunction of two theses:

  > *Synchronicity*. What attitudes you ought to have at a time does not directly depend on what attitudes you have at other times.
  > *Impartiality.* In determining what attitudes you ought to have, your beliefs about what attitudes you have at other times play the same role as your beliefs about what attitudes other people have.

  Synchronicity conflicts with the conservation principles, while *Impartiality* conflicts with the reflection principles.

- *Replacements*. Given the problems for these principles, he suggests we replace Conditionalization and Reflection with:

  1. **Synchronic Conditionalization**. Let π be the uniquely rational ur-prior. If at time t you have total evidence E, your credence at t in each proposition H should equal $\pi(H|E)$.
  2. **Expert Deference**. If p is your credence function and $[p^A_{exp}(H) = r]$ is the proposition that A is an expert with credence r in H, then it is a requirement of rationality that, for all H,
     $p(H|[p^A_{exp}(H) = r]) = r$ (provided $p([p^A_{exp}(H) = r]) > 0$).

  Both these principles have problems.

- *The Problem for Synchronic Conditionalization*. Synchronic Conditionalization is incompatible with *permissivisim*, i.e., the view that relative to a specific body of total evidence, there needn't be a unique attitude that is rational to adopt towards every proposition.

  Rosen ((2001), p. 71): "It should be obvious that reasonable people can disagree, even when confronted with the same body of evidence. When a jury or a court is divided in a difficult case, the mere fact of disagreement does not mean that someone is being unreasonable."

  – Permissivism can involve rejecting two kinds of uniqueness theses.

    > *Intrapersonal Uniqueness*. For any agent S, any proposition H, and any time time t, there is a unique attitude that S is permitted to adopt towards H at t.
    > *Interpersonal Uniqueness*. For any two agents S, S^*, any proposition H, and any time time t, if S and S^* have the same total evidence at t, then there is a unique attitude that both S and S^* are permitted to adopt towards H at t.

    Permissivists themselves think that *Intrapersonal Uniqueness* is much more plausible than *Interpersonal Uniqueness*.

  – *Intrapersonal Uniqueness* seems plausible in light of cases like this:

> I don't want to believe that Jo is guilty. I have just received some evidence that strongly supports the claim that Jo is guilty relative to my current epistemic standard, so I should believe it. But I am aware that if I switch to another epistemic standard, I won't have to believe it. Should I switch?
>
> It seems to me that I shouldn't; this seems like a bad way of avoiding relevant evidence.

But it's extremely hard to explain this intuition without appealing to something like the reflection principles.

- *The Problem for Expert Deference*. As Hedden notices, the Principle of Expert Deference cannot hold, at least in some cases where an agent assigns a non-zero probabity to the possibility that two experts disagree (see Gallow (2018)). But the way he resolves this problem appeals to *Synchronic Conditionalization*.

References

Briggs, R. (2009). Distorted reflection. *Philosophical Review*, 118(1):59–85.

Christensen, D. (2000). Diachronic coherence versus epistemic impartiality. *Philosophical Review*, 109(3):349–371.

de Finetti, B. (1937). Foresight: Its logical laws, its subjective sources. In Henry E. Kyburg, J. and Smokler, H. E., editors, *Studies in Subjective Probabilitiy*, pages 94–158. London, New York, Sydney: John Wiley and Sons.

Earman, J. (1992). *Bayes or Bust?* Cambridge, MA: MIT Press.

Fraassen, B. C. V. (1984). Belief and the will. *Journal of Philosophy*, 81(5):235–256.

Fraassen, B. C. V. (1989). *Laws and Symmetry*. Oxford: Oxford University Press.

Gaifman, H. and Snir, M. (1982). Probabilities over rich languages, testing and randomness. *Journal of Symbolic Logic*, 47(3):495–548.

Gallow, J. (2018). No one can serve two epistemic masters. *Philosophical Studies*, 175(10):2389–2398.

Greaves, H. and Wallace, D. (2006). Justifying conditionalization: Conditionalization maximizes expected epistemic utility. *Mind*, 115(459):607–632.

Hajek, A. (2008). Dutch book arguments. In *The Oxford Handbook of Rational and Social Choice*. Oxford University Press.

Hedden, B. (2015). Time-slice rationality. *Mind*, 124(494):449–491.

Howson, C. and Urbach, P. (1993). *Scientific Reasoning: The Bayesian Approach*. Chicago and La Salle, Illinois: Open Court. Third Edition.

Joyce, J. (2009). Accuracy and coherence: Prospects for an alethic epistemology of partial belief. In *Degrees of Belief*, volume 342, pages 263–297. Synthese.

Joyce, J. M. (1998). A nonpragmatic vindication of probabilism. *Philosophy of Science*, pages 575–603.

Keynes, J. M. (1929). *The Treatise on Probability*. London: Macmillan and Co., Limited.

Kolmogorov, A. (1950). *Foundations of the theory of probability*. New York: Chelsea Publishing Company.

Lasonen-Aarnio, M. (2010). Unreasonable knowledge. *Philosophical Perspectives*, 24(1):1–21.

Leitgeb, H. and Pettigrew, R. (2010a). An objective justification of bayesianism i: Measuring inaccuracy. *Philosophy of Science*, 77(2):201–235.

Leitgeb, H. and Pettigrew, R. (2010b). An objective justification of bayesianism ii: The consequences of minimizing inaccuracy. *Philosophy of Science*, 77(2):236–272.

Lewis, D. (1980). A subjectivist's guide to objective chance. In *Studies in Inductive Logic and Probability*, volume 2, pages 263–293. University of California Press.

Lewis, D. (1994). Humean supervenience debugged. *Mind*, 103(412):473–490.

Lewis, D. (1999). Why conditionalize? In *Papers on Metaphysics and Epistemology*, pages 403–7. Cambridge: Cambridge University Press.

Pettigrew, R. (2016). *Accuracy and the Laws of Credence*. Oxford University Press UK.

Pettigrew, R. (2020). What is conditionalization, and why should we do it? *Philosophical Studies*, 177(11):3427–3463.

Ramsey, F. P. (1990/1926). Truth and probability. In Mellor, D. H., editor, *Philosophical Papers*, pages 52–94. Cambridge: Cambridge University Press.

Rosen, G. (2001). Nominalism, naturalism, epistemic relativism. *Noûs*, 35(s15):69–91.

Teller, P. (1973). Conditionalization and observation. *Synthese*, 26(2):218–258.

White, R. (2010). Evidential symmetry and mushy credence. In Gendler, T. and Hawthorne, J., editors, *Oxford Studies in Epistemology*. Oxford: Oxford University Press.