Probability

Outline:

1. Axioms of the Probability Calculus
2. Key Principles
3. Bayes' Theorem: computations and applications
4. Interpretations of probability:
    - objective
    - subjective
5. Bayes' Theorem: philosophical issues

1. The Probability Calculus

- Notation: &, $\vee$, ~, $\leftrightarrow$,
- *P(p)*: the probability of *p*
- *Equivalent formulations*: statements or *outcomes*

Axioms of unconditional and conditional probability

(1) (*Positive*): For any p [and q], $0 \leq P(p)$ [and $0 \leq P(p / q)$].

(2) (*Tautology*): If t is a tautology, then $P(t) = 1$ [and $P(t / q) = 1$].

(3) (*Additivity*): If p and q are mutually exclusive (can't both be true), then:

- $P(p \text{ or } q) = P(p) + P(q)$;
  [$P(p \text{ or } q / r) = P(p / r) + P(q / r)$, for any statement r]

(4) (*Conjunction*): $P(p \,\&\, q) = P(q)\, P(p / q)$
$= P(p)\, P(q / p)$

Def: p and q are independent if $P(p / q) = P(p)$

Note: Conditional probability is generally either defined by (4) or taken to be primitive.

2. Key principles

Proposition 1*:* If p and q are mutually exclusive and $P(q) \neq 0$, then $P(p \mathbin{/} q) = 0$.

Proposition 2: $P({\sim}p) = 1 - P(p)$

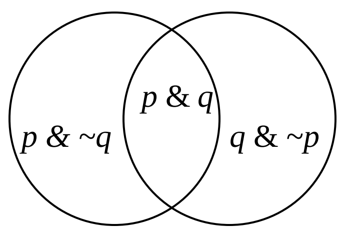
> *Corollary*: If f is a logical contradiction, then $P(f) = 0$ (since ${\sim}f$ is a tautology, so that $P({\sim}f) = 1$).

Proposition 3: If p and q are equivalent, then $P(p) = P(q)$.

Proposition 4: If $p \quad q$, then $P(p) \leq P(q)$.

> *Corollary*: $0 \leq P(p) \leq 1$ for any p (since $p \quad t$ and $P(t) = 1$).

Proposition 5: $P(p \vee q) = P(p) + P(q) - P(p \mathbin{\&} q)$



Proposition 6: If $P(q) \neq 0$, then

$$P(p \mathbin{/} q) = P(p \mathbin{\&} q) \mathbin{/} P(q).$$

Proposition 7: If p and q are independent, then $P(p \mathbin{\&} q) = P(p)\,P(q)$.

Question: Why can't we define conditional probability as the probability of a conditional:

$$P(p \mathbin{/} q) = P(q \supset p)$$

Proposition 8: Suppose $p_1, \ldots, p_n$ are mutually exclusive, i.e., p_i $\sim p_j$ for all i, j.

Then $P(p_1 \vee \ldots \vee p_n) = P(p_1) + \ldots + P(p_n)$.

Proposition 9 (Theorem of Total Probability): Suppose $q_1, \ldots, q_n$ are mutually exclusive and exhaustive. Then for any sentence p,

$$P(p) = P(p \,\&\, q_1) + \ldots + P(p \,\&\, q_n)$$

Corollary: Under the same assumptions,

$P(p) = P(p / q_1)P(q_1) + \ldots + P(p / q_n)P(q_n)$ (just apply the conjunction axiom).

Countable additivity

Consider an infinite disjunction: $H_1 \vee H_2 \vee \ldots$ where H_i means you get your first heads on toss i.

Then $H \leftrightarrow H_1 \vee H_2 \vee \ldots$ (where H is 'coin eventually comes up heads').

If coin is fair, $P(H) = 1$, and $P(H_i) = 1/2^i$. So we have

$$P(H) = \Sigma P(H_i) = P(H_1) + P(H_2) + \ldots$$

In general, P is *countably additive* if $P(p_1 \vee p_2 \ldots) = \Sigma P(p_i)$.

(De Finetti's lottery: a counterexample to countable additivity as a requirement for subjective probability)

3. Bayes' Theorem

Version A:

$$P(p/q) = \frac{P(p) \cdot P(q/p)}{P(q)}$$

Version B:

$$P(p/q) = \frac{P(p) \cdot P(q/p)}{P(p) \cdot P(q/p) + P(\sim p) \cdot P(q/\sim p)}$$

Version C:

$$P(p_1/q) = \frac{P(p_1) \cdot P(q/p_1)}{P(p_1) \cdot P(q/p_1) + P(p_2) \cdot P(q/p_2) + \ldots + P(p_n) \cdot P(q/p_n)}$$

Terminology

- The *posterior probability* of *p* relative to *q*, $P(p/q)$, appears on the left.
- The *prior probability* of *p*, $P(p)$.

  - Connection to plausibility
  - *Open-mindedness* about *p*: $0 < P(p) < 1$:
- The *likelihood* of the evidence *q* given *p*, $P(q/p)$

  - If $P(q/\mathrm{p}) = 0$, no evidence will raise probability of *p*
  - If $P(q/p) = 1$, *q* provides strong evidence for *p* (unless *E* is totally unsurprising even if ~*p*)
- The *expectedness* of the evidence *q*, $P(q)$.

  - Suprising evidence has greater power to raise probability of *p*

c) Prior probabilities

Objection: where do prior probabilities come from?

- Symmetry of a chance set-up: lotteries, dice, card games
- Statistical frequencies

But what if neither of these applies?

Example: Some coins are biased to 75% tails; others are fair. No idea of the frequency of biased coins.

Convergence Argument: Bayesians say you should use your best guess as the prior probability. Widely divergent prior probabilities will "wash out" or converge, given a large amount of data.