A Brief Primer on Probability

Ryan Doody

The Probability Axioms

Consider a language $\mathcal{L}$ of propositions, closed under truth-functional connectives. Define a real-valued function c over $\mathcal{L}$ to represent the *credence* (or *degree of belief*) that an agent assigns to the propositions in $\mathcal{L}$. According to **Probabilism**, rationality requires c to be a probability function—and, thus, to obey the following axioms:

Truth-functional Connectives

$\neg p$	...	It's not the case that p.
$p \wedge q$	...	p and q.
$p \vee q$	...	p or q.
$p \supset q$	...	If p, then q.
$p \equiv q$	...	p if and only if q.

For any $X \in \mathcal{L}$, $c(X) \in \mathbb{R}$.

The Probability Axioms

NON-NEGATIVITY. Every $X \in \mathcal{L}$ is assigned a non-negative number.

$$c(X) \geq 0 \tag{1}$$

NORMALITY. Every tautology $\top \in \mathcal{L}$ is assigned 1.

$$c(\top) = 1 \tag{2}$$

FINITE ADDITIVITY. For any mutually exclusive $X, Y \in \mathcal{L}$, the number assigned to their disjunction equals the sum of the numbers assigned to them.

$$\text{If } (X \wedge Y) \vDash \bot, \text{ then } c(X \vee Y) = c(X) + c(Y) \tag{3}$$

Interesting Fact: These are known as the Kolmogorov Axioms, named after Andrey Kolmogorov, the Soviet mathematician who introduced them in 1933.

A tautology (which we'll abbreviate as '$\top$') is any proposition that is guaranteed to be true as a matter of logic: e.g., $(p \vee \neg p)$.

Two propositions are *mutually exclusive* if it's impossible that they *both* be true. Each one refutes the other. And so, their conjunction entails a contradiction (which we'll abbreviate as '$\bot$').

Here are three interesting and useful facts.

The Negation Rule: For any $X \in \mathcal{L}$, $c(\neg X) = 1 - c(X)$.

The Overlap Rule: The probability of a disjunction equals the sum of the probabilities of its disjuncts minus the probability of its disjuncts' overlap.

$$c(X \vee Y) = c(X) + c(Y) - c(X \wedge Y)$$

*The Logical Consequence Rule:** If $X \vDash Y$, then $c(X) \leq c(Y)$.

[*]*The Conjunction Fallacy.* In a famous study, Tversky and Kahneman (1983) presented subjects with the following story:

> Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.

The subjects were then asked to rank the probabilities of the following propositions:

- Linda is active in the feminist movement.
- Linda is a bank teller.
- Linda is a bank teller and is active in the feminist movement.

A large majority of the subjects ranked the third option as more probable than the second!

Conditional Probability

In addition to the three axioms above, we introduce the notion of *conditional probability*.

THE RATIO FORMULA: For any $X, Y \in \mathcal{L}$ with $c(Y) > 0$,

$$c(X \mid Y) = \frac{c(X \wedge Y)}{c(Y)} \tag{4}$$

This is the probability that X is the case *conditional* on Y being the case. Given the notion of conditional credence, here are two more useful facts.

$c(X \mid Y)$, your credence in X *given* Y, is not your current actual opinion about X—rather, it's your assessment of X on the supposition that Y is true.

The Law of Total Probability: For any $X, Y_1, Y_2, \ldots, Y_n \in \mathcal{L}$, where $Y_1, Y_2, \ldots, Y_n$ form a partition (i.e., are mutually exclusive and jointly exhaustive),

$$c(X) = c(X \mid Y_1) \cdot c(Y_1) + c(X \mid Y_2) \cdot c(Y_2) + \ldots c(X \mid Y_n) \cdot c(Y_n)$$

The Multiplication Rule: For any $X, Y \in \mathcal{L}$, if $c(Y) > 0$, then

$$c(X \wedge Y) = c(X \mid Y) \cdot c(Y)$$

And here's a useful definition:

When $c(X \mid Y) > c(X)$, we say that Y is *positively relevant* to X—X and Y are taken to be positively correlated.

Independence:* X and Y are probabilistically independent just in case $c(X \mid Y) = c(X)$.

Updating by Conditionalization

How should your degrees of belief evolve over time? Let c_t be your credences at time t, and c_{t^+} be your credences at some later time t^+.

CONDITIONALIZATION. If $E \in \mathcal{L}$ is everything you learn between t and t^+, then, for any $X \in \mathcal{L}$, $c_{t^+}(X) = c_t(X \mid E)$.

The conditional function $c(\bullet \mid X)$ satisfies the Kolmogorov Axioms, and thus is itself a probability function. So, updating by Conditionalization won't lead you to violate Probabilism.

Bayes' Theorem

Calculating $c(X \mid E)$ can be made significantly easier by making use the following famous theorem.

BAYES' THEOREM. For any $X, E \in \mathcal{L}$, where $c(E) > 0$,

$$c(X \mid E) = \frac{c(E \mid X) \cdot c(X)}{c(E)} \tag{5}$$

Given *The Law of Total Probability*, the theorem can be rewritten as follows:

$$c(X \mid E) = \frac{c(E \mid X) \cdot c(X)}{c(E \mid X) \cdot c(X) + c(E \mid \neg X) \cdot c(\neg X)}$$

And, where $X, Y_1, Y_2, \ldots, Y_n \in \mathcal{L}$ form a partition,

$$c(X \mid E) = \frac{c(E \mid X) \cdot c(X)}{c(E \mid X) \cdot c(X) + c(E \mid Y_1) \cdot c(Y_1) + \cdots + c(E \mid Y_n) \cdot c(Y_n)}$$