Topic 12: Introduction to Probability

02-680: Essentials of Mathematics and Statistics

October 31, 2024

Probability is the mathematical language for quantifying uncertainty

1 Basics

The ***sample space*** (Ω), sometimes called the universe, is the set of all possible *mutually exclusive* outcomes of an event.

$$\Omega_{coin} = \{\texttt{Heads}, \texttt{Tails}\}$$

$$\Omega_{twocoin} = \Omega_{coin} \times \Omega_{coin} = \{\langle \texttt{Heads}, \texttt{Heads} \rangle, \langle \texttt{Heads}, \texttt{Tails} \rangle, \langle \texttt{Tails}, \texttt{Heads} \rangle, \langle \texttt{Tails}, \texttt{Tails} \rangle\}$$

A sample ***outcome***, or atomic element, $\omega \in \Omega$ is one of the possible things that can happen.

A set of events is a subset of the sample space with some condition. For instance if the event is that the first of two coin tosses is heads.

$$A_{firstheads} = \{\langle \texttt{Heads}, \texttt{Heads} \rangle, \langle \texttt{Heads}, \texttt{Tails} \rangle\} \subset \Omega_{twocoin}$$

We say the ***probability*** $P(A)$ of an event A is the fraction of all outcomes that A covers. So in the example above $P(A_{firstheads}) = \frac{1}{2}$.

An ***event space***, $\mathcal{A}$, is the set of all possible events. For discrete events (such as coin flips) this can be thought of typically as the power set of Ω. In continuous spaces it is typically thought of as the Borel field of Ω (details of this are beyond the scope of the class for now).

Another way of saying this: consider non-empty Ω and $\mathcal{A}$:

(1) $A \in \mathcal{A} \to \overline{A} \in \mathcal{A}$

(2) $A_1, A_2, ... \in \mathcal{A} \to \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$.

1.1 Three Axioms

Axiom 1: Positive Probability. For any event A, $p(A) \geq 0$.

That is, we cannot have a negative probability. In the previous example if $A_{threetails}$ is the event that you get 3 tails when a coin is tossed twice, $p(A_{threetails}) = 0$ but this can't be negative.

Axiom 2: Total Probability. For any event space Ω, $p(\Omega) = 1$. That is, the probability of something in the sample space happening is 1. So for the example above the probability of a single coin flip being either `Heads` or `Tails` is 1, (as defined) there are no other possible outcomes.

Axiom 3: Disjoint Event Space. For *disjoint* event spaces $A_1, A_2, ..., A_n$,

$$p(A_1 \cup A_2 \cup ... \cup A_n) = p(A_1) + p(A_2) + ... + p(A_n).$$

So in a single coin flip if A_{heads} and A_{tails} are the events of a single coin flip being heads and tails respectively, we can see the event spaces are disjoint (they don't share any outcomes) so

$$p(A_{heads} \cup A_{tails}) = p(A_{heads}) + p(A_{tails}) = \frac{1}{2} + \frac{1}{2} = 1.$$

As a counter example lets define

$$A_{lastheads} = \{\langle \texttt{Heads}, \texttt{Heads} \rangle, \langle \texttt{Tails}, \texttt{Heads} \rangle\} \subset \Omega_{twocoin}.$$

We can see that

$$p(A_{firstheads} \cup A_{lastheads}) \neq p(A_{firstheads}) + p(A_{lastheads})$$

because $A_{firstheads} \cap A_{lastheads} = \{\langle \texttt{Heads}, \texttt{Heads} \rangle\} \neq \emptyset$ thus they are not disjoint.

1.2 Example Consequences of the Three Axioms

1.2.1 $p(\emptyset) = 0$

$\Omega \cup \emptyset = \Omega$.

Since Ω is disjoint from $\emptyset$, we know $p(\Omega \cup \emptyset) = p(\Omega) + p(\emptyset)$.

Put these two together we see that $p(\Omega) = p(\Omega) + p(\emptyset)$.

Since $p(\Omega) = 1$, then $1 = 1 + p(\emptyset)$, and thus $p(\emptyset) = 0$.

1.2.2 $p(\overline{A}) = 1 - p(A)$

We know by definition that $\overline{A} = \Omega \setminus A$.

From Axiom 3, since A and $\overline{A}$ are disjoint, we know $p(A \cup \overline{A}) = p(A) + p(\overline{A})$.

But we also know that $A \cup \overline{A} = \Omega$.

So that means $p(A) + p(\overline{A}) = p(\Omega) = 1$. With some algebra we find the original statement.

1.2.3 $0 \leq p(A) \leq 1$.

From axiom 1: $0 \leq p(A)$.

We also know $p(\overline{A}) \geq 0$ from axiom 1. (or really $-1p(\overline{A}) \leq 0$)

From above we know $p(A) = 1 - p(\overline{A}) \leq 1 - 0 = 1$.

1.2.4 If $A \subseteq B$ then $p(A) \leq p(B)$.

$$
\begin{aligned}
p(B) &= p(A \cup (B \setminus A)) && B = A \cup (B \setminus A) \\
&= p(A) + p(B \setminus A) && A \text{ and } B \setminus A \text{ are disjoint (Axiom 3)} \\
&\geq p(A) && p(B \setminus A) \geq 0 \text{ (Axiom 1)}
\end{aligned}
$$

2 Perspectives

For a coin toss effect, there are two methods for explaining $p(\texttt{Heads}) = \frac{1}{2}$:

Frequentists. You flip a coin 100 times, you get `Heads` approximately 50 of those tosses. Frequentist are very objective in the explanation of things.

Bayesian. You *believe* you will get tails 50% of the times you flip the coin. Bayesian logic is subject to its interpretation.