Chapter 2

The Probability Function

Theoretical Probability Calculation

When you say you "know" that the toss of a fair coin has a 50% of being heads, you are reasoning that:

- the outcome is either a head or a tail
- outcomes are equally likely
- the two probabilities must add to 100%

Theoretical Probability Calculation

The basic idea is to combine

- some general properties that should be true of all probability situations (called **axioms**) with
- some additional assumptions about the situation at hand

To use this method, we need axioms.

Axioms of Probability

Definition 2.1 (Axioms of probability) Let S be a sample space for a random experiment. A probability assignment for S is a function P mapping events to the real line such that:

A1 $P(E) \geq 0$ for any event E

A2 $P(S) = 1$

A3 The probability of a *disjoint* union is the sum of probabilities.

- $P(E_1 \cup E_2) = P(E_1) + P(E_2)$, provided $E_1 \cap E_2 = \phi$
- $P(E_1 \cup E_2 \cup \cdots \cup E_k) = P(E_1) + P(E_2) + \cdots + P(E_k)$ provided $E_i \cap E_j = \phi$ whenever $i \neq j$.
- $P(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i)$ provided $E_i \cap E_j = \phi$ whenever $i \neq j$.

Axioms are not concerned with interpretations of probability. There are, however, two common interpretations given to probability calculations:

- $P(A)$ is the long run proportion of times event A occurs in repetitions of the experiment. (Frequentist)
- $P(A)$ measures an observer's strength of belief that A is true. (Bayesian)

Corollaries of the axioms

Theorem 2.1 If P is a probability function and A is any set in S then:

a. $P(A^c) = 1 - P(A)$ (Rule of Complements)

Corollaries of the axioms

Theorem 2.1 If P is a probability function and A is any set in S then:

b. $P(A) \leq 1$

Corollaries of the axioms

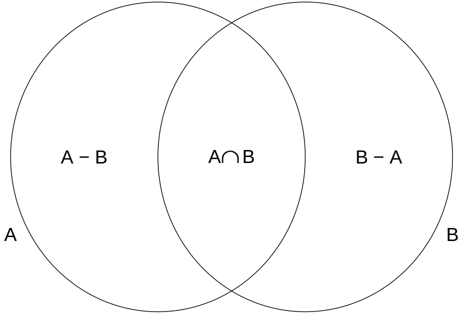
Theorem 2.1 If P is a probability function and A is any set in S then:

c. $P(\emptyset) = 0$

Corollaries of the axioms

Theorem 2.2 (Addition rule, two event version) If P is a probability function, and A and B are any sets in S

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$



Proof of Theorem 2.2

We begin by writing:

$$A \cup B = (A - B) \cup (A \cap B) \cup (B - A).$$

Now we can use axiom [A3] to write:

$$\begin{aligned} P(A \cup B) &= P\left((A - B) \cup (A \cap B) \cup (B - A)\right), \\ &= P(A - B) + P(A \cap B) + P(B - A), \end{aligned}$$

Next, we use axiom [A3] repeatedly to find $P(A - B)$ and $P(B - A)$:

- $P(A) = P\left((A - B) \cup (A \cap B)\right) = P(A - B) + P(A \cap B)$, so

$$P(A - B) = P(A) - P(A \cap B).$$

Proof of Theorem 2.2

We begin by writing:

$$A \cup B = (A - B) \cup (A \cap B) \cup (B - A).$$

Now we can use axiom [A3] to write:

$$\begin{aligned} P(A \cup B) &= P\left((A - B) \cup (A \cap B) \cup (B - A)\right), \\ &= P(A - B) + P(A \cap B) + P(B - A), \end{aligned}$$

Next, we use axiom [A3] repeatedly to find $P(A - B)$ and $P(B - A)$:

- $P(B) = P\left((B - A) \cup (A \cap B)\right) = P(B - A) + P(A \cap B)$, so

$$P(B - A) = P(B) - P(A \cap B).$$

Proof of Theorem 2.2

Combining these, we have:

$$
\begin{aligned}
P(A \cup B) &= P\left((A - B) \cup (A \cap B) \cup (B - A)\right), \\
&= P(A - B) + P(A \cap B) + P(B - A), \\
&= \underbrace{P(A) - P(A \cap B)}_{P(A-B)} + P(A \cap B) + \underbrace{P(B) - P(A \cap B)}_{P(B-A)}, \\
&= P(A) + P(B) - P(A \cap B).
\end{aligned}
$$

Example 2.1

Let A and B be two events defined on a sample space S such that $P(A) = 0.3$, $P(B) = 0.5$, and $P(A \cup B) = 0.7$. Find:

a. $P(A \cap B)$

Example 2.1

Let A and B be two events defined on a sample space S such that $P(A) = 0.3$, $P(B) = 0.5$, and $P(A \cup B) = 0.7$. Find:

b. $P(A^c \cup B^c)$

Example 2.1

Let A and B be two events defined on a sample space S such that $P(A) = 0.3$, $P(B) = 0.5$, and $P(A \cup B) = 0.7$. Find:

c. $P(A^c \cap B)$

- The addition rule

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

gives two useful identities:

  - Union bound (aka Boole's inequality)

$$P(A \cup B) \leq P(A) + P(B)$$

  - Bonferroni's inequality

$$P(A \cap B) \geq P(A) + P(B) - 1$$

Example 2.2

A new therapy for a disease will be approved if it is shown to be effective in two different studies. Suppose each study has a 95% probability of claiming that the treatment is effective. What can you say about the probability that the therapy is approved?