Chapter 2

The Probability Function

References: Pruim 2.1, Larsen & Marx 2.3

When a random experiment is performed, the outcome we observe is an element of the sample space. Some outcomes may be more likely than others. How do we quantify this? In this chapter we describe the assignment of a measure, called the probability measure, to an experiment's outcomes - and more generally to an event. Specifically, if E is an event in the sample space S, the symbol $P(E)$ will denote the probability of E and we refer to P as the probability function. It provides a measure of how likely the event E is to occur.

How do we come up with this function? The idea is to combine some general properties which should be true of all probability situations (called **probability axioms**) with some additional assumptions about the situation at hand, and then use deductive reasoning to reach conclusions.

For example, we would not find it surprising if we were told that the probability of a head when tossing a fair coin is 0.5. This is because we are probably reasoning something like the following:

- the outcome will be a head or a tail
- these two outcomes will be equally likely (since the coin is fair)
- the two probabilities must add to 1 because there is a 100% chance of getting one of these two outcomes.

This kind of reasoning is an example of the theoretical method for assigning probabilities. To use this method, we first need some self-evident axioms that any measure of likelihood must satisfy.

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. That is, for any events E_i

- $P(E_1 \cup E_2) = P(E_1) + P(E_2)$, 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 $i \neq j$.

Note that the axioms are not concerned with interpretations of the probability, only that it satisfy certain mathematical properties. 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. (probability as relative frequency)
- $P(A)$ measures an observer's strength of belief that A is true.

For either interpretation, axioms A1-A3 must hold. However, they lead to two schools of statistical inference: the Frequentist and Bayesian which we will learn more about in STAT 342.

One can derive many basic properties of P from just these three axioms. We don't include them in our list of axioms because we can derive them from the axioms. But once we have shown them, we can use them just like the axioms in subsequent calculations.

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)*

b. $P(A) \leq 1$

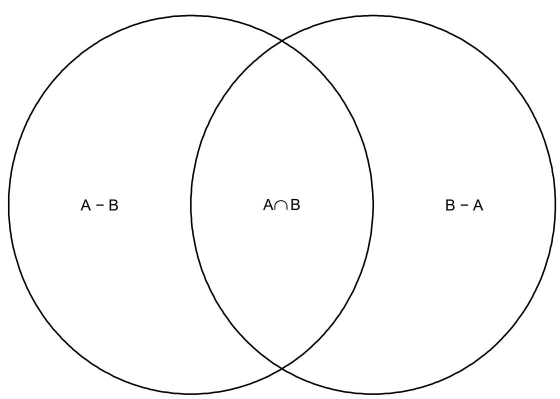
c. $P(\emptyset) = 0$

Theorem 2.1 states some very obvious consequences of the axioms. The result in the next theorem is a little less obvious.

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. We begin by writing $A \cup B$ as the union of three disjoint events as shown below.



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

Now we can use axiom [A3] repeatedly.

- $P(A) = P\left((A - B) \cup (A \cap B)\right) = P(A - B) + P(A \cap B)$, so $P(A - B) = P(A) - P(A \cap B)$.
- $P(B) = P\left((B - A) \cup (A \cap B)\right) = P(B - A) + P(A \cap B)$, so $P(B - A) = P(B) - P(A \cap B)$.
- Combining these, we have:

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

$\square$

. . . .

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. $P(A \cap B)$

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

c. $P(B - A)$

. . . .

Comment: The addition rule in Theorem 2.2 gives two useful identities:

- Union bound (aka Boole's inequality) $P(A \cup B) \leq P(A) + P(B)$
- Bonferroni's inequality $P(A \cap B) \geq P(A) + P(B) - 1$

Bonferroni's inequality is particularly helpful when it is difficult (or even impossible) to calculate the intersection probability, but some idea of the size of this probability is desired. Note that unless the probabilities of the individual events are large, the Bonferroni bound is a useless (but correct!) negative number.

. . . .

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?

. . . .

2.1 Practice Problems

1. An experiment has two possible outcomes with probabilities p and p^2. Find p.
2. True or false: $P(E) = P(E \cap F) + P(E \cap F^c)$.
3. Given that $P(A) = 0.4$, $P(A \cup B) = 0.7$, and $P(B^c) = 0.55$, find the probability $P(A \cap B)$.
4. Given that $P(A) = 0.4$, $P(A \cap B) = 0.1$ and $P((A \cup B)^c) = 0.2$. Find $P(B)$.

5. Let E and F be two events such that $P(E) = 0.35$, $P(F) = 0.15$ and $P(E \cap F) = 0.05$. Find $P(E^c \cap F^c)$.