

1.2: Probability Measures

Now we are ready to formally define probability.

Definition 1.2.1

A **probability measure** on the sample space S is a function, denoted P , from subsets of S to the real numbers \mathbb{R} , such that the following hold:

1. $P(S) = 1$
2. If A is any event in S , then $P(A) \geq 0$.
3. If events A_1 and A_2 are disjoint, then $P(A_1 \cup A_2) = P(A_1) + P(A_2)$.

More generally, if $A_1, A_2, \dots, A_n, \dots$ is a sequence of *pairwise disjoint* events, i.e., $A_i \cap A_j = \emptyset$, for every $i \neq j$, then

$$P(A_1 \cup A_2 \cup \dots \cup A_n \cup \dots) = P(A_1) + P(A_2) + \dots + P(A_n) + \dots$$

So essentially, we are defining probability to be an **operation** on the events of a sample space, which assigns numbers to events in such a way that the three properties stated in Definition 1.2.1 are satisfied.

Definition 1.2.1 is often referred to as the **axiomatic definition of probability**, where the three properties give the three **axioms** of probability. These three axioms are all we need to assume about the operation of probability in order for many other desirable properties of probability to hold, which we now state.

Properties of Probability Measures

Let S be a sample space with probability measure P . Also, let A and B be any events in S . Then the following hold.

1. $P(A^c) = 1 - P(A)$
2. $P(\emptyset) = 0$
3. If $A \subseteq B$, then $P(A) \leq P(B)$.
4. $P(A) \leq 1$
5. **Addition Law:** $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Exercise 1.2.1

Can you prove the five properties of probability measures stated above using only the three axioms of probability measures stated in Definition 1.2.1?

Answer

- (1) For the first property, note that by definition of the complement of an event A we have

$$A \cup A^c = S \quad \text{and} \quad A \cap A^c = \emptyset.$$

In other words, given any event A , we can represent the sample space S as a disjoint union of A with its complement. Thus, by the first and third axioms, we derive the first property:

$$\begin{aligned} 1 = P(S) &= P(A \cup A^c) = P(A) + P(A^c) \\ &\Rightarrow P(A^c) = 1 - P(A) \end{aligned}$$

- (2) For the second property, note that we can write $S = S \cup \emptyset$, and that this is a disjoint union, since anything intersected with the empty set will necessarily be empty. So, using the first and third axioms, we derive the second property:

$$\begin{aligned} 1 = P(S) &= P(S \cup \emptyset) = P(S) + P(\emptyset) = 1 + P(\emptyset) \\ &\Rightarrow P(\emptyset) = 0 \end{aligned}$$

(3) For the third property, note that we can write $B = A \cup (B \cap A^c)$, and that this is a disjoint union, since A and A^c are disjoint. By the third axiom, we have

$$P(B) = P(A \cup (B \cap A^c)) = P(A) + P(B \cap A^c). \quad (1.2.1)$$

By the second axiom, we know that $P(B \cap A^c) \geq 0$. Thus, if we remove it from the right-hand side of Equation 1.2.1, we are left with something smaller, which proves the third property:

$$P(B) = P(A) + P(B \cap A^c) \geq P(A) \Rightarrow P(B) \geq P(A)$$

(4) For the fourth property, we will use the third property that we just proved. By definition, any event A is a subset of the sample space S , i.e., $A \subseteq S$. Thus, by the third property and the first axiom, we derive the fourth property:

$$P(A) \leq P(S) = 1 \Rightarrow P(A) \leq 1$$

(5) For the fifth property, note that we can write the union of events A and B as the union of the following two disjoint events:

$$A \cup B = A \cup (A^c \cap B),$$

in other words, the union of A and B is given by the union of all the outcomes in A with all the outcomes in B that are *not* in A . Furthermore, note that event B can be written as the union the following two disjoint events:

$$B = (A \cap B) \cup (A^c \cap B),$$

in other words, B is written as the disjoint union of all the outcomes in B that are also in A with the outcomes in B that are *not* in A . We can use this expression for B to find an expression for $P(A^c \cap B)$ to substitute in the expression for $A \cup B$ in order to derive the fifth property:

$$P(B) = P(A \cap B) + P(A^c \cap B) \Rightarrow P(A^c \cap B) = P(B) - P(A \cap B)$$

$$P(A \cup B) = P(A) + P(A^c \cap B) \Rightarrow P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Note that the axiomatic definition (Definition 1.2.1) does not tell us how to *compute* probabilities. It simply defines a formal, mathematical behavior of probability. In other words, the axiomatic definition describes how probability should theoretically *behave* when applied to events. To compute probabilities, we use the properties stated above, as the next example demonstrates.

Example 1.2.1

Continuing in the context of Example 1.1.5, let's define a probability measure on S . Assuming that the coin we toss is *fair*, then the outcomes in S are **equally likely**, meaning that each outcome has the *same probability* of occurring. Since there are four outcomes, and we know that probability of the sample space must be 1 (first axiom of probability in Definition 1.2.1), it follows that the probability of each outcome is $\frac{1}{4} = 0.25$.

So, we can write

$$P(hh) = P(ht) = P(th) = P(tt) = 0.25.$$

The reader can verify that this defines a probability measure satisfying the three axioms.

With this probability measure on the outcomes we can now compute the probability of any event in S by simply *counting* the number of outcomes in the event. Thus, we find the probability of events A and B previously defined:

$$P(A) = P(\{hh, ht, th\}) = \frac{3}{4} = 0.75$$

$$P(B) = P(\{ht, th\}) = \frac{2}{4} = 0.50.$$

We consider the case of equally likely outcomes further in [Section 2.1](#).

There is another, more empirical, approach to defining probability, given by using *relative frequencies* and a version of the Law of Large Numbers.

Relative Frequency Approximation

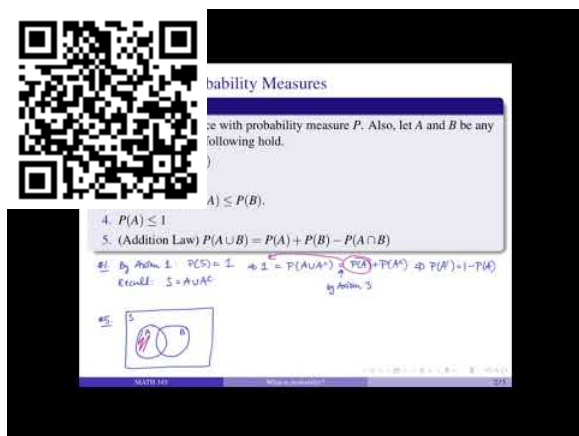
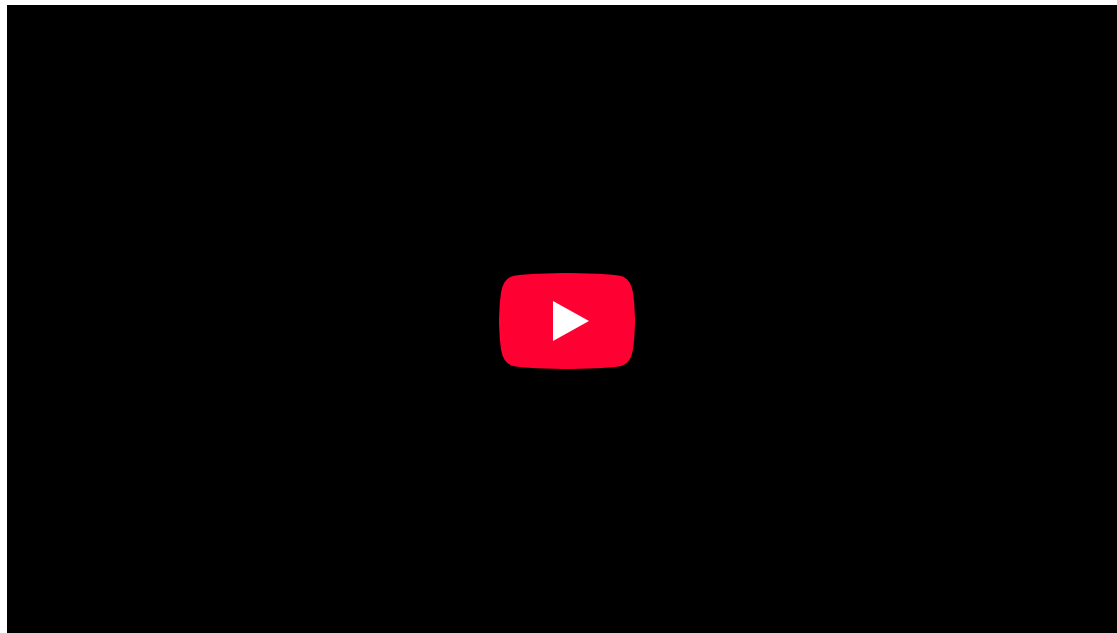
To *estimate* the probability of an event A , repeat the random experiment several times (each repetition is called a *trial*) and count the number of times A occurred, i.e., the number of times the resulting outcome is in A . Then, we approximate the probability of A using **relative frequency**:

$$P(A) \approx \frac{\text{number of times } A \text{ occurred}}{\text{number of trials}}.$$

Law of Large Numbers

As the number of trials increases, the relative frequency approximation approaches the theoretical value of $P(A)$.

This approach to defining probability is sometimes referred to as the **frequentist definition of probability**. Under this definition, probability represents a *long-run average*. The two approaches to defining probability are equivalent. It can be shown that using relative frequencies to define a probability measure satisfies the axiomatic definition.



https://youtu.be/Y_GAW22n1Ms

This page titled [1.2: Probability Measures](#) is shared under a [not declared](#) license and was authored, remixed, and/or curated by [Kristin Kuter](#).