CHAPTER 1

FOUNDATIONS OF PROBABILITY

INTRODUCTION

In this chapter, we develop the foundations of probability theory. The material presented in this chapter is fundamental. There is nothing mathematically complex or difficult - all that is required is simple algebra. Furthermore, there are only a few concepts that the reader needs to master before a whole new world of understanding is opened up in how to deal with uncertainty and randomness in engineering, science, and nature. By the end of this chapter, the reader should begin to appreciate the importance of probability and see how and when it appears in many different contexts and applications.

This chapter begins by looking at the notion of randomness and uncertainty by asking some probing questions about what events should be considered random, and what should be considered, or deterministic. This leads to the definition of a few simple terms: *events*, *sample spaces*, and *experiments*. These terms set the framework within which basic concepts of probability theory may be developed. Three simple and intuitive *axioms* are then introduced that are the foundation of probability theory.

1-1 RANDOMNESS AND UNCERTAINTY

Let us begin our journey into the world of probability with the following question: "What phenomena or events in nature should we consider to be random?" For example, should the outcome of the flip of a coin be taken to be a random event, with its outcome unknown until the coin is flipped and comes to rest on the table? Perhaps it should be, unless we are given precise initial conditions at the time the coin is released from the hand in order to compute the trajectory and orientation of the coin throughout its flight until it comes to final rest on the table. Since this information is rarely or never available, it is certainly easier, and more realistic, to

assume that the outcome of the flip of the coin is a random event, with an outcome that is equally likely to be either *Heads* or *Tails*.

As another example, should the time and location of the next earthquake be considered a random event? Perhaps the answer to this question should be "no," since an earthquake is the outcome of a complex set of interactions among many (unknown) terrestrial forces and celestial dynamics and, therefore, could be predicted if the exact state of the earth's crust were known, and if we understood all of the forces or conditions that influence the triggering of an earthquake. However, since this information is impossible to obtain, or certainly outside the current state of today's seismic technology, we have no option other than to assume that earthquakes are *random events*, and attempt to use whatever information we might have available to model the state of the earth's surface in order to *predict* (to some degree of confidence or reliability) when and where the next earthquake might occur.

As yet another example, consider the measurement of the current in a resistor that is connected to a constant DC power supply.[1] Should the current through the resister be taken to be a random number, or should it be considered to be simply an unknown value that needs to be measured? And if the current was to be measured, would there be any uncertainty or randomness in the measurement? Ignoring the fact that the current through the resister is a result of the electrons moving randomly in a given direction, looking at the device that measures current (an ammeter) we would note that it has finite resolution, i.e., is only capable of measuring current to a certain level of precision. If, for example, the ammeter measures current to the nearest milliamp,[2] and if the reading is 23 mA, then all that is known (if we believe the meter) is that the current is somewhere between 22.5 mA and 23.5 mA. In other words, due to *quantization errors* there is some uncertainty in the measurement. Beyond this, however, there may be some *stray currents* that the meter picks up that adds further uncertainty or randomness in our measurement. It should be clear that this discussion is applicable to virtually any process that involves the measurement of some quantity, such as the measurement of fluid flow within a pipe, the measurement of the temperature within a gas, the measurement of the depth of the ocean floor, or the recording of an image on photographic paper or in a memory chip.

Let us now take a slightly different look at randomness, and consider the

[1] For non-electrical engineers, this means that the current that we would like to measure is, at least on some scale, a constant.

[2] For non-electrical engineers, it is sufficient to note that this is simply a unit of current much like millimeter is a unit of distance.

following sequence of fifteen decimal digits [Ref: Kalman]:

$$S_1 = \{3,\ 7,\ 3,\ 0,\ 9,\ 5,\ 0,\ 4,\ 8,\ 8,\ 0,\ 1,\ 6,\ 8,\ 8\}$$

Now let us ask ourselves the following question: "Is this a random sequence?" Before we begin to find an answer this question, perhaps we should first ask a more fundamental question: "What do we mean by random?" The concept of randomness in a sequence of numbers may be formalized in many different ways. For example, we might say that

- A sequence of numbers is *random* if there is no structure or observed patterns in the sequence.

The difficulty with this is in determining how the term *structure* should be defined, and quantifying precisely what is meant by *patterns*. And what happens if the patterns are too subtle so that they miss our detection? It may then be better to say that

- A sequence of numbers is *random* if it is impossible to *predict* the next number in the sequence from the previous numbers.

This, too, is not a very satisfying or precise definition because what criteria should be used to decide whether or not the next number can be predicted? And how would one quantify how accurate the prediction should be before it is decided whether or not the next number is predictable (surely we cannot expect to be correct all the time)? Perhaps we should just say that

- A sequence of numbers is *random* if, at any point in the sequence, any one of the ten possible digits are equally likely to occur.

If this idea is applied to the sequence S_1, we would note that there are three 8's (20%), three 0's (again 20%), and not a single 2 in the sequence. One may then be tempted to conclude that the distribution of digits is not uniform "enough" for this sequence to be truly random. However, is there any reason to believe that this sequence of numbers was not randomly generated by rolling a "fair" ten-sided die fifteen times? And in the rolling of such a die, would the sequence

$$S_2 = \{1,\ 1,\ 1,\ 1,\ 1,\ 1,\ 1,\ 1,\ 1,\ 1,\ 1,\ 1,\ 1,\ 1,\ 1\}$$

be any less likely to occur? We will see very shortly that if we assume that the die is *fair*, i.e. has no a priori biases towards landing on one number versus another, then

both sequences are *equally probable* outcomes in the *experiment* of rolling a ten-sided die fifteen times. But, for some reason, this seems contrary to our "intuition." We generally consider the first sequence to be "more random" than the second.

Thought Problem

Why do people not pick 1, 2, 3, 4, 5, 6 in the game of lotto?[3] Because they feel that this is less likely than 7, 11, 19, 25, 32, 43?

In many cases, when a particular event or phenomenon is examined, what is considered random and what is considered deterministic is often a matter of what information is given about the underlying event. For example, with regards to the sequence S_1, if no information is given on how this sequence is generated, then one might be inclined to view it as sequence of fifteen random decimal digits in the sense that, if only these fourteen digits are given, then it is not reasonable to assume that we can predict the value of the fifteenth digit. Interestingly, however, if we were given the sequence S_2, one would be inclined to say that we have a reasonably good chance of correctly predicting that the fifteenth digit in the sequence will be a "one." This feeling would probably prevail even if one were told that each digit in the sequence is chosen by randomly selecting one of ten numbered balls out of a jar and common sense would say that the next digit is equally likely to be any one of the digits from one to ten.

Taking this example one step further, suppose that we were told that the sequence S_1 represents the fifteen digits in the decimal expansion of $\sqrt{2}$, beginning with the tenth digit, would this make the sequence any less random? In this case, the answer would most likely be "yes" since, with this extra information, there is no longer any uncertainly or randomness to the sequence in the sense that any and all digits in the sequence is known or perfectly predictable (computable). On the other hand, suppose that we are told only that this sequence is a fifteen digit expansion of $\sqrt{n}$ beginning with the qth digit after the decimal point, where both n and q are integers that are chosen at random with n not a perfect square. In this case, since there is no feasible way to determine what the sequence is going to be before it is generated, and no practical way of predicting the $(k+1)$st decimal digit in

[3]Lotto is a game in which a person pays a certain amount of money in order to select (typically) six numbers between one and 59 (or some other number) in the hopes of winning money in a random drawing. In the drawing, six numbers are selected at random and, if all six numbers are the same as those selected by the person playing the game, a very large amount of money is paid for the winning ticket. Typically, smaller pay-offs are made if only four or five numbers are chosen correctly.

the sequence from the first k digits, then we would probably be forced to view the sequence as random.

It is clear from the previous discussions that formulating a definition that captures the notion of randomness is not an easy task. In fact, determining what is and what is not random is sometimes difficult, and the answer may depend upon what information is given or what assumptions are made. Furthermore, if we were to take a close look at just about any natural event or measurement in the real world, it is difficult to find one that does not have some form of randomness or uncertainty associated with it.

Challenge

Find an event that is *deterministic*, i.e., has no randomness associated with it at all.

For now, we leave the discussion of randomness, and begin thinking about how to describe and mathematically characterize events that are random. The formal development of probability theory begins by looking at how to set up a framework for describing events that are assumed to have some form of randomness or uncertainty associated with them, such as the roll of a die, the flip of a coin, the time to failure of a mechanical device, the number of fractures that occur in a steel beam over a given period of time, or the life expectancy of a human.

1-2 PROBABILITY FRAMEWORK

The primary goal of this chapter is to introduce a formalism for defining a probability measure for experimental outcomes and rules on how to manipulate these probabilities. However, before introducing probability measures, it is necessary to first set up a framework for dealing with randomness by introducing the concept of an ***experiment***, defining what is meant by a ***sample space***, and discussing what is meant by an ***event*** within the sample space.

1-2.1 EXPERIMENT

Fundamental to any discussion of probability is the concept of an ***experiment***. One view of the world is that every outcome, every observation, and every measurement, is the outcome of some underlying experiment, either real or conceived, that has

some form of randomness or uncertainty associated with the outcome.[4] Thus, it is common to use the terminology ***random experiment***. Examples include flipping a coin, counting the number of photons that hit a photodetector in a given period of time, measuring the flow of fluid in a pipe, measuring the lifetime of a light bulb, or selecting a person at random for an opinion poll. It is important to distinguish the experiment from the experimental outcomes. For example, in the coin flipping experiment, an experimental outcome would be "Heads" and in the photodetector experiment an experimental outcome would be 8,345,242 photons. Similarly, in the fluid flow experiment an outcome would be 3.2 m^3/sec whereas in the opinion poll experiment an experimental outcome would be the selection of a specific person, e.g., H. Tian, out of a pool of potential candidates.

1-2.2 SAMPLE SPACE

In dealing with random experiments, two concepts that are important to understand are sample space and event. The ***sample space***, sometimes referred to as the ***certain event***, is the set of all possible outcomes in a given experiment. However, a little care needs to be taken in how one defines "the set of all possible experimental outcomes." A simple example will illustrate the point. Consider the experiment of rolling a die. It is certainly reasonable to say that the sample space consists of six possible outcomes, each one corresponding to one of the six numbers on the die. On the other hand, it is also possible to consider (model) this experiment as one that consists of only two possible outcomes. The first is an even number (either 2, 4, or 6), and the second is an odd number (either 1, 3, or 5). These two outcomes cover all possible outcomes in the sense that no matter what number is rolled, the outcome will either be *even* or *odd*. However, there are three different ways (outcomes) that the outcome *even* may occur, and similarly for *odd*. What we are looking for in the definition of a sample space is given in the simple yet precise definition given by Drake [**??**]:

Definition

The sample space, denoted by Ω, is the *finest grained*, *mutually exclusive*, *collectively exhaustive* listing of all possible outcomes of an experiment.

[4] By an experiment, we should not be thinking of test tubes in a chemistry lab, or test subjects evaluating the effectiveness of certain drugs or medications. Here, we think of an experiment in a much more general and sometimes abstract context.

In this definition, there are three important characteristics for the list of outcomes in a sample space. The first, finest grained, imposes the constraint that none of the outcomes are collections (unions) of other outcomes. These finest grained outcomes are called ***elementary events*** and will be denoted by ω_i. The second, mutually exclusive, means that every outcome in the sample space is unique and distinct for all of the other outcomes.[5] The third, collectively exhaustive, requires that every possible experimental outcome be accounted for.

In some experiments, it is possible to specify the sample space by making a list of all possible outcomes. For example, in the coin flipping experiment, there are two elementary events,

$$\omega_1 = \{ \text{ Heads } \} \quad ; \quad \omega_2 = \{ \text{ Tails } \}$$

and the sample space is

$$\Omega = \{\text{Heads}, \text{Tails}\}$$

A simple *listing* of the elementary events may be used for any sample space that has a *discrete* set of outcomes, even when there are an infinite number of possible outcomes. For example, consider the experiment of recording (counting) the number of photons to hit a photodector over a specific period of time. In this experiment, the sample space is the set of all non-negative integers,[6]

$$\Omega = \{0, 1, 2, 3, \ldots\}$$

The coin-flipping experiment and the photon counting experiment both have a ***discrete sample space*** since it is possible to label each outcome as ω_i for some integer i. The coin-flipping experiment has a finite number of outcomes in the sample space, whereas the photodetector experiment has a countably infinite[7] number of possible outcomes.

It is also possible to have an experiment that consists of an ***uncountably infinite*** number of outcomes. For example, in the experiment of measuring the time to failure of a system, the sample space is the set of all real numbers greater than or

[5] In the notation of set theory that will be discussed in Sect. 1-3, mutually exclusive means that $\omega_i \cap \omega_j = \varnothing$ for all $i \neq j$.

[6] Although it is physically impossible for an infinite number of photons to hit the detector, it may not be appropriate to place an upper bound on the number. Therefore, it is common to leave the number unbounded, with the understanding that the likelihood (probability) of very large numbers being observed or measured may be close to zero.

[7] By countably infinite we mean that we may associate each outcome with a number, beginning with one, and going out to infinity.

equal to zero,[8]

$$\Omega = \{t \mid t \geq 0\}$$

The sample space for this experiment is said to be ***continuous*** since it is not possible to enumerate the set of all possible outcomes. Instead, the set of all possible outcomes is defined implicity by stating that it consists of all non-negative real numbers. Other examples of continuous sample spaces include the temperature in the cooling tower of a nuclear power plant, the voltage across a capacitor just prior to an electrical discharge across the plates, and the peak sound intensity from a jet engine during takeoff.

In working with experiments that involve uncertain or random events, it is very important to understand what the underlying sample space is that one is working in. Misunderstanding or misrepresenting the sample space may sometimes lead to erroneous answers or faulty analysis. Therefore, before beginning to solve any problem, it is recommended that the first step be to ***draw a picture*** of the sample space.

Useful Tip!

The first step in solving any problem in probability should be drawing a picture of the underlying sample space.

Some examples of sample spaces are given in the following examples.

Example 1-1: FLIPPING TWO COINS

An example of a discrete sample space with a finite number of outcomes is the experiment of flipping two coins. However, there are two ways that the experiment may be performed, and each one generates it own unique sample space. The first way to perform the experiment is to flip the two coins together, without any regard as to which coin is which. If we assume that the two coins are indistinguishable so that it is impossible to associate the outcomes of the flips to specific coins, then there are three possible outcomes for this experiment:

(a) Both coins are *Heads*,

(b) One coin is *Heads* and one coin is *Tails*, and

[8] The notation $\{t|t \geq 0\}$ means the set of all values of t *given* that $t \geq 0$. Here, any expression to the right of the vertical line is the condition that restricts the values to the variable to the left of the vertical line.

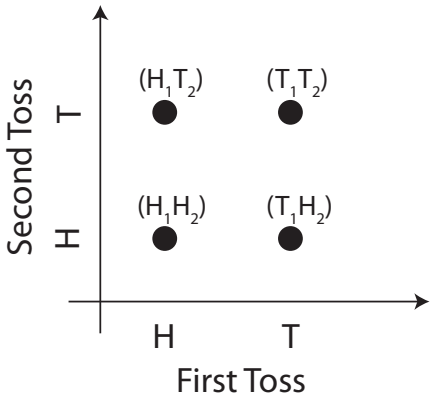


Figure 1-1: Sequential sample space for the experiment of flipping a coin twice.

(c) Both coins are *Tails*.

The second way to perform the experiment is to flip the two coins sequentially, and note the outcome of each flip individually.[9] For example, the outcome of the toss of the first coin may be denoted by H_1 and T_1, depending on whether the flip results in *Heads* or *Tails*. Similarly, H_2 and T_2 may be used to indicate the two possible outcomes of the toss of the second coin. Thus, $\{H_1T_2\}$ would be used to represent an outcome of *Heads* for the first coin and *Tails* for the second coin. For this experiment, the sample space has four possible outcomes:

$$\{H_1T_2\},\ \{T_1T_2\},\ \{H_1H_2\},\ \text{and}\ \{T_1H_2\}$$

Whether or not the coin is fair and whether or not the outcome of the flip of the first coin has any effect on the outcome of flip of the second coin does not affect the sample space or how it is represented.

A simple way to represent this sample space graphically is illustrated in Fig. 1-1. Along each axis are the two possible outcomes, H for *Heads* and T for *Tails*, and each of the four dark circles in the figure represents one of the elementary events. ■

Example 1-2: STATE OF A SYSTEM

Consider the experiment of checking the state of a system every hour until it fails. Once it fails, the experiment is terminated (the system is replaced and the

[9] Equivalently, we may label the coins as *coin 1* and *coin 2*, or we may simply flip a single coin twice.

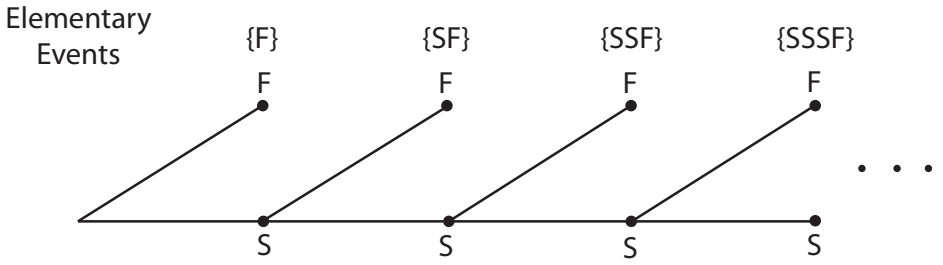


Figure 1-2: Sequential sample space for the experiment of checking the state of a system.

experiment is repeated). Each hour that the system is checked, if it is working, the state is recorded as S, and if it has failed an entry of F is entered into the log. Thus, each experimental outcome is a sequence of S's followed by a single F. For example,

$$\omega = \{SSSF\}$$

represents the event that the system is working for four days, and fails on the fifth. This sample space for this experiment, referred to as a ***sequential sample space*** may be conveniently represented using a sequential tree shown in Fig. 1-2.[10]

Note that another way to characterize this experiment is to let N be the number of days that the system is working until it fails. In this case, the sample space is the set of non-negative integers and may be represented graphically by marking the non-negative integers along the real number line. ■

The previous example is an experiment that is referred to is ***repeated trials***. More specifically, the experiment involves repeatedly performing the same experiment until some stopping condition is satisfied. In this experiment, the stopping condition is the failure of the system. The experiment of flipping a coin twice, as in the previous example, may also be considered one of repeated trials with the stopping condition being the given number of flips, in this case two. As will be seen in Chapter 4, repeated trials occur in many applications, and are often used to model or describe a number of simple yet useful random sequences.

Example 1-3: TARGET SHOOTING

Consider the experiment of shooting a rifle at a target that is 100m away. Assume that the target has a ***bulls eye*** at the center, and that the performance of the shooter

[10] Note that the sequentially flipping of two coins may also be represented in this way but, in this case, the tree would terminate after two branches.

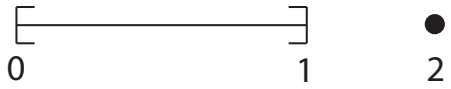


Figure 1-3: Sample space for the target shooting experiment.

is measured by the distance the bullet lands away from the bulls eye. If the target has a radius of one meter, then the set of possible outcomes from a single shot from the rifle (assuming an infinitely precise measurement of the point of impact of the bullet) is any real number between zero and one, $[0, 1]$. If the bullet misses the target entirely, then the shooter is assigned a distance of 2 meters from the bulls eye. In this experiment, the sample space is the union of the interval $[0, 1]$ and the number two,

$$\Omega = \{[0, 1], 2\}$$

This sample space is illustrated in Fig. 1-3. ■

1-2.3 EVENTS

Another important concept in probability theory is that of an ***event***. We have already introduced the term "elementary events," which are the finest grained outcomes in a sample space. In a more general context, the term event refers to any specific outcome or a set of outcomes of an experiment. More specifically, events are subsets of the sample space. For example, in the experiment of rolling a die, the following are examples of events:

$$A = \{3\}$$

$$B = \{\text{An even number is rolled}\}$$

$$C = \{\text{The number six is not rolled}\}$$

Note that A is an elementary event whereas B and C are more general events.

As another example, consider the experiment of counting the number of sunspots over a twelve month period.[11] The elementary events in this experiment

[11] Sunspot activity is of interest as they are believed to be correlated with terrestrial activity (some have gone so far as to assert that they affect human behavior). The Solar Physics Branch of the NASA Marshall Space Flight Center has studied the sunspot records to look for characteristic behaviors that might help in predicting future sunspot activity. Although sunspots themselves produce only minor effects on solar emissions, the magnetic activity that accompanies the sunspots can produce dramatic changes in the ultraviolet and soft x-ray emission levels, and these changes over the solar cycle have important consequences for the Earth's upper atmosphere.

are the non-negative integers

$$\Omega = \{0, 1, 2, \ldots\}$$

If we let n denote the number of sunspots that are counted, then

$$A = \{0 \leq n < 10\}$$

is an event that contains ten elementary events, and represents the event that *less than ten sunspots are counted*. Similarly, the event

$$B = \{n \geq 100\}$$

contains an infinite number of elementary events, and represents the event that one hundred or more sunspots are counted.

There are three special events that occur frequently in probability. The first, which is one we have already encountered, is the ***elementary event***. The second is the ***certain event***, which is the set of all possible experimental outcomes in the sample space. Thus, the event $A = \Omega$ is the certain event. Finally, there is the ***impossible event***, which is the null set or empty set, $\varnothing$. Since the impossible event contains no experimental outcomes, this event will never occur when an experiment is performed. Although it may seem to be a bit silly or absurd to talk about an event that is impossible and contains no experimental outcomes, it is important to be able to refer to an event that contains no experimental outcomes. This event will appear in the following section when operations on sets (events) are discussed. The impossible event, for example, is the complement of the certain event.

Example 1-4: **DIGITAL TRANSMITTER**

Suppose we have a transmitter that transmits, at specified times, a binary digit (either a zero or a one) across a channel to a receiver. If we consider the experiment of transmitting a binary digit at a specified time then there are two possible outcomes: either a zero or a one is transmitted. Thus, the sample space for this experiment contains only two elementary events,

$$\omega_1 = \{0\} \quad \text{and} \quad \omega_2 = \{1\}$$

Thus, there are only four events, in total: the impossible event, the two elementary events, and the certain event,

$$\varnothing,\ \{0\},\ \{1\},\ \Omega$$

■

Example 1-5: TRANSMISSION OF TWO BITS

Now consider the same transmitter as in the previous example, and let the experiment be the transmission of two binary digits. Assuming that the order of the transmission of bits is important, so that transmitting a zero and then a one is different and distinct from transmitting a one followed by a zero, then there are four elementary events,

$$\omega_1 = \{00\}\,;\ \ \omega_2 = \{01\}\,;\ \ \omega_3 = \{10\}\,;\ \ \omega_4 = \{11\}$$

This sample space has many more events than the sample space in the previous example. For example,

$$A = \{\text{Both bits the same}\} = \{00 \text{ or } 11\}$$

is an event consisting of the two elementary events $\{00\}$ and $\{11\}$, and

$$B = \{\text{ first transmitted bit is zero }\} = \{00 \text{ or } 01\}$$

is an event consisting of the two elementary events $\{00\}$ and $\{01\}$.

Exercise: Find the total number of distinct events that there are in this sample space. (Hint: The answer is either 15, 16, or 17. Don't forget the impossible event and the certain event.) ■

The previous example may be made a bit more complex and interesting by considering the experiment in which a digital transmitter produces a sequence of eight binary digits (a byte). In this experiment, there are $2^8 = 256$ elementary events in the sample space (again assuming that the ordering in which the bits are transmitted is important). One of these elementary events is

$$\omega_i = \{01100111\}$$

As discussed before in Example 1-1 in the context of flipping two coins, there are two ways to view or model this experiment. The first is to let the sample space Ω consist of a set of 2^8 elementary events, ω_i, consisting of all possible sequences of eight binary digits. The other is to view this experiment as one of ***repeated trials*** consisting of a sequence of repeated experiments with each experiment being the transmission of a single bit. With this view, a counter may be used to index the outcome of each experiment. For example, we may let $b(n)$ be the binary digit that is transmitted at time n. In this context, we are dealing with a sequence of random outcomes, or what is more generally refer to as a discrete random sequence as we will see later in Chapter 21. An advantage of this second approach of viewing

the experiment is that a number of important generalizations may be made easily within this framework, such as introducing statistical dependencies between the bits or allowing the probabilities to change between one time and the next.

Example 1-6: FAX MACHINES AND RUN LENGTH ENCODING

An interesting example related to the previous two examples is the transmission of a black and white document by a fax machine. When a document is scanned by the fax machine, the scanner determines whether a small square or rectangular area within a particular scan line in the document should be represented by a *black pixel* or a *white pixel*. With white pixels represented by zeros and black pixels by ones, the transmission of the scanned document involves the sending of a sequence of zeros and ones to the machine that is to receive the fax.

Since it is generally true that most documents are predominately white, except for certain areas where there may be text, the sequence of bits generated by the scanner typically have long runs of zeros, representing long stretches of the document where there is no text. Therefore, rather than transmitting every single output from the scanner (the sequence of zeros and ones) it may be much more efficient to transmit the ***run length*** of the zero pixels. For example, rather than transmitting the sequence of 32 bits,

$$00000000000000010000000110000001$$

the fax machine would send the following sequence of numbers:

$$14,\ 8,\ 0,\ 6$$

which would be decoded as

- fourteen white pixels followed by a black pixel,
- eight white pixels followed by a black pixel,
- zero white pixels followed by a black pixel, and
- six white pixels followed by a black pixel.

Depending on how these numbers are encoded, this may be a much more efficient method of transmitting the output of the scanning device. For example, with the simple representation of each of the run-length numbers with a four-bit number, the transmitted sequence would be

$$1110, 1000, 0000, 0110$$

for a total of sixteen bits, half the number in the original sequence. Such encoders are called ***run-length encoders***. In this example, all run-lengths are assumed to be

equally likely, and each run length is encoded with the same number of bits, i.e., four. In many applications, long run lengths will be more likely than short ones, and variable length encoding may significantly increase the efficiency of the encoder.

Once we have introduced the concept of a probability measure and are able to assign probabilities to the run lengths, we will be in a position to a position to understand how a ***Huffman coder*** may be used to efficiently encode the run lengths. Huffman coders are important in data compression, and are found in many compression systems such as JPEG (images) and MPEG (video). ■

1-3 SET THEORY

In probability theory, one often encounters events that are defined in terms of other events. For example, consider the event A that someone over the age of fifty gets the flu, and the event B that someone who receives a flu vaccination and gets the flu. If we are interested in the event that someone over the age of fifty gets the flu ***and*** has been vaccinated, then we are dealing with an event that is a combination (intersection) of the events A and B. In order to work with events such as this, it is necessary to introduce a few basic concepts from set theory that allow us to perform operations on sets. In particular, four set operations that will be useful are: union, intersection, complement and difference. In the following paragraphs, each of these operations are defined and discussed in the context of events, which are ***sets of experimental outcomes***. To illustrate and help explain these operators, it will be convenient to use a graphical device called a ***Venn diagram***. With a Venn diagram, the sample space, Ω, is represented abstractly as a rectangle, and events are represented as regions inside this rectangle. The shapes of these regions are irrelevant, and only their relationship to each other is important. For example, shown in Fig. 1-4 is a Venn diagram representing a sample space Ω along with three events, A, B, and C. Note that A is shown as being separate and distinct from events B and C indicating that A has no experimental outcomes in common with either B or C. Events B and C, on the other hand, are shown intersecting each other, indicating that these two events have experimental outcomes in common.

The first operator of interest is the union. Given two sets, A and B, the ***union*** is the set C that contains the elements that are in either ***A or B***. The union of A and B is denoted either by[12]

$$C = A \cup B \tag{1.1}$$

[12]In this book, both ways of expressing the union will be used. Although the first is generally preferred, the summation sign is often convenient since it often makes expressions simpler and more intuitive to read and to understand. The same comment applies to the intersection, which is defined next.

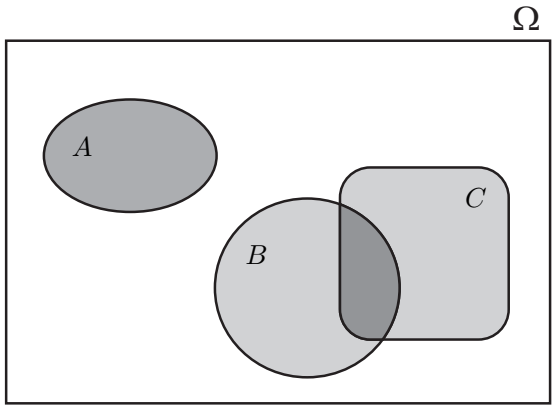


Figure 1-4: A Venn diagram representing a sample space Ω and three events, A, B, and C.

or

$$C = A + B \tag{1.2}$$

Thus, if A and B are events, then the event $C = A \cup B$ occurs if event A occurs, event B occurs, or both events A and B occur (there may be outcomes that are common to both A and B). For example, in the transmission of a digital image across a network, if A is the event that no bits are received in error and the event B is the event that one bit is received in error, then $C = A \cup B$ is the event that no more than one bit is received in error. A picture illustrating the union of two events is given in Fig. 1-5a.

The next operator is the ***intersection*** of two sets, A and B, which is denoted either by

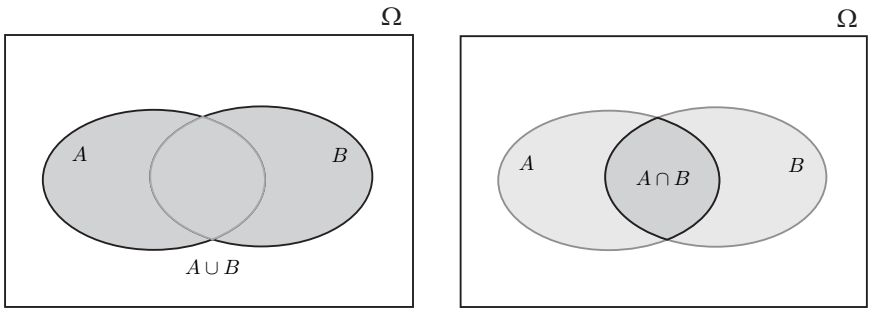
$$C = A \cap B$$

or

$$C = AB$$

The intersection of A and B is the set that contains all elements that are in both A ***and*** B. In other words, an element is contained in the set C if it is in both set A and in set B. Thus, for any two events, A and B, the event $C = A \cap B$ will occur only if both event A occurs and event B occurs. A picture illustrating the intersection of two sets is given in Fig. 1-5b.

An example of the intersection of two events is the following. Let A be the event that there are an equal number of zeros and ones in the transmission of eight binary digits across a digital communication channel, and let B be the event that



(a) Union (b) Intersection

Figure 1-5: Set operations of union and intersection. (a) The union of the sets A and B includes the entire shaded area. (b) The intersection of the sets A and B consists of the dark shaded area, which is common to both A and B.

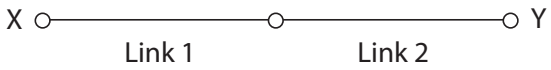


Figure 1-6: Series connection of two communication links in a computer network.

the first four bits are zero. Then $A \cap B$ contains a single elementary event,

$$A \cap B = \{00001111\}$$

since this is the only event that has an equal number of zeros and ones with the first four being equal to zero. As another example, consider a computer network that consists of two links that are connected in series as illustrated in Fig. 1-6. If A is the event

$$A = \{\text{Link 1 is available}\}$$

and B is the event

$$B = \{\text{Link 2 is available}\}$$

then the event $C = A \cap B$ is the event

$$C = \{\text{Both links are available}\}$$

and, therefore, communication between Node X and Node Y is possible.

If the intersection of two sets is empty,

$$A \cap B = \varnothing$$

then A and B are said to be ***disjoint*** sets, or that A and B are ***mutually exclusive***. In the experiment of measuring the time T in hours until a light bulb fails, the events $A = \{T > 1000\}$ and $B = \{T \leq 1000\}$ are mutually exclusive.

In some cases one event will be contained in or be a subset of another. For example, in the experiment of measuring the number of inches of rain that fall over the ocean near the island of Hawaii over a twelve month period, consider the events

$$A = \{20 < R \leq 40\} \quad ; \quad B = \{25 < R \leq 30\}$$

It is clear that if event B occurs, then event A also occurs since B is included in A. This relationship is denoted by $B \subset A$. Note that if $B \subset A$ then

$$A \cup B = A \quad \text{and} \quad A \cap B = B$$

The next set operation is the complement. For any set A, the ***complement*** of A, denoted by A^c, is defined to be the set of all elements that are *not* in A. In terms of events, if A is an event in the sample space Ω of all possible experimental outcomes, then A^c is the set of all outcomes in Ω that are not in A. Therefore, if the event A occurs, then A^c does not occur. A picture illustrating the relationship between a set A and its complement is given in Fig. 1-7. Note that for any event A

$$A \cap A^c = \varnothing$$

i.e., A and A^c are mutually exclusive events, and

$$A \cup A^c = \Omega$$

The last operation of interest is the ***set difference***, which is defined as follows. If A and B are two sets, then the set difference, denoted by

$$C = A - B$$

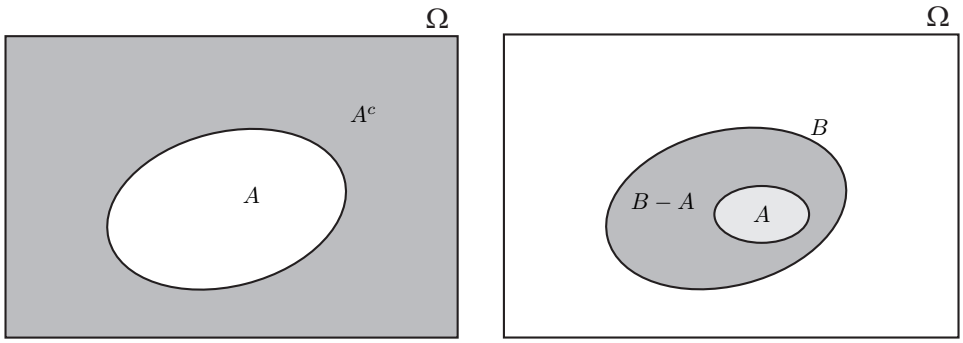
is the set of elements in A that are not in B. More formally, the set difference is

$$A - B = \{x \in A \,|\, x \notin B\}$$

Thus, think of $A - B$ as the set of elements in A that remain after the removal of all elements in B that are contained in A. For example,

$$\{1, 2, 3\} - \{2, 3, 4\} = \{1\}$$

As another example, if $\mathbb{R}$ is the set of real numbers and $\mathbb{Q}$ is the set of rational numbers, then $\mathbb{R} - \mathbb{Q}$ is the set of irrational numbers. A picture illustrating the difference of two sets A and B, sometimes called the ***relative complement of B in A***, is given in Fig. 1-7b.



(a) Complement (b) Set Difference

Figure 1-7: Set operations of complement and difference.

1-3.1 THE ALGEBRA OF EVENTS AND DEMORGAN'S LAWS

The previous section defined the basic set operations of union, intersection, complement, and difference. In probability theory, and event of interest may be defined in terms of a sequence or combination of these set operators that are applied to one or more sets. Therefore, it is important to understand the rules under which these operators may be manipulated or simplified. Formally, there are seven laws or ***axioms*** that fully define the ***algebra of events*** and provide the tools necessary to manipulate expressions that involve the set operations of union, intersection, and complement. Although interesting in their own right, a complete and thorough development of these axioms is not essential for a solid understanding of probability theory. However, an awareness of what these axioms are, what they mean, and how to use them is important. These seven axioms are listed below.

Axioms for the Algebra of Events

1.	Commutative	$A \cup B = B \cup A$
2.	Associative	$A \cup (B \cup C) = (A \cup B) \cup C$
3.	Distributive	$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
4.	Double complement	$(A^c)^c = A$
5.	Mutual exclusion	$A \cap A^c = \varnothing$
6.	Inclusion	$A \cap \Omega = A$
7.	DeMorgan	$(A \cap B)^c = A^c \cup B^c$

With the possible exception of Axiom 7, which is known as ***DeMorgan's law***, these axioms should be obvious and self-evident, and the reader should study them to get an intuitive feel for what each one means. DeMorgan's Law, which is less obvious, states that

$$(A \cap B)^c = A^c \cup B^c \tag{1.3}$$

A similar expression that may be derived from Eq. (1.3) is

$$(A \cup B)^c = A^c \cap B^c \tag{1.4}$$

To see how Eq. (1.4) may be derived from Eq. (1.3), note that if A and B in Eq. (1.3) are replaced by their complement, A^c and B^c, respectively, then

$$(A^c \cap B^c)^c = A \cup B \tag{1.5}$$

and taking the complement of both sides of (1.5) gives Eq. (1.4). The pair of equations, Eq. (1.3) and Eq. (1.4), are commonly referred to as DeMorgan's Laws.

It is instructive to visualize DeMorgan's Laws graphically. For example, DeMorgan's Laws are illustrated graphically in Fig. 1-8 for the case in which $A \cap B \neq \varnothing$. In this figure, note that $(A \cap B)^c$ corresponds to those elements or outcomes that are outside the shaded region labeled $A \cap B$. Note that this region consists of all outcomes that are outside of A plus all of those that are outside of B, i.e., in the set $A^c \cup B^c$.

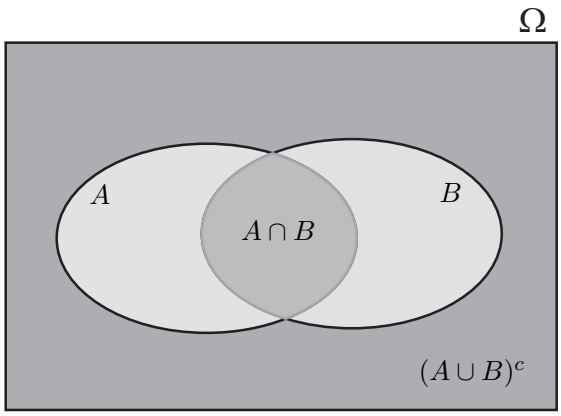


Figure 1-8: Graphical illustration of DeMorgan's Laws, $(A \cap B)^c = A^c \cup B^c$ and $(A \cup B)^c = A^c \cap B^c$.

DeMorgan's Laws may be generalized to unions and intersections of more than two sets. Specifically, given sets $A_1, \ldots, A_n$, it follows by induction that

$$\left[\bigcap_{i=1}^{n} A_i\right]^c = \bigcup_{i=1}^{n} A_i^c \tag{1.6}$$

and

$$\left[\bigcup_{i=1}^{n} A_i\right]^c = \bigcap_{i=1}^{n} A_i^c \tag{1.7}$$

We now leave the algebra of sets and events, and turn to the question of how to assign probabilities to experimental outcomes and events.

1-4 PROBABILITY MEASURE

Every day life is filled with expressions and statements that are probabilistic in nature, and the term "probability" is frequently found in articles on the web, in magazine and newspaper reports, and in casual conversations. Examples include statements such as "The probability of getting an 'A' in this class is small," a weather report that states that "There is a chance (probability) of scattered thunderstorms developing in the evening," and a stock analyst's report asserting that "Given the strong earnings growth of the company and the low P/E ratio, the

stock price is expected to double in the next twelve months."[13] Although each of these are statements about the likelihood or not of some random or uncertain event out outcome, they lack a quantitative measure of our belief in the likelihood of one outcome versus another. We come a bit closer to defining something quantitative with a statement such as "There is a 90% chance of rain showers tomorrow," or "The probability of getting the flu is ten times higher if one is over the age of sixty and has not received a flu vaccination." However, what is needed is a procedure for assigning a probability measure to a random event, and a process for evaluating the probability of one event that may be defined in terms of other events. Therefore, the next step in our journey into the world of probability is to define a quantitative measure for events in a sample space of an experiment.

With our eyes set on a quantitative measure for the probability of a random event, we first turn to one of the early approaches to probability known as the ***classical theory***. Generally attributed to the French mathematician and astronomer Pierre Simon Laplace (1749-1827), the classical approach assigns a number to the probability of an event E that is the ratio of the number of ***favorable outcomes*** in E, i.e., the number of possible outcomes associated with the event E, to the ***total number of possible outcomes***. In other words, if out of a total of N possible outcomes in some experiment there are N_E favorable outcomes in an event E, then the probability of E, denoted by $\mathrm{P}\{E\}$, is the number

$$\mathrm{P}\{E\} = \frac{N_E}{N} \tag{1.8}$$

Note that since $0 \leq N_E \leq N$, then the probability of any event is non-negative and bounded by one,

$$0 \leq \mathrm{P}\{E\} \leq 1$$

As a specific example, let us return to the experiment of rolling a die, and consider the problem of assigning a number to the probability of the event E that the outcome of the roll is an even number. Since the total number of favorable outcomes (either a two, a four, or a six) is equal to three, then $N_E = 3$, and since the total number of possible outcomes is six, then $N = 6$. Therefore,

$$\mathrm{P}\{E\} = \frac{N_E}{N} = \frac{3}{6} = 0.5$$

This is certainly very reasonable since half the time we expect to roll an even number, and half the time we expect an odd number. There are,

[13] Here the term ***expected*** is a term rooted in probability as will be seen in Chapter 6 when the concept of ***expectation*** and ***expected value*** are introduced.

however, two problems with this approach. The first is that it assumes that all outcomes are equally likely, an assumption that is referred to as the ***principle of indifference*** [Ref.]. To better understand why this is a problem, suppose that we would like to assign a probability to the event that the outcome of the flip of a coin is *Heads*. With the classical approach, since there is only one favorable outcome, *Heads*, and a total of two possible outcomes, *Heads* and *Tails*, then

$$\mathrm{P}\{H\} = \tfrac{1}{2}$$

However, this approach assumes that the coin is "fair." It does not allow for the possibility that we may have an unfair coin, i.e., one that is weighted so that it is more likely to land either on *Heads* or *Tails* (or has two *Heads*). In addition, suppose that we would like to allow for the (extremely unlikely) event that the coin will land on its edge. In this case, the total number of outcomes becomes three, and the probability of *Heads* becomes

$$\mathrm{P}\{H\} = \tfrac{1}{3}$$

As another example, consider the experiment of selecting a book at random, opening it to a random page, and then randomly selecting a letter on that page. Since the total number of possible outcomes is twenty-six, then $N = 26$. If we want to assign a probability to the event that we select one of the letters *x, q, j* or *z*, then there are four *favorable outcomes*. Therefore, with the classical approach we have

$$P\{x,\, q,\, j,\, z\} = 4/26 = 0.1538,$$

or slightly more than a 15% chance. However, we know from our experience with the English language that all letters are not equally likely, and given the unlikelihood of each of these four letters, we would view this probability as being far too high. In fact, from experiments designed to estimate the probability, or frequency of occurrence, of the letters in the English alphabet in normal text, the probability of one of these four letters being selected is approximately [**??**]:

$$P\{x,\, q,\, j,\, z\} \approx 0.0044$$

The second problem with the classical approach is that it cannot handle experiments that have sample spaces with an infinite number of outcomes. For example, suppose that we would like to develop a probabilistic description for the *time to failure* of a specific device, such as a light bulb that is manufactured by a particular company. In order to promote the "long-life" of these light bulbs, we may want to show that the probability is very high that a light bulb will last more

than 10,000 hours. Since the time to failure may be any real number $t \geq 0$, the number of possible outcomes is infinite, and it is not possible to evaluate the ratio of the number of favorable outcomes to the total number of possible outcomes. Therefore, the classical approach is generally limited to experiments that have finite sample spaces.

Another approach to probability is known as the ***relative frequency approach***. A simple example that illustrates the basic idea is the following. In a weather forecast we may hear a statement such as "there is a 50/50 chance of rain tomorrow," and one generally interprets this statement to mean that fifty times out of a hundred, for the given atmospheric conditions, rain can be expected.[14] A 50% probability, or a probability of one-half, is then this ratio of fifty "successes" out of 100 chances or "trials". Thus, the relative frequency approach to assigning probabilities is based on the idea of performing n ***independent*** experiments and recording the number of times, n_E, that the event E occurs. The probability that is assigned to the event E is then given by

$$\mathrm{P}\{E\} = \lim_{n\to\infty} \frac{n_E}{n} \tag{1.9}$$

Although the term independence has not yet been defined, for now we may use the literal definition of independence and take it to mean that the outcome of one experiment has no effect or influence on the outcome of any experimental outcome.

Since it is not reasonable to assume that an experiment may be performed an infinite number of times, no matter how patient we are, assigning probabilities using Eq. (1.9) is not feasible. Therefore, it is generally assumed that if the experiment is performed a sufficiently large number of times, then n_E/n should be close to $\mathrm{P}\{E\}$,

$$\mathrm{P}\{E\} \approx \frac{n_e}{n} \tag{1.10}$$

and this approximation is then taken as the probability $\mathrm{P}\{E\}$.

An interesting question to ask is this: How many times does an experiment need to be performed in order for this approximation to be "good enough"? Another interesting question is: For a given number of times that an experiment is performed, how certain can one be that the approximation Eq. (1.10) is within a certain precision of the "true" probability? Chapter **??** addresses these questions and examines how good of an estimate the approximation in Eq. (1.10) is for $\mathrm{P}\{E\}$.

[14] Paul Harvey, who was a well-known American radio broadcaster for the ABC Radio Networks, had another interpretation for such a statement when it deals with the likelihood of something going wrong. He said that "If there is a 50/50 chance that something can go wrong, then nine times out of ten it will."

Unlike the classical approach, the relative frequency approach does not require that the outcomes be equally likely. The relative frequency approach also provides a mechanism for assigning probabilities to events that are difficult or impossible to assign using the classical approach. For example, consider the assignment of a probability to the event of having a dropped phone call within a cellular telephone network. With the classical approach, it is not clear how one would assign a probability to this event. Since there are only two possible outcomes (dropped call or no dropped call), then the classical approach would set the probability of having a dropped call equal to $1/2$, which clearly is not what the probability should be, particularly in light of the fact that this probability would be the same for any region, for any cellular network, and at any time of day. Alternatively, using the relative frequency approach, suppose that over some period of time we make n phone calls where the outcome of each call (dropped call or no dropped call) is independent of the others. Then, if E is the event that there is a dropped call, and if out of the n phone calls there are n_E dropped calls, then the probability of event E would be given by Eq. (1.9).

In spite of its advantages over the classical approach, the relative frequency approach also has some problems. The first is that it is necessary to assume that the ratio n_E/n approaches a limit as n goes to infinity, and that this limit corresponds to what we call the probability of event E. However, it is not clear in what sense this ratio might converge, or even that it will converge, especially given that we are dealing with a sequence of numbers that is not deterministic.[15]

The second problem, as already mentioned, is that it is not possible to perform an experiment an infinite number of times, so it is necessary to assume that for sufficiently large n, the ratio n_E/n is close to $\mathrm{P}\{E\}$, However, in some cases it may not be feasible to perform an experiment a sufficient number of times for this approximation to be used, and in some cases it may not even be possible to perform the experiment even once. For example, consider the case of assigning a probability to the event that a specific volcano will erupt within the next 100 years, or the probability that life exists on another planet. Using the relative frequency approach to assign probabilities to these events is not realistic, and it becomes necessary to resort to experience, historical data, or some other means for the assignment of a probability.

Given the difficulties with both the classical and the relative frequency approaches, an alternative is to simply assign probabilities to events based on some reasonable set of criteria. For example, a mathematical model or empirical

[15]Chapter **??** looks more closely at issues related to the convergence of a sequence of random numbers.

data from an experiment may be available that may be used to make probability assignments. Alternatively, we may have sufficient experience with an experiment that allows us to assign probabilities to events. For example, consider the experiment of flipping a coin, and the task of assigning probabilities to the two outcomes *Heads* and *Tails*. Our experience would indicate that it is equally likely for a coin toss to result in *Heads* or *Tails*, assuming that the coin is *fair*. In this case we would simply set

$$\mathrm{P}\{\mathrm{Heads}\} = \mathrm{P}\{\mathrm{Tails}\} = 1/2 \tag{1.11}$$

Alternatively, we may take this as the definition of a *fair coin*, i.e., a coin is fair if Eq. (1.11) holds. But this approach raises a number of important and difficult questions.

1. What do we do for an unfair coin? How do we determine what probability to assign to *Heads* and *Tails* in this case? How do we know that a coin is fair? What test can we use to determine whether or not a coin is fair?

2. What do we do for more complex systems? For example, how would we assign probabilities for the occurrence of an event such as the time to failure of a device or the arrival of a packet of information over a network? Or how would we assign a probability that the outcome of a particular medical trial is positive?

3. What rules should we place on making probability assignments? More importantly, what constraints, if any, must we impose on the assignment of probabilities so that we have a self-consistent framework upon which to build a theory of probability?

The first two questions are difficult ones, and will not be considered here. Therefore, we turn our attention to the third question, so that instead of worrying about *what* the probabilities should be that are assigned to events, we will concern ourselves with the question of *how* these probabilities should be assigned. In other words, what are the rules that should be used when these assignments are made? The answer to this question lies in the ***axiomatic theory*** of probability. This theory is founded upon three *axioms* that probability assignments must satisfy in order to build a consistent theory of probability. It will then up left up to the systems engineer, the scientist, or the data analyst to decide how to assign probabilities to events of interest that are consistent with and satisfy these axioms.

1-4.1 THE PROBABILITY AXIOMS

The axiomatic theory of probability is elegant and powerful. And yet, this theory is built upon three very simple (and intuitive) axioms, just as electromagnetic field theory is built upon four fundamental (not so intuitive) equations, called Maxwell's equations, and just as the foundation for Boolean Algebra is based on seven axioms introduced by Boole in 1854.[16] As long as probabilities are assigned to events in such a way that they satisfy these three axioms, we are guaranteed to have a ***legitimate*** and ***self-consistent*** probability space to work in. These three axioms are:

Probability Axioms

(1) For any *event* A, the probability of A is non-negative,

$$\mathrm{P}\{A\} \geq 0$$

(2) The probability of the *certain event* Ω is equal to one,

$$\mathrm{P}\{\Omega\} = 1$$

(3) For any two mutually exclusive events, A and B, the probability of the *union* is the sum of the probabilities of the individual events,

$$\mathrm{P}\{A \cup B\} = \mathrm{P}\{A\} + \mathrm{P}\{B\}$$

It is important to point out that these axioms have no connection to or association with any natural or physical system or to any experiment. They only provide a framework upon which a self-consistent theory of probability can be built. Furthermore, no rules are given on how to assign probabilities to events. This is the job for the scientist, the engineer, the mathematician, the statistician, the data analyst or the probability expert. However, whatever probabilities are assigned to events in Ω, they must be made in such a way so that the probability of any event in

[16] Beginning with the premise that there is a set B and two operators, $+$ and $*$, the seven axioms are: closure, cardinality, commutative, associative, existence of an identity element, distributive, and the existence of a complement element.

Ω may be found. When this is done, the probability assignments are said to ***provide a complete probabilistic description of the experiment***.

The first axiom places a measure on probabilities that prevents them from being negative. The second axiom states that $\mathrm{P}\{\Omega\} = 1$, which is a consequence of the fact that all possible outcomes are contained within the sample space Ω and, therefore, the probability that some outcome in Ω occurs when the experiment is performed, must equal one. The third axiom, called the ***additivity axiom***, is the most restrictive, and may be generalized to unions of any finite number of mutually exclusive events. Specifically, if $A_1, A_2, \ldots, A_m$ are mutually exclusive events, $A_i \cap A_j =$ for $i \neq j$, then it follows by induction that

$$\mathrm{P}\{A_1 \cup A_2 \cup \cdots \cup A_m\} = \mathrm{P}\{A_1\} + \mathrm{P}\{A_2\} + \cdots + \mathrm{P}\{A_m\}$$

Many experiments in a variety of applications have sample spaces with an infinite number of possible outcomes. Examples include the number of bits that are transmitted across a digital communication channel before the first error in transmission occurs, the selection of a radioactive particle at time $t = 0$ and recording the time at which the first radioactive emission occurs, and the distance a new car travels before it breaks down. For sample spaces such as these, it may be necessary to consider an infinite union of mutually exclusive events. In this case, it is necessary to strengthen Axiom 3 and require that

$$\mathrm{P}\left\{\bigcup_{k=1}^{\infty} A_k\right\} = \sum_{k=1}^{\infty} \mathrm{P}\{A_k\} \quad ; \quad A_i \cap A_j = \varnothing \text{ for all } i \neq j \tag{1.12}$$

Eq. (1.12) is referred to as the ***countable additivity axiom***.

1-4.2 Consequences of the Probability Axioms

We now turn our attention to a few important consequences that follow from these probability axioms. This will mark the beginning of our development of a powerful and useful theory of probability. The first consequence is the following:

Consequence 1

If an event A has probability $\mathrm{P}\{A\}$, then the probability of the complement, A^c, is

$$\mathrm{P}\{A^c\} = 1 - \mathrm{P}\{A\} \tag{1.13}$$

This follows directly from Axioms 2 and 3. Specifically, since

$$A \cup A^c = \Omega$$

then

$$\mathrm{P}\{A \cup A^c\} = \mathrm{P}\{\Omega\} = 1$$

Since A and A^c are mutually exclusive, $A \cap A^c = \varnothing$, then it follows from Axiom 3 that

$$\mathrm{P}\{A \cup A^c\} = \mathrm{P}\{A\} + \mathrm{P}\{A^c\}$$

Therefore,

$$\mathrm{P}\{A\} + \mathrm{P}\{A^c\} = 1$$

and Eq. (1.13) follows.

Based on our everyday experience with probabilities and uncertainties, the property given in Eq. (1.13) is certainly intuitive. For example, when the weatherman says that there is a 95% chance of rain, he is saying that the probability of rain is 0.95. It then follows that there is a 5% chance that it will not rain, or

$$\mathrm{P}\{\text{No Rain}\} = 1 - \mathrm{P}\{\text{Rain}\} = 1 - 0.95 = 0.05$$

Although very simple, Eq. (1.13) can be extremely useful in finding the solution to what seems to be a difficult problem. For example, to find the probability of some event A, it may be much easier to find the probability of the complement A^c and then use Eq. (1.13). An illustrative example is given below.

Example 1-7: RANDOM POINTS IN TIME

A number of applications involve experiments that involve random points in time. An example is the phenomenon of radioactive decay. Although it is not possible to know the precise moments at which radioactive emissions occur, it is reasonable to assume that an emission is equally likely to occur at any point in time. With this in mind, consider the experiment of counting the number of emissions over a one second time interval. The sample space of this experiment is the set of all non-negative integers,

$$\Omega = \{0, 1, 2, \ldots\}$$

With the emissions assumed to be equally likely at any point in time, it will be shown in Chapter **??** that the number of emissions over a one second time interval follows a ***Poisson Probability Law***, which is given by

$$\mathrm{P}\{k \text{ emissions}\} = \frac{\lambda^k}{k!} e^{-\lambda} \quad ; \quad k = 0, 1, 2, \ldots$$

where $\lambda > 0$ is the ***rate parameter*** that represents the average number of emissions per second that can be expected to occur. It may be shown that this probability assignment satisfies the three probability axioms, and provides a complete probabilistic description of the experiment. Specifically, it is clear that $\mathrm{P}\{k \text{ emissions}\} \geq 0$ for all k, and that

$$\sum_{k=0}^{\infty} \mathrm{P}\{k \text{ emissions}\} = 1$$

which follows from the Taylor series expansion of e^{λ} given by

$$\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{\lambda}$$

The third axiom holds because of the way in which the probabilities as assigned, i.e., to the elementary events $\{k \text{ emissions}\}$ for all k, with

$$\mathrm{P}\{(k \text{ emissions}) \cup (l \text{ emissions})\} = \mathrm{P}\{k \text{ emissions}\} + \mathrm{P}\{l \text{ emissions}\}$$

when $k \neq l$. Finally, it is clear that the probability of any event $A \in \Omega$ may be found by summing the probabilities of all elementary event that lie within A, and thus we have a complete probabilistic specification of the experiment.

Given this model, suppose that we would like to find the probability that there is more than one arrival in a one second interval. To simplify notation, let N be the number of emissions that are counted in one second, and let the event $\{k \text{ emissions}\}$ be denoted by $\{N = k\}$.[17] Since $\{N = k\}$ and $\{N = l\}$ are mutually exclusive events if $k \neq l$, using the countable additivity axiom in Eq. (1.12) it follows that

$$\mathrm{P}\{N > 1\} = \sum_{k=2}^{\infty} \mathrm{P}\{N = k\} = \sum_{k=2}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda}$$

Although it is possible to evaluate this sum, it is much easier to find this probability using Eq. (1.13) as follows:

$$\begin{aligned} \mathrm{P}\{N > 1\} &= 1 - \mathrm{P}\{N \leq 1\} \\ &= 1 - \big[\mathrm{P}\{N = 0\} + \mathrm{P}\{N = 1\}\big] \\ &= 1 - e^{-\lambda} - \lambda e^{-\lambda} \end{aligned}$$

[17] Here, N represents what we call a ***random variable***, a concept to be presented in Chapter 6.

which is the probability that we wanted to find. ■

Useful Tip!

If it is difficult to find the probability of an event A, consider finding the probability of A^c and then use Eq. (1.13) to find $\mathrm{P}\{A\}$.

A special case of Eq. (1.13) follows when $A = \Omega$. In this case, $A^c = \varnothing$ and

$$\mathrm{P}\{\varnothing\} = 1 - \mathrm{P}\{\Omega\} = 1 - 1 = 0$$

This, of course, is certainly reasonable since when an experiment is performed, some outcome or event must occur. Since $\varnothing$ is the empty set, then this probability should be zero.

The next consequence of the probability axioms establishes the relationship between the probabilities of two events when one is a subset of the other.

Consequence 2

Probabilities are monotonic in the sense that if A is a subset of B, $A \subseteq B$, then

$$\mathrm{P}\{A\} \leq \mathrm{P}\{B\} \tag{1.14}$$

This is an intuitive result that should also be obvious. Since any outcome in A will also be an outcome in B when $A \subseteq B$, then the probability of event A will be at least as large as the probability of event B. And since B may contain outcomes are *not* contained in A, then $\mathrm{P}\{B\}$ may, in fact, be larger than $\mathrm{P}\{A\}$. For example, let A be the event that the temperature T of a semiconductor device is greater than 40 degrees C,

$$A = \{T \geq 40\}$$

and B the event that the temperature T is greater that 30 degrees C,

$$B = \{T \geq 30\}$$

Since $A \subseteq B$, then

$$\mathrm{P}\{T \geq 40\} \leq P\{T \geq 30\}$$

A useful and important corollary follows from Eq. (1.14) by setting $B = \Omega$. Since any set A is a subset of Ω, and since $\mathrm{P}\{\Omega\} = 1$, then

$$\mathrm{P}\{A\} \leq 1$$

In other words, the probability of any event is never larger than one. This result, combined with the second axiom, constrains the probability of any event A to be between zero and one.

Important Check

In solving any probability problem, always check to make sure that any calculated probabilities are between zero and one, i.e., for any event A

$$0 \leq \mathrm{P}\{A\} \leq 1$$

The third axiom states that if A and B are mutually exclusive events, then the probability of ***either A or B*** is the sum of the probabilities of A and B. What happens if A and B are not mutually exclusive? This answer is given below.

Consequence 3

For any two events A and B,

$$\mathrm{P}\{A \cup B\} = \mathrm{P}\{A\} + \mathrm{P}\{B\} - \mathrm{P}\{A \cap B\} \tag{1.15}$$

Note that $P\{A \cap B\} = 0$ when $A \cap B = \varnothing$ and Eq. (1.15) is then equivalent to Axiom 3. The third term in Eq. (1.15) accounts for any outcomes that are common to both A and B. Since the probability of these events would be counted twice if the probability of A was added to the probability of B, then this term performs the necessary correction. As an illustration, consider the experiment of rolling a single die once, and let A be the event that the outcome is an even number and B be the event that the outcome is greater than or equal to three. If the die is fair, then we would assume that all outcomes are equally likely, and we would have

$$\mathrm{P}\{A\} = 1/2, \quad \mathrm{P}\{B\} = 2/3$$

Since

$$A \cup B = \{2,4,6\} \cup \{3,4,5,6\} = \{2,3,4,5,6\}$$

then

$$\mathrm{P}\{A \cup B\} = 5/6$$

Note that if we were to add $\mathrm{P}\{A\}$ to $\mathrm{P}\{B\}$ then we would be double counting the elementary events $\{4\}$ and $\{6\}$ since these events are common to both A and B. Since $A \cap B = \{4,6\}$ then $P\{A \cap B\} = 1/3$, and using Eq. (1.15) we correctly find the probability of $A \cup B$ as

$$P\{A \cup B\} = P\{A\} + P\{B\} - P\{A \cap B\} = 1/2 + 2/3 - 1/3 = 5/6$$

1-4.3 PROBABILITY ZERO

Axiom 1 imposes the requirement that the probability of an event must be greater than or equal to zero. Although most events of interest will have a non-zero probability, it is also possible for an event to have a probability of zero. For example, we have seen that the probability of the empty set (the null event) is zero, $\mathrm{P}\{\varnothing\} = 0$. However, it is also possible for a non-empty set to have a probability of zero, and one of the confusing and subtle points in probability theory is the notion that if the probability of an A is equal to zero,

$$\mathrm{P}\{A\} = 0$$

this does not necessarily mean that A will never occur, or that it is an ***impossible event***. In other words, even when $\mathrm{P}\{A\} = 0$, in some cases it is possible that the event A may occur, but it is extremely unlikely that it will. This seemingly contradictory statement will be explored later, but for now the following example will serve as a useful illustration.

Example 1-8: INFINITE PRECISION ROULETTE WHEEL

Suppose that we have a roulette wheel that is infinitely calibrated between zero and one, i.e., when the wheel is spun any real number between zero and one may appear (we are assuming that we have a device that is able to measure where the wheel lands to infinite precision). Also assume that any number between zero and one is equally likely to occur. In this case, the probability of the roulette wheel landing on some number, such as $1/\sqrt{2}$, must be equal to zero since there is an infinite number of other values that are equally likely to occur. To clarify this point, suppose that the probability of the wheel landing on any given number is some small but nonzero value, $\epsilon > 0$. Since each number is equally likely to occur, then

for any N distinct numbers between zero and one, the probability that the wheel will land on any one of these will be $N\epsilon$ by Axiom 3. Since there are an infinite number of values between zero and one, if we try to find the probability that the roulette wheel lands on any number between zero and one (the sample space) we will find that we violate Axiom 2 for any $\epsilon > 0$. Therefore, ϵ must equal zero, and the probability that the wheel lands on any given number must be zero. This does not mean, however, that it is *impossible* for the roulette wheel to land on $1/\sqrt{2}$. In fact, each time that the wheel is spun, it lands on some number, and no matter what the number is, it has a probability of zero of occurring! ■

In situations when the probability of an event is equal to zero, but the event is not impossible, the event is said to ***almost surely*** never happen. The following example further illustrates the concept of an event tat ***almost surely*** never happens.

References

1. Alvin W.Drake, *Fundamentals of Applied Probability Theory*, McGraw-Hill, New York, 1967.

2. Harold J. Larson and Bruno O. Schubert, *Random Variables and Stochastic Processes, Volume 1*, John Wiley & Sons, 1979.

3. A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hi, Second Edition, 1984