Chapter 1

Introduction to Probability and Set Theory

1.1 What is probability?

As we begin, we will make an attempt to answer some questions, such as, 'What is probability?' and 'As engineers, why do we need to study it?'. Be it in our daily lives or in the engineering area of interest, we encounter countless circumstances where the outcome of a situation cannot be deterministically predicted. The reasons for the unpredictability in a situation may lie in the nature of the matter under investigation or it may simply be due to lack of sufficient information about the matter. In either case, the uncertainty prevails. Despite the abundance of uncertainty in individual situations, the occurrences around us, technical or otherwise, are not totally chaotic, although it may seem so at times. 'In the long run' or 'on the average', there is a certain 'order' or 'law', often called statistical regularity, that reigns the domain of the unpredictable. Probability Theory is the branch of applied mathematics which expresses and organizes this statistical regularity observed in unpredictable occurrences. It provides the mechanism to describe and cope with the unpredictable (non-deterministic or random) situations and if possible use them to our advantage.

We encounter non-deterministic or random quantities and occurrences in virtually all areas of electrical and computer engineering as well as other branches of engineering. Hence, an adequate knowledge of probability and random processes has been an essential component of the electrical and computer engineering curriculum since 1960's. If anything, the reliance on probabilistic methods has been increasing over time in many areas, such as communication
systems, digital signal processing, control systems, signal and image processing, microwave systems, remote sensing, computer networks, performance evaluation and manufacturing. Hence, there should not be any doubt in the reader's mind regarding the importance of probability and random processes in electrical and computer engineering.

Despite the abundance of examples of the use of probability and random processes in different areas of electrical and computer engineering, a meaningful exposition of these applications requires a good understanding of the specific topic, which is generally not expected in a book/course like this one. Providing such an understanding within the context of probability would require too much "overhead" which cannot be afforded in terms of time and space available. Hence, in most probability books, as it is in this one, examples and problems are chosen from games of chance or layperson-type problems or simply as mathematical exercises. This is a difficulty which tends to undermine the motivation on the part of students, yet one which does not seem to have a simple solution. In order to apply probabilistic techniques to a specific engineering problem, one needs to extract the "probabilistic essence" of the problem, that is, to reduce it to a probability problem, solve it as such and carry the results back to the context.

Below is a list of phenomena where the outcome is unpredictable and therefore a probabilistic approach is necessary:

1. A coin is tossed and the outcome is observed.

2. A die is thrown and the outcome is observed.

3. A card is drawn from a deck of 52 cards and the outcome is observed.

4. A hand of 5 cards is dealt from a deck of 52 in the game of poker and it is observed whether or not the hand contains a "three of a kind" or "two pairs".

5. The status of a communication network is observed at a given time and it is noted whether or not there is a path connecting nodes A and B.

6. At the output of a radio receiver the signal \( y(t) = s(t) + n(t) \), where \( s(t) \) is the desired signal and \( n(t) \) is the random noise, is observed.

7. The performance of a digital communication system operating in a noisy environment is observed.
1.1. **WHAT IS PROBABILITY?**

![Diagram of a communication network]

**Figure 1.1:** A communication network

8. The number of users on a computer network $k(t)$ is observed.

9. The number of particles emitted from a radioactive material $\alpha(t)$ is observed.

We now describe in a little more detail some of the random situations described above and discuss some of the probabilistic concerns in these situations.

**A Communication Network.** The communication network consisting of many links, shown in Fig. 1.1, is to be investigated. Suppose we are given that at any given time the link $l_i$ is operational with probability $p_i$. Then we wonder: What is the probability that there is communication between nodes A and B? How is this probability affected if $l_7$ is removed? If another link were to be added to the network, where should it be placed to maximize the probability of communication between A and B? Or more generally, given a certain number of nodes and a certain number of links, how should the network be arranged so that the probability of communication between any two nodes is not less than a certain lower bound? As the problems of interest get more involved, naturally their solutions get more complicated.

**An Analog Communication System.** Suppose it is desired to transmit a signal $s(t)$ from point A to point B using a certain channel (transmission medium), e.g., the atmosphere. We often have to transform the signal $s(t)$ to make it suitable for transmission and then of course to transform it back to its original form. These two transformations are called modulation and demodulation. Fig. 1.2 shows the block diagram of such a communication system. Even if the modulator-demodulator pair works perfectly, corruptive noise is added to the signal during transmission through the channel. Unwanted noise is also added to the signal at the transmitter and the receiver. The noise $n(t)$ is of course unknown, hence is appropriately modeled as
random. However, certain things about the noise, called its statistical properties are usually known. So, the problem then is: knowing the statistical properties of the noise and also of the signal, to design the optimum modulator-demodulator pair such that the distortion $|s(t) - \hat{s}(t)|$ as minimized in some statistical sense. For example, in commercial radio broadcasting, we know that AM is more susceptible to noise than FM. In order to analyze and design such a system, one needs to know how to deal with random processes, such as $n(t)$.

This is a generic problem: we have $y(t) = x(t) + n(t)$, where $x(t)$ is the desired signal and $n(t)$ is the corruptive random noise. How should $y(t)$ be processed to obtain the best estimate/approximation of $x(t)$? This could be a problem in control or microwave or communication systems.

**A Digital Communication Problem.** Suppose a ternary communication system is transmitting the symbols $\{0, +1, -1\}$, but errors occur in receiving them as they were transmitted. A diagram of such a system is shown in Fig. 1.3. If we are given or are able to determine the probabilities that a 0 sent was received as a 0, or as a +1, or as a −1, and the same when a +1 or a −1 was sent, then can we compute the overall probability of error of this system? Or, given that a 0 is received, what is the probability that a 0 was sent? One might wonder, if one can improve the system so that the probability of accurate transmission of each symbol would be increased, or, if the probability of error can be decreased, by coding blocks of symbols into other symbols and decoding them at the receiver?

**Service Facility.** Consider a service facility consisting of a "buffer" of length $B$ (i.e., a waiting area) and $N$ servers. Suppose customers arrive in a random fashion, but on the average $\lambda$ customers per hour; also suppose that the service time is random, but on the average each server serves $\mu$ customers per hour. Fig. 1.4 depicts the schema for such a service facility.
Upon further specifying the nature of the randomness in the arrivals and the service time, one can hope to answer questions like: After this facility has been running for a while, what is the probability that at an arbitrary time, the facility is completely empty, or all servers are busy and the buffer is full, or a customer arriving receives service, i.e., does not find the servers and buffer full? What is the expected number of customers served per hour? Given that a customer receives service, what is the expected time he/she spends in the facility? Given a certain fee and pay schedule, what is the expected revenue of the facility per hour? If some further investment is to be made, which would be more profitable, expanding the buffer or increasing the number of servers?

There are three issues concerning unpredictable occurrences that we need to elaborate further before concluding this section:

A. Randomness. In the above paragraphs, several situations are described where the
outcomes cannot be deterministically predicted and therefore are characterized as random. In some of these cases the unpredictability is the result of a lack of sufficient information to predict the outcome or the physical situation being too complicated to analyze deterministically. For example, in the coin tossing or die throwing experiment, if we knew precisely the position and the velocity of the coin or the die as it leaves our hand and the characteristics of the material it lands on, by invoking sophisticated techniques of mechanics, theoretically we could calculate, that is, predict the outcome of this "random" experiment, as difficult as that may be. In some cases, however, the phenomenon is intrinsically random, such as in the case of the particles emitted from a radioactive material. In these cases, there is no known way of removing the unpredictability even if one is willing to do complicated computations for it. In many instances, one can argue one way or the other whether the situation is intrinsically random or specified as such for convenience, for example, the atmospheric or electronic noise in a communication system or the customers arriving in a service facility. There is no practical benefit in trying to resolve this philosophical question. Whether it is for convenience or by necessity, once a situation is characterized as random, then it is necessary to deal with it using probabilistic methods. If some information which can be characterized as deterministic is available, then such information can usually be incorporated into the probabilistic formulation to reduce the "degree of uncertainty" in the problem. Hence, we are not particularly concerned about the argument on what is intrinsically random and what is specified as random by necessity or for convenience.

B. Statistical Regularity. We mentioned earlier that, although individual occurrences may be unpredictable, on the average, random phenomena are ruled by an "order", often called statistical regularity. For example, in a coin tossing experiment, we cannot predict the outcome of a single toss, but if the coin is tossed 100 times and the ratio of tosses resulting in "heads" to the total number of tosses is formed, we would see that this ratio would likely be close to 0.5, although not necessarily so in each case. Specifically, the probability that the ratio (for 100 tosses) falls into the interval \([0.4, 0.6]\) is 0.964; if it is tossed 200 times, this probability is 0.996. Similarly, we cannot predict the time the next particle will be emitted from a radioactive material, but if we measure the time interval between 100 consecutive emissions and compute their average, we would see that that average interarrival period is quite stable from one 100 run to another. Probability theory is the mathematical model which describes this statistical regularity observed in random phenomenon and provides the means to manage it.
C. Scientific Investigation. The methodology called the scientific investigation used in
the study of all physical and other sciences, consists of three major steps: (1) The observation
of the phenomenon under study and the identification of any patterns in it; (2) The building of
a mathematical model describing the patterns observed; and (3) The verification of the results
obtained using the model with the phenomenon. The model has to be consistent in itself,
although it may involve some idealization and simplifying assumptions of the phenomenon it
represents. The goodness of the model depends on the extent of the simplifying assumptions
and it manifests itself in how well the model predictions match the actual phenomenon in the
third step of scientific investigation.

In electrical circuits, for example, we observe that the current that flows through a resistor
is proportional to the voltage applied across the resistor. That is the observation (step 1). The
Ohm's Law, \( I = \frac{V}{R} \), is the mathematical model describing this phenomenon (step 2). It is an
idealization, in that, there are instances where Ohm's Law may not hold exactly. In those cases,
one may need more sophisticated models to attain a better match with the actual situation.
One can then use "the model" (the Ohm's Law and other laws of circuit theory, in this case)
to analyze and solve complicated circuits. The results can then be compared with the actual
measurements to verify the model as well as its implementation (step 3). Of course, any
inaccuracies in the input data, such as in the values of the resistances or voltages applied, will
result in inaccuracies in the output values, which is not (necessarily) due to a deficiency of the
model, circuit theory.

Similarly, we observe, as did Sir Isaac Newton, that objects around us fall when they are
dropped. Upon observing the pattern in falling objects, Newton provided the model called the
Theory of Gravitation, that describes this phenomenon. According to this model, on planet
Earth, a free falling object travels a distance of \( d = \frac{1}{2}gt^2 \) meters in \( t \) seconds, where \( g \) is the
gravitational acceleration in \( \text{meters/seconds}^2 \). Using the laws of this mathematical model, one
can calculate the expected behavior of flying objects, satellites and the like. One can then
verify whether or not the actual behavior agrees with that predicted by the model. A common
characteristic of these two examples of the scientific investigation is that all quantities involved
in these two phenomena and the corresponding models are physical quantities not requiring
any abstraction.

The third example of scientific investigation we will discuss is the random phenomena.
We observe the random phenomena around us and see the pattern, called statistical regularity, that appears to govern these phenomena. Probability theory is the mathematical model that represents statistical regularity. It is a consistent set of axioms, definitions and theorems which describes the interrelations of probabilities and other probabilistic quantities. The results obtained through the model can then be verified. What distinguishes this case from the previous two is that in this case the quantities of interest, i.e., the probabilities, are non-physical quantities and are not easily measurable. This fact does not diminish the significance of the model or alter the methodology. However, it makes the specification of the probabilities that are input to the model and the verification of the probabilities that are output by the model somewhat difficult. (Here, we are using the term “probability” loosely; “probabilistic quantities” would perhaps be a more appropriate term.) We cannot simply go in with a “probability meter” and measure the probabilities of various events to be used as input or for verification with the output probabilities. In the next section, we discuss some methods for probability determination, but the general problem falls in the domain of statistics rather than probability theory. Probability theory provides the mechanism for obtaining the output probabilities from the input probabilities. It is not so much concerned with how the input probabilities are obtained or their accuracy or how the output probabilities are verified, except for the fact that, through the result known as the weak law of large numbers, probability theory provides a method for determining or verifying probabilities by performing the random experiment many times.

1.2 Historical development of probability

A brief discussion of the historical development of probability theory is in order here, particularly because, as individuals (without formal training in it) our notions of probability follow a path similar to that of the historical development of probability theory. Moreover, knowledge of this development contributes to a better understanding of the present day probability theory. The interest in probability theory goes back to the 17th century when many scientists of the era were trying to determine the odds in some games of chance. Thus, interest in the games of chance was the main motivation in the development of probability. Renowned scientists of the 17th and 18th centuries, such as Pascal, Fermat, Huygens, Bernoulli, De Moivre and later Laplace, made significant contributions to probability. The approach to probability they helped develop is now called the classical approach.
1.2. HISTORICAL DEVELOPMENT OF PROBABILITY

A. Classical Approach. This approach relies on two major assumptions on the random phenomenon, also called experiment, under investigation: that the experiment has finite number of outcomes and that they are all equally likely. According to this approach, the probability of an event \( A \) is defined as

\[
P(A) = \frac{N_A}{N}
\]

where \( N \) denotes the total number of outcomes in the experiment and \( N_A \) denotes the number of outcomes that are favorable to event \( A \). By "an outcome being favorable to an event", we mean, "the showing of that outcome as a result of the experiment implies the occurrence of the event". For example, if the random experiment is drawing a card from a deck and event \( A = \text{drawing of an ace} \), then \( P(A) = \frac{4}{52} \), since there are a total of 52 cards in the deck and 4 of them are aces.

This approach is a very reasonable one; indeed, our intuitive thinking on probability frequently follows precisely this line of reasoning. It is particularly useful when one can readily make the argument that there is no reason for any of the outcomes to be more or less probable than any other, hence they have to be equally likely and that there are a finite number of outcomes. This approach can be characterized as "determining probabilities by pure reasoning", as opposed to experimentation. However, one may run into some difficulty in using this method, if one is not careful in identifying the equally likely outcomes properly. For example, in the random experiment "two dice are tossed and the sum of the two faces is observed", one may choose the outcomes as \( \{2, 3, 4, \ldots, 12\} \) and one may think that they should be equally likely and each has probability \( \frac{1}{11} \). But as any student of craps knows, a sum of 2 and a sum of 7 are certainly not equally likely. However, if one identifies the outcomes of this random experiment as \( \{(1,1), (1,2), \ldots, (6,6)\} \), such that they are equally likely and that there are 36 of them, then using the classical approach one would find the correct probabilities as \( P(2) = \frac{1}{36} \) and \( P(7) = \frac{6}{36} = \frac{1}{6} \).

Despite the reasonableness and the intuitive appeal of the classical approach, due to its reliance on the two assumptions, it is too restrictive, hence unsatisfactory as a general theory of probability. First, there are many random phenomenon which have infinite outcomes; second, even those with finite number of outcomes, in many instances, the outcomes are not equally likely, or we cannot readily assume that they are. Hence, the classical approach fails to provide a general theoretical framework for a very large class of random phenomena. Recognition of this
deficiency of the classical approach led to the development of the alternative called the *relative frequency approach*. This corresponds to one of the two ways our intuition on probability works.

**B. Relative Frequency Approach.** Developed in the early part of this century by Von Mises, the *relative frequency approach* is an *empirical* approach which attempts to exploit the statistical regularity observed in random phenomenon. According to this approach, to determine the probability of an event $A$, the random experiment is performed $n$ times and the number of times event $A$ occurs, $n_A$, is noted. Then, the probability of event $A$ is defined as

$$P(A) = \frac{n_A}{n}. \quad (1.2)$$

This ratio defined as the probability of $A$ is called the *relative frequency*. For example, if the random experiment is that “a loaded (or not a perfect cube) die is thrown and the outcome is observed”. Clearly, the classical approach cannot be employed in this case, since the outcomes cannot be assumed as equally likely. If we want to “determine” the probability of a certain face showing, say “six”, most would agree that indeed the reasonable thing to do would be to throw the die $n$ times, say $n = 100$, observe the number of times the face “six” shows, denote it as $n_6$, and form the relative frequency. We would tend to believe that the probability of a “six” showing is (somehow) represented by the corresponding relative frequency. This certainly is a reasonable belief, although at some point it needs to be substantiated. Then some may feel that the number of throws, 100, is not large enough to feel confident with the result and may be willing to do more experiments, say throw the die 500 or 1,000 times, and form the corresponding relative frequency. As the number of throws is increased, the relative frequency will vary, possibly significantly. Our intuition suggests that we would feel more confident about the “accuracy” of the result, that is, the relative frequency representing the true probability, obtained with say 1,000 throws than with 100 throws. This, too, is a reasonable intuitive belief, which will be shown to have theoretical justification.

Despite the reasonableness and the intuitive appeal of the relative frequency approach to probability, the discussion in the previous paragraph points out to the difficulty with it. Namely, ‘What should $n$ be?’ As noted, the larger $n$ is, the more confident we are with the result, but how large is large enough? Clearly there is no unique $n$ with which we can definitively say, now that is the definition. To remedy this ambiguity about $n$, we may be tempted to define the
1.2. HISTORICAL DEVELOPMENT OF PROBABILITY

probability as the limit of the relative frequency as \( n \to \infty \), that is,

\[
P(A) = \lim_{n \to \infty} \frac{n_A}{n},
\]

Unfortunately, this too does not resolve the difficulty, because except for an intuitive hunch, there is no guarantee that this limit exists. The ratio does actually converge to a limit, but there is no way of proving that without a consistent probability theory structure in place. Another difficulty with the relative frequency approach is that in many instances it is impossible or impractical to repeat a random experiment even a few times. Therefore, despite its strong intuitive appeal, the relative frequency approach is also unsatisfactory as a mathematically complete theory to describe the statistical regularity observed in random phenomenon. This leads us to the third approach in the development of probability theory, called the axiomatic approach.

C. Axiomatic Approach. This is a consistent mathematical model complete with axioms, definitions and theorems, which describes statistical regularity and allows us to compute probabilities and related quantities. The major contribution to its development was due to Kolmogorov in 1930's. At the present time, it is the only universally accepted, modern approach to probability. In other words, it is not one of three alternate approaches in studying probability; rather, it is the last stage in the development of probability.

Fortunately, selecting the axiomatic approach as the approach of choice does not mean that all the intuitive and other benefits of the classical and the relative frequency approaches have to be discarded. On the contrary, both these approaches continue to be very useful intuitive tools. Moreover, the classical and relative frequency approaches find complete explanations in the axiomatic approach. The classical approach is simply a restriction of the axiomatic approach to a very special class of random experiments, namely those with finite and equally likely outcomes. The relative frequency approach also finds a complete theoretical explanation within the axiomatic approach through the result called the weak law of large numbers. Thus, these two former approaches are not merely intuitive hunches to be abandoned; rather, they are parts of a more comprehensive theory called the axiomatic approach to probability. Here, we do not need to elaborate on the axiomatic approach any more, since the whole course/book is a detailed exposition of it.
1.3 Basic Probabilities

As mentioned above, the axiomatic theory of probability is a mathematical model devised to describe random phenomena. Therefore, as in mathematical models describing other phenomena, certain basic quantities have to be provided as input, in order to use the model and compute other quantities. For example, in circuit theory, to use Ohm's Law to compute \( I \), you need to know the values of \( V \) and \( R \). Same is true in probability theory. We need to know the basic probabilities to be able to use the model and compute other probabilities and related quantities. The specification of the basic probabilities have to be complete and consistent. What we mean by complete and consistent will be explained in the next chapter.

So the question is: Where do the basic probabilities come from? How are they determined? They are given to us somehow, possibly determined by pure reasoning as in the classical approach or by experimentation as in the relative frequency approach or by subjective judgment as in “guessing”. How they are determined does not so much concern the inner workings of the Probability Theory. As long as they are specified completely and consistently, the theory/model will go to work using them, much like when \( R = 1 \, k\Omega \) is given in a circuit, we don’t worry about where it came from, but simply use it in the formulas. Naturally, for the sake of the “accuracy” of the results output by the model, we are concerned with the “accuracy” of the basic probabilities input to the model. Therefore, in each instance, we use all means at our disposal to get an accurate (also complete and consistent) specification of the basic probabilities. We note again that the classical and relative frequency approaches are two invaluable methods for obtaining these specifications.

Subjective Statements. In our daily lives, we often make subjective statements like: “With 80% probability, New England Patriots will win the next Super Bowl”. No one can prove or disprove such a statement. It is an experiment that cannot be tried more than once. Hence, technically speaking, it is not proper to talk about such a probability. It is a statement that reflects one’s biases, subjective judgment, lack of information and so on. Similarly, a statement like: “There is 70% chance that it will rain today.”, as a probabilistic statement, can only be interpreted to mean: “If we could have the exact meteorological conditions as today (as we know them and as much as we know them) many many times, it would rain in 70% of those days.” We will not be concerned with such statements and random experiments, except perhaps
1.4. SET THEORY

in specifying the basic probabilities sometimes.

Characteristics of Random Phenomena. We will assume that the random phenomena that we investigate have the following characteristics: First, the random experiment must be repeatable, at least conceptually. Second, the outcome in each trial may be unpredictable, but we should be able to specify all outcomes of the experiment. This may involve some idealization of the actual physical experiment by discarding some of the physically possible but probabilistically uninteresting outcomes from the set. The set $S$ of all possible outcomes of the random experiment, called the sample space, corresponds to the universe set in set theory. Certain (sometimes all) subsets of $S$ represent "events" of the random experiment and set operations on these subsets produce other subsets which are also events. Hence, set theory is a necessary tool for studying probability theory.

1.4 Set Theory

Events in probability theory are represented by sets; and set operations on events produce other events of interest. Hence, set theory is an essential tool in studying probability. Therefore, in this section, we will review the basic concepts of set theory. We start out with some definitions.

Definition 1.1 A set is a collection of objects. Objects that make up a set are called the elements or members of that set.

For example, if $A = \{a, b, c\}$, then $a$ is an element of $A$, denoted by $a \in A$, but $d$ is not an element of $A$, which is denoted by $d \notin A$. In general, sets are specified in two ways:

1. By listing or tabulating all of its elements. For example, $A = \{a, b, c\}$, $B = \{0, 1\}$, and $C = \{1, 2, 3, 4, \ldots \}$.

2. By stating the common (hence, defining) property of all of its elements. For example, $D = \{x \in \mathbb{R} \mid 0 \leq x < 4\}$, which in short can be written as $\{0 \leq x < 4\}$ or as $[0, 4)$, and $E = \{k \in \mathbb{I} \mid 0 < k\}$, where $\mathbb{R}$ denotes the real numbers and $\mathbb{I}$ denotes the integers.

As is customary, we denote intervals of the real line by $[a, b]$ (or $[a, b), (a, b], (a, b)$), to represent the real numbers between $a$ and $b$, where each endpoint is included in the set if it is demarcated by a square bracket and not included if demarcated by a paranthesis.
CHAPTER 1. INTRODUCTION TO PROBABILITY AND SET THEORY

The order in which the elements of a set are written is of no consequence; writing of an element in a set more than once is not a proper practice. Given a set, it should be unambiguously determinable whether a certain object is an element in that set or not.

Definition 1.2 Two sets $A$ and $B$ are said to be equal, denoted as $A = B$, if and only if they contain exactly the same elements. If they are not equal, then they are distinct sets, which is expressed as $A \neq B$.

There are two special sets of interest:

1. The set that contains, as its elements, all objects of interest (in a certain context) is called the universe set, denoted here by $S$. [Despite Russell's Paradox, which proves by contradiction that nothing contains everything, the qualifier "all objects of interest in a certain context" allows us to define the universe set $S$.]

2. The set that contains no elements is called the empty set, the null set or the void set, denoted by $\emptyset$. For example, $A = \{x \in \mathbb{R} \mid x^2 + 1 = 0\} = \emptyset$.

Sometimes sets are classified according to the number of elements in them:

1. $\emptyset$, the null set, contains zero elements.

2. A set which contains one element, e.g., $\{a\}$, is called a singleton.

3. A set with finite number of elements, e.g., $\{a, b, c\}$, is called a finite set.

4. A set whose elements can be put in a one-to-one correspondence with the positive integers, e.g., $\{0, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, \ldots\}$, is called a countable or countably infinite set. Note that a finite set is also countable.

5. A set which cannot be put in a correspondence with the positive integers, e.g., $\{0 \leq x \leq 1\}$, is called an uncountable or uncountably infinite set.

It might require some work to establish certain sets as countable or uncountable. It is good exercise for example to show that the rational numbers are countable. Showing that irrational numbers or real numbers on any non-zero length interval are uncountable, is more involved.

Some definitions and facts on set relationships follow:
1.4. SET THEORY

Definition 1.3 If every element of set $A$ is also an element of set $B$, then we say that $A$ is contained in $B$ or that $A$ is a subset of $B$, denoted by $A \subseteq B$ or by $B \supseteq A$.

To prove $A \subseteq B$, one needs to show that every element of $A$ is an element of $B$, or that $x$, any (arbitrary) element of $A$, is an element of $B$. The subset relationship is sometimes denoted by the symbol "\subset". Here we reserve that symbol to denote the proper subset relationship which is defined next.

Definition 1.4 If $A \subseteq B$ (i.e., every element of $A$ is also an element of $B$) and at least one element of $B$ is not in $A$, then we say that $A$ is a proper subset of $B$, denoted by $A \subset B$ or by $B \supset A$.

By proper subset, we are precluding the possibility of equality. So, $A \subset B$ is equivalent to $A \subseteq B$ and $A \neq B$.

The following theorem is useful in proving the equality of sets:

Theorem 1.1 $A = B$ if and only if $A \subseteq B$ and $B \subseteq A$.

So, frequently, $A = B$ is proved by showing

1. $x \in A$ implies $x \in B$, and
2. $x \in B$ implies $x \in A$.

Some facts on the subset relation and the special sets, $S$ and $\phi$ are: For any set $A$, $\phi \subseteq A$, $A \subseteq A$ and $A \subseteq S$ are always true.

It is sometimes convenient to represent sets schematically. Such a representation, called the Venn Diagram, is often helpful in obtaining relationships between sets. (Venn diagram representations are generally not accepted as formal proofs of set relationships, although they are useful in formulating or verifying formal proofs.) See Fig.1.5.a for the Venn diagram depiction of $A \subseteq B \subseteq S$.

Now, we define some set operations.
Figure 1.5: Venn diagram representation of the subset relation and the union operation.

Definition 1.5 The union of any two sets $A$ and $B$ is the set that contains elements that are in $A$ or in $B$, or in both. It is denoted by $A \cup B = \{x \mid x \in A \text{ or } x \in B\}$.

For example, if $A = \{a, b, c\}$, $B = \{a, b\}$ and $C = \{c, d, e\}$, then $A \cup B = \{a, b, c\}$, $A \cup C = \{a, b, c, d, e\}$ and $B \cup C = \{a, b, c, d, e\}$. Following are some facts regarding the union operation and the special sets $S$ and $\phi$: For any set $A$, $A \cup \phi = A$, $A \cup A = A$, and $A \cup S = S$. The Venn diagram depiction of $A \cup B$ is given in Fig. 1.5.b. The union operation corresponds to the logical or (not exclusive or) operation.

Definition 1.6 The intersection of any two sets $A$ and $B$ is the set that contains elements that are in $A$ and in $B$. It is denoted by $A \cap B = \{x \mid x \in A \text{ and } x \in B\}$. The intersection of two sets is sometimes denoted by the juxtaposition of the set symbols, such as $AB$.

For example, if $A = \{a, b, c\}$, $B = \{c, d, e\}$ and $C = \{f\}$, then $A \cap B = \{c\}$ and $A \cap C = \phi$. Fig. 1.6.a depicts the Venn diagram of the intersection of two sets. For any set $A$, the following are true: $A \cap \phi = \phi$, $A \cap A = A$, and $A \cap S = A$. The intersection operation corresponds to the logical and operation.

Definition 1.7 If $A \cap B = \phi$, then $A$ and $B$ are called disjoint or mutually exclusive sets.

More set operations follow.

Definition 1.8 Let $S$ be the universe set and $A$ any subset of $S$. The complement of $A$ is the set of all elements (in $S$) that are not in $A$. It is denoted by $\overline{A} = \{x \in S \mid x \notin A\}$. The complement of a set is sometimes denoted by $A'$ or $A^c$. 
1.4. SET THEORY

Figure 1.6: Venn diagram representation of the intersection and complementation operations.

Note that, unlike the union and intersection operations, the complement set $\overline{A}$ is defined in reference to the universe set $S$. Hence, to determine $\overline{A}$, one needs to know the universe set $S$. For example, if $S = \{a, b, c, d\}$ and $A = \{a, b, c\}$, then $\overline{A} = \{d\}$. Venn diagram depiction of the complement set is given in Fig. 1.6.b. The following are some set equations involving the complement operation: $\overline{\emptyset} = S$, $\overline{S} = \emptyset$, $\overline{A} = A$, $A \cup \overline{A} = S$ and $A \cap \overline{A} = \emptyset$ (that is, $A$ and $\overline{A}$ are disjoint sets). The complementation operation corresponds to the logical negation operation.

**Definition 1.9** Set subtraction. $A - B$ is the set of all elements that are in $A$ but not in $B$. That is, $A - B = \{x \mid x \in A \text{ and } x \notin B\}$.

Note that $A - B = A \cap \overline{B}$ and $\overline{A} = S - A$ are some set identities. Venn diagram representation of set subtraction is shown in Fig. 1.7. From the Venn diagram one can easily see that

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

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

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

Following are some facts (or theorems) involving set operations:

1. If $A \subseteq B$, then $A \cup B = B$ and $A \cap B = A$.

2. For any $A$ and $B$, $A \subseteq A \cup B$ and $A \cap B \subseteq A$.

3. If $A \subseteq B$ and $B \subseteq C$, then $A \subseteq C$. 
4. If \( B \subseteq C \), then \( A \cap B \subseteq A \cap C \).

5. If \( A \subseteq B \) and \( C \subseteq B \), then \( A \cup C \subseteq B \).

6. If \( A \subseteq C \) and \( B \subseteq D \), then \( A \cup B \subseteq C \cup D \).

7. If \( A \) and \( B \) are disjoint sets, then \( A \cap C \) and \( B \cap C \) are also disjoint, for any set \( C \).

To give an example of a formal proof, we will prove the fourth fact above:

**Example 1.1** Given \( B \subseteq C \), show that \( A \cap B \subseteq A \cap C \).

Since \( B \subseteq C \), any element of \( B \) is also an element of \( C \). Now take \( x \) an arbitrary element of \( A \cap B \), assuming for the moment that such an \( x \) exists. That means \( x \in A \) and \( x \in B \). By the initial premise, then \( x \in C \). Therefore, \( x \in A \) and \( x \in C \); hence \( x \in A \cap C \). If an \( x \in A \cap B \) does not exist, that is, \( A \cap B = \emptyset \), then \( A \cap B \subseteq A \cap C \) holds trivially.

The union and intersection operations obey the following laws:

**Commutativity Law**

\[
A \cup B = B \cup A,
\]

\[
A \cap B = B \cap A.
\]

**Distributivity Law**

\[
A \cap (B \cup C) = (A \cap B) \cup (A \cap C),
\]

\[
A \cup (B \cap C) = (A \cup B) \cap (A \cup C).
\]
1.4. SET THEORY

Associativity Law

\[ A \cup (B \cup C) = (A \cup B) \cup C = A \cup B \cup C, \]
\[ A \cap (B \cap C) = (A \cap B) \cap C = A \cap B \cap C. \]

Note that successive application of the commutativity, distributivity and associativity laws yields the finite and countable versions of these laws.

Suppose \( A_1, A_2, A_3, \ldots \) are a collection of sets, then \( \bigcup_{i=1}^{k} A_i \) is the (finite) union of \( A_i \) sets, and \( \bigcup_{i=1}^{\infty} A_i \) is the countable union of \( A_i \) sets. Similarly, \( \bigcap_{i=1}^{k} A_i \) and \( \bigcap_{i=1}^{\infty} A_i \) are the finite and countable intersections of \( A_i \) sets, respectively.

**Definition 1.10** If \( A \cup B = S \), then \( A \) and \( B \) are collectively exhaustive. Similarly, for finite or countably infinite unions of sets, if \( \bigcup_{i=1}^{k} A_i = S \), then \( \{ A_i \}_{i=1}^{k} \) is said to be collectively exhaustive; if \( \bigcup_{i=1}^{\infty} A_i = S \), then \( \{ A_i \}_{i=1}^{\infty} \) is collectively exhaustive.

**Definition 1.11** If a collection of sets, such as \( \{ A_i \}_{i=1}^{k} \) or \( \{ A_i \}_{i=1}^{\infty} \), is collectively exhaustive and sets in the collection are pairwise disjoint, then the collection of sets is called a partition of \( S \).

Note that elements of sets can be sets themselves. Such a set is called “a set of sets” or “a class of sets” or “a family of sets”.

**Definition 1.12** The set of all subsets of a set \( A \) is called the power set of \( A \), denoted by \( \mathcal{P}_A \).

For example, for \( A = \{ a, b, c \} \), the power set of \( A \) is:

\[ \mathcal{P}_A = \{ \emptyset, \{ a \}, \{ b \}, \{ c \}, \{ a, b \}, \{ a, c \}, \{ b, c \}, \{a,b,c\} \}. \]

It can be shown by mathematical induction that, the power set of a set with \( n \) elements, contains \( 2^n \) elements. \( \emptyset \) and the set itself are always elements of the power set.

There is a pair of set identities called the *De Morgan’s Laws* that facilitates the determining and proving of other set equalities:

\[ \overline{A \cup B} = \overline{A} \cap \overline{B}, \]  \hspace{2cm} (1.4)

\[ \overline{A \cap B} = \overline{A} \cup \overline{B}. \]  \hspace{2cm} (1.5)
To demonstrate a formal proof of a set equation, let us prove the first De Morgan's law: To prove \( \overline{A \cup B} = \overline{A} \cap \overline{B} \), by Theorem 1.1, it suffices to prove that the set on each side of the equation is a subset of the set on the other side. First, to show \( \overline{A \cup B} \subseteq \overline{A} \cap \overline{B} \). Let \( x \) be an arbitrary element in \( \overline{A \cup B} \). That means \( x \) is not in \( A \cup B \), which in turn means \( x \) is not in \( A \) and it is not in \( B \). Therefore, \( x \) is in \( \overline{A} \cap \overline{B} \). Thus \( \overline{A \cup B} \subseteq \overline{A} \cap \overline{B} \). If such an \( x \) in \( \overline{A \cup B} \) does not exist, that is, if \( \overline{A \cup B} = \phi \), then the same result follows immediately. This completes the first part of the proof. For the second part, we need to show \( \overline{A} \cap \overline{B} \subseteq \overline{A \cup B} \). Let \( x \) be an arbitrary element in \( \overline{A} \cap \overline{B} \). That means \( x \) is not in \( A \) and not in \( B \), which in turn means \( x \) is not in \( A \) or \( B \). Therefore, \( x \in \overline{A \cup B} \). Again the same is true if such an \( x \) does not exist. This completes the second part of the proof and hence the proof itself.

It follows from De Morgan's laws that \( A \cup B = \overline{\overline{A} \cap \overline{B}} \), which implies that the union operation "\( \cup \)" can be expressed in terms of the intersection and complementation operations. Similarly, it follows from De Morgan's laws that \( A \cap B = \overline{\overline{A} \cup \overline{B}} \), which implies that the intersection operation can be expressed in terms of the union and complementation operations. In other words, any expression involving set operations can be expressed in terms of the complementation and one of the union or intersection operations, although for convenience all three are used.

Successive application of the De Morgan's laws yields the finite and countable versions of the laws, which can be helpful in various probability calculations. For example,

\[
A_1 \cup A_2 \cup \cdots \cup A_k = \overline{\overline{A_1} \cap \overline{A_2} \cap \cdots \cap \overline{A_k}}.
\]

(1.6)

Despite the more complicated appearance, the expression in the RHS of this set equation may be easier to deal with.

We now define the Cartesian product of sets, which will be useful in studying Bernoulli trials.

Definition 1.13 The Cartesian product of two sets \( A \) and \( B \), denoted by \( A \times B \), is the set of all ordered pairs such that the first component of the pair is an element of \( A \) and the second component is an element of \( B \). In other words, \( A \times B = \{(a, b) \mid a \in A, b \in B\} \).

Some examples of the Cartesian product of sets are: If \( A = \{h, t\} \) and \( B = \{1, 2, 3, 4, 5, 6\} \), then \( A \times B = \{(h, 1), (h, 2), (h, 3), (h, 4), (h, 5), (h, 6), (t, 1), (t, 2), (t, 3), (t, 4), (t, 5), (t, 6)\} \) and
1.5. PROBLEMS

\[ A \times A = \{(h, h), (h, t), (t, h), (t, t)\} \]
Note that \( A \times B \neq B \times A \), unless \( A = B \). It is possible to define the Cartesian product of several sets. For example,

\[ A_1 \times A_2 \times \cdots \times A_n = \{(a_1, a_2, \ldots, a_n) \mid a_1 \in A_1, a_2 \in A_2, \ldots, a_n \in A_n\} \tag{1.7} \]
is the set of (ordered) \( n \)-tuples, such that the \( i \)th component is an element of the \( i \)th set \( A_i \), for \( i = 1, 2, \ldots, n \). A special case is when the sets in the Cartesian product are identical. For example, \( R \times R \times R \), also expressed as \( R^3 \), denotes the three-dimensional Euclidean space.

Finally, we state a result known as the duality principle, which yields set identities from other set identities: In a set identity, if we interchange \( \cup \) with \( \cap \), and \( S \) with \( \phi \), and vice versa, the identity is preserved. Note for example that, if the duality principle is applied on the first De Morgan's law, the second De Morgan's law is obtained.

1.5 Problems

1. Describe a few phenomena (other than those described in the text) where the outcomes are unpredictable. State some random quantities of interest in these phenomena. Can you say whether or not these phenomena are intrinsically random?

2. Describe a few random phenomena where you can identify a pattern that corresponds to statistical regularity.

3. Describe a few physical phenomena which are represented by mathematical models. For each, identify the mathematical model.

4. Describe a few random phenomena which can be adequately treated with (a) the classical approach, (b) the relative frequency approach.

5. By throwing a "fair" die 200 times, compute and plot the relative frequency of the face "six" showing for \( n = 10, 20, 30, \ldots, 200 \). Comment on how the relative frequency in your experiment varies with \( n \).

6. Give a few examples of countable sets.

7. Show that rational numbers are countable.

8. Depict the Venn diagram representation of the following sets: \( (A \cup B) \cap \overline{C} \), \( A \cup (B \cap \overline{C}) \), \( (A \cup \overline{B}) \cap C \), \( A \cup B = (A \cap B) \), and \( (A \cup B \cup C) = (A \cap B \cap C) \).

9. Write expressions for the sets shown in Fig. 1.8. Simplify the expressions as much as possible.

10. Draw a Venn diagram representation of four arbitrary sets \( A, B, C, D \) showing all possible intersections. Comment on the difficulty in doing this, if that is so.
11. For $A$, $B$, and $C$ three subsets of $S$, prove the following set equalities:
   a) $A - B = A - (A \cap B)$.
   b) $A - (B \cap C) = (A - B) \cup (A - C)$.
   c) $(A \cap B) \cup (\overline{A} \cap C) \cup (B \cap C) = (A \cap B) \cup (\overline{A} \cap C)$.

12. Prove the following facts:
   a) If $A \subseteq B$ and $C \subseteq B$, then $A \cup C \subseteq B$.
   b) If $A \subseteq C$ and $B \subseteq D$, then $A \cup B \subseteq C \cup D$.
   c) If $A$ and $B$ are disjoint sets, then so are $A \cap C$ and $B \cap C$, for any set $C$.

13. Prove your answers to the following.
   a) If set $A$ and $B$ are mutually exclusive and collectively exhaustive, then are $\overline{A}$ and $\overline{B}$ mutually exclusive?
   b) If sets $A$ and $B$ are mutually exclusive but not collectively exhaustive, then are $\overline{A}$ and $\overline{B}$ mutually exclusive? Are they collectively exhaustive?
   c) If sets $A$ and $B$ are collectively exhaustive, but not mutually exclusive, then are $\overline{A}$ and $\overline{B}$ mutually exclusive? Are they collectively exhaustive?

14. Give a Venn diagram representation and write down a set expression for the sets described below. Simplify the expressions as much as possible.
   a) Of the three sets $A$, $B$, and $C$, only $A$,
   b) At least one of $A$, $B$, or $C$,
   c) Exactly one of $A$, $B$, or $C$,
   d) At most one of $A$, $B$, or $C$,
   e) $A$, but not in $B$ or $C$,
   f) None of $A$, $B$, or $C$,
   g) $A$ or $B$, but not $C$,
   h) $A$, but neither $B$ nor $C$,
   i) Two or more of $A$, $B$, and $C$.

15. Give two examples of partitions of (a) the set of integers, and (b) the set of real numbers.
16. Write down the power set of the following sets, (a) $\phi$, the null set, and (b) $B = \{a, b, c, d\}$.

17. Using mathematical induction, prove that the power set of a set with $n$ elements has $2^n$ elements.

18. Give examples of the Cartesian product of two sets. Repeat for three sets.

19. For $A = \{h, l\}$ and $B = \{1, 2, 3, 4\}$, write down explicitly the Cartesian product set $A \times A \times B$.

20. (Larson) Express $\overline{A} \times \overline{B}$ in terms of the Cartesian products involving $A, \overline{A}, B, \overline{B}, S$ and union operations.