EEE 549- Probability and Stochastic Processes

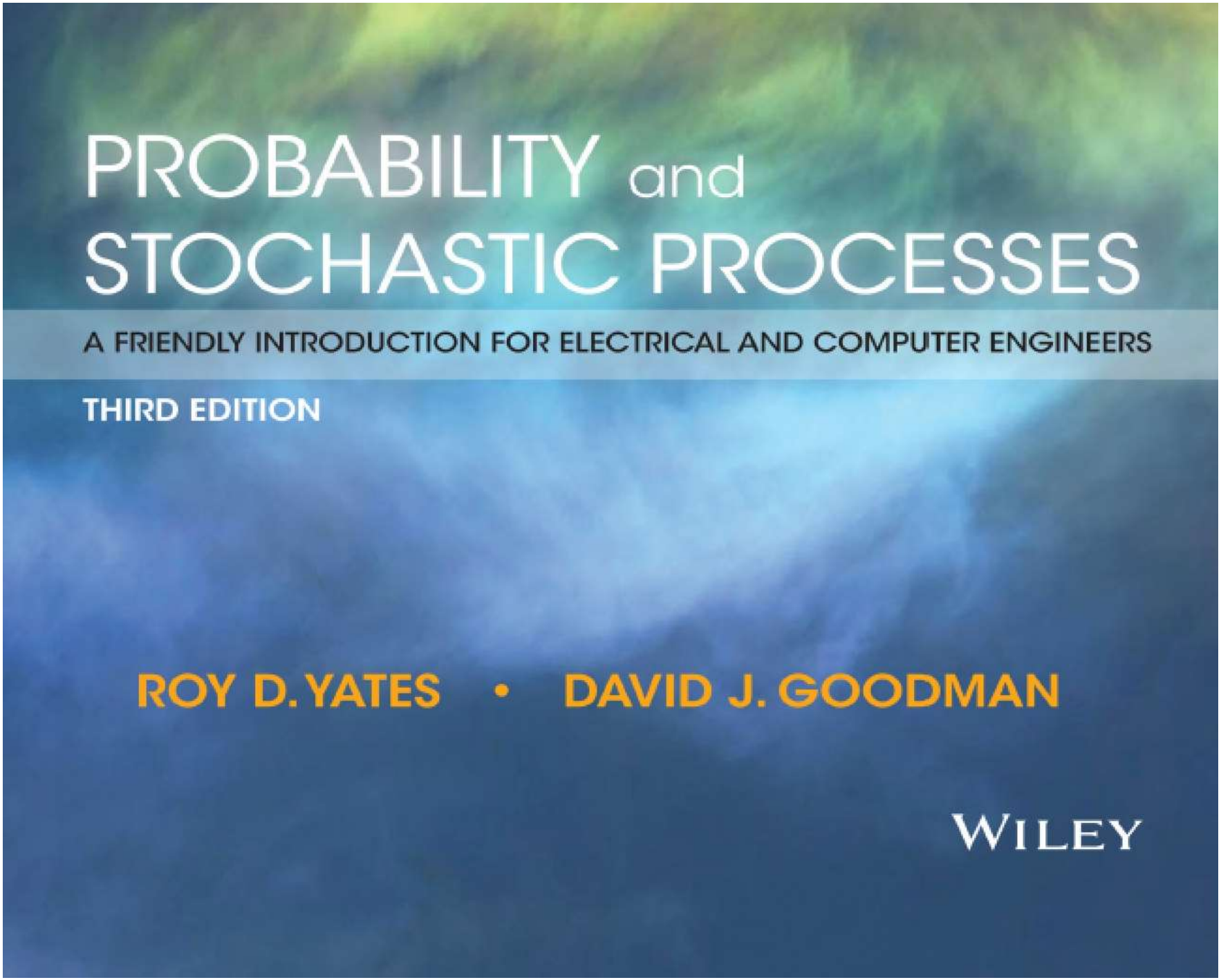
Prof.Dr. Bayram ÜNAL

Istanbul Gelisim University

Faculty of Engineering and Architecture,

Electrical and Electronics Engineering

2024-2025 Fall Semester


PROBABILITY and
STOCHASTIC PROCESSES
A FRIENDLY INTRODUCTION FOR ELECTRICAL AND COMPUTER ENGINEERS
THIRD EDITION
ROY D. YATES • DAVID J. GOODMAN
WILEY

Probability and Stochastic Processes

1) Experiments, Models, and Probabilities [3E]
2) Sequential Experiments [3E]
3) Discrete Random Variables [3E]
4) Continuous Random Variables [3E]
5) Multiple Random Variables [3E]
6) Random Vectors [3E8]
7) Sums of Random Variables [3E9]
8) The Sample Mean [3E10]
9) Hypothesis Testing [3E11]
10) Estimation of a Random Variable [3E12]
11) Stochastic Processes [3E13]
12) *Random Signal Processing* [2E11]
13) *Markov Chains* [2E12]
14) *Conditional Probability Models* [3E7]

Introduction to Probability

Probability is a measure quantifying the likelihood that an event will occur. It is expressed as a number between 0 and 1, where 0 indicates an impossible event, 1 indicates a certain event, and values in between represent degrees of likelihood. In other words, the probability of an event A is denoted by P(A), and it satisfies:

1. $0 \leq P(A) \leq 1$: The probability of any event is always between 0 and 1, inclusive.
2. $P(\text{Impossible event}) = 0$: If an event cannot occur, its probability is 0.
3. $P(\text{Certain event}) = 1$: If an event is certain to occur, its probability is 1.
4. $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$: The probability of the union of two events is the sum of their individual probabilities minus the probability of their intersection.
5. $P(\text{Not } A) = 1 - P(A)$: The probability of the complement of event A is 1 minus the probability of A.

Probabilities can be expressed as fractions, decimals, or percentages. For example, a probability of 0.25, 1/4, or 25% all represent the same likelihood.

Probability theory is widely used in various fields, including mathematics, statistics, economics, physics, and artificial intelligence, to model uncertainty and randomness in different phenomena.

Introduction to Stochastic Processes

A stochastic process, also known as a random process, is a mathematical concept used to model a sequence of random variables that evolve over time. It represents how a system or phenomenon changes or evolves in a probabilistic or random manner as time progresses. In essence, it's a collection of random variables indexed by time.

Key elements of a stochastic process include:

- **Time Parameter:** A stochastic process is observed or modelled over a specific time interval or index set. This parameter represents the progression of time.
- **Sample Paths:** Each random variable in the stochastic process corresponds to the outcome or state of the system at a specific point in time. These random variables collectively form a sequence, and each sequence is called a sample path.
- **Probability Distribution:** At each point in time, the random variable takes on values according to a probability distribution. The type of distribution and its parameters may change with time or remain constant, depending on the specific stochastic process.

Stochastic processes can be categorized into two different :

- **Discrete-Time Stochastic Processes:** In these processes, the time parameter takes on discrete values (e.g., integers). Examples include random walks, Poisson processes, and Markov chains.
- **Continuous-Time Stochastic Processes:** Here, the time parameter is continuous (e.g., real numbers). Examples include Brownian motion, continuous-time Markov processes, and diffusions.

Stochastic processes used in various fields and applications:

- **Finance:** Modeling stock prices, interest rates, and portfolio returns.
- **Physics:** Describing particle movements, diffusion of molecules, and radioactive decay.
- **Electronics :** Analyzing signal processing, control systems, and reliability analysis.
- **Economics:** Modeling economic variables like inflation rates and exchange rates.
- **Statistics:** Generating random samples for simulations and Monte Carlo methods.
- **Biostatistics:** Studying the spread of diseases and population dynamics.

Probability and Stochastic processes in electronics

Probability and stochastic processes play a significant role in electronics, helping engineers and researchers analyze and model various random or uncertain phenomena in electronic systems.

Probability in Electronics:

- **Failure Analysis:** Probability theory is used to assess the likelihood of electronic components or systems failing due to factors like manufacturing defects, wear and tear, or environmental stress. Reliability predictions are essential for designing electronics with acceptable failure rates.
- **Component Variability:** Electronic components (e.g., resistors, capacitors) often exhibit manufacturing variations. Probability distributions are used to describe the variability in component values and to ensure that designs can accommodate these variations.
- **Signal Integrity:** In high-speed digital and analog circuits, signal integrity issues can arise due to noise, crosstalk, and other factors. Probability theory is employed to analyze the probability of signal distortion or errors.
- **Monte Carlo Simulations:** Probability-based Monte Carlo simulations are used to analyse electronic circuits and systems under various operating conditions, taking into account parameter uncertainties and variability. This is particularly useful in integrated circuit design.

Stochastic Processes in Electronics-1:

- **Noise Modelling:** Stochastic processes are used to model different types of noise in electronic circuits, including thermal noise, shot noise, and flicker noise. Understanding the stochastic nature of noise is crucial for designing low-noise circuits and systems.
- **Random Telegraph Noise (RTN):** RTN in semiconductor devices, caused by the random capture and release of charge carriers by individual traps, is modelled as a stochastic process. Engineers use stochastic models to predict its impact on circuit performance and reliability.
- **Communication Systems:** In wireless communication systems, fading due to multipath propagation is often modelled using stochastic processes. This allows engineers to design communication systems that can cope with signal variations caused by fading.
- **Random Signals:** Stochastic processes are used to describe random signals encountered in various applications. For instance, in radar and sonar systems, stochastic models help analyze and detect targets in the presence of noise and interference.

Stochastic Processes in Electronics-2:

- **Reliability Analysis:** Stochastic processes are used to analyse the reliability of electronic systems over time. This includes modelling the aging and degradation of components and predicting failure rates.
- **Random Vibration and Mechanical Shock:** In electronic packaging and reliability testing, stochastic processes are used to model random vibrations and mechanical shocks that electronic components may experience during their lifetime.

Probability and Stochastic Processes

Chapter 1- Experiments, Models, and Probabilities

1.1 Set Theory

1.2 Applying Set Theory to Probability

1.3 Probability Axioms

1.4 Conditional Probability

1.5 Partitions and the Law of Total Probability

1.6 Independence

Experiments, Models, and Probabilities

1.1 Set Theory

People and machines use probability theory to describe uncertainty about whether an event occurs and to make decisions in the presence of this uncertainty.

Most people who study probability have already encountered set theory and are familiar with such terms as ***set, element, union, intersection***, and ***complement***.

A = {Gelişim University, İstanbul University, the planet Mercury}

B = {all Gelişim student who weigh more than 80kg}

In engineering, we often use mathematical rules for generating all of the elements of the set:

$C = \{x^2 | x = 1, 2, 3, 4, 5\}$ C = {1, 4, 9, 16, 25}

Some sets have an infinite number of elements: $D = \{x^2 | x = 1, 2, 3, ...\}$

We continue the sequence indefinitely, forming an infinite set containing all perfect squares except 0. The definition of D implies that $144 \in D$ and $10 \notin D$.

A is a subset of *B* if every member of *A* is also a member of *B*.

The symbol $\subset$ denotes a subset. Thus $A \subset B$ means that "the set *A* is a subset of the set *B*."

Using the definitions of sets *C* and *D,* we observe that $C \subset D$. If *I* = {all positive integers, negative integers, and 0} it follows that $C \subset I$, and $D \subset I$.

The definition of set equality, $A = B$, is $A = B$ if and only if $B \subset A$ and $A \subset B$.

A and *B* are identical if and only if every element of *A* is an element of *B* and every element of *B* is an element of *A*.

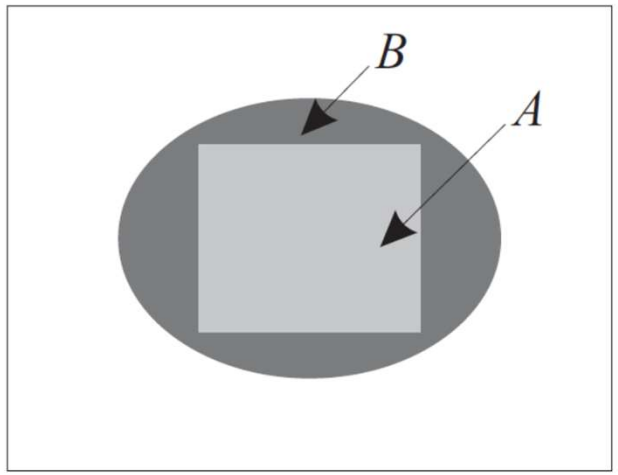
It is necessary to define a ***universal set*** **S**.

For example, the universal set for *A* could be *S* = {all universities in Türkiye, all planets}. The universal set for *C* could be $S = I = \{0, 1, 2, \ldots\}$.

By definition, every set is a subset of the universal set. That is, for any set *X*, $X \subset S$.

The *null set* may seem like it is not a set at all. By definition it has no elements. The notation for the null set is ∅.

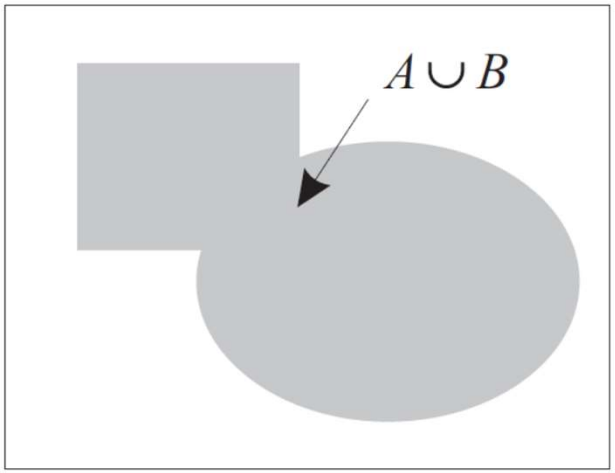
By definition ∅ is a subset of every set. For any set *A*, $\emptyset \subset A$.



It is customary to refer to Venn diagrams to display relationships among sets.
By convention, the region enclosed by the large rectangle is the universal set *S*.
Closed surfaces within this rectangle denote sets.
A Venn diagram depicting the relationship $A \subset B$ is shown on the left.

When we do set algebra, we form new sets from existing sets. There are three operations for doing this: ***Union, Intersection, and Complement***.

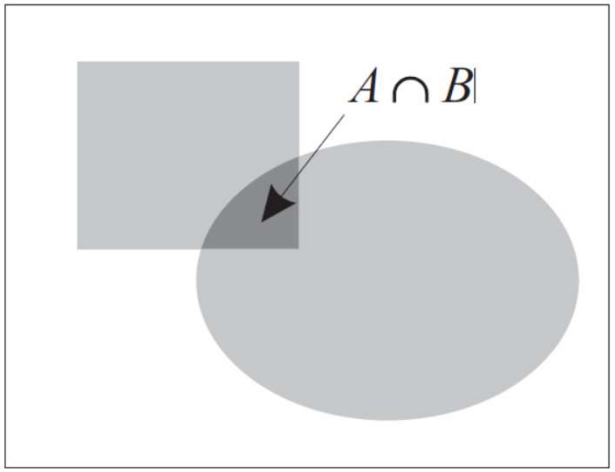
Union and intersection combine two existing sets to produce a third set. The complement operation forms a new set from one existing set.



The ***Union*** of sets *A* and *B* is the set of all elements that are either in *A* or in *B*, or in both. The union of *A* and *B* is denoted by $A \cup B$. In this Venn diagram, $A \cup B$ is the complete shaded area. Formally,

$x \in A \cup B$ if and only if $x \in A$ or $x \in B$.

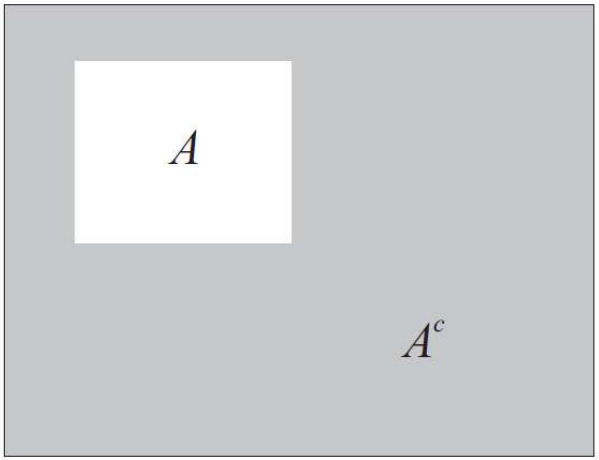
The set operation union corresponds to the logical "OR" operation.



The ***intersection*** of two sets *A* and *B* is the set of all elements that are contained both in *A* and *B*. The intersection is denoted by *A*∩*B*. Another notation for intersection is *AB*. Formally,

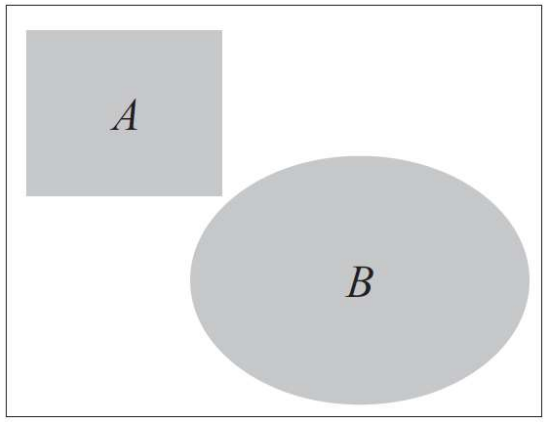
$x \in A \cap B$ if and only if $x \in A$ and $x \in B$.

The set operation intersection corresponds to the logical "AND" function.



The ***complement*** of a set *A*, denoted by *Ac*, is the set of all elements in *S* that are not in *A*. The complement of *S* is the null set ∅. Formally,

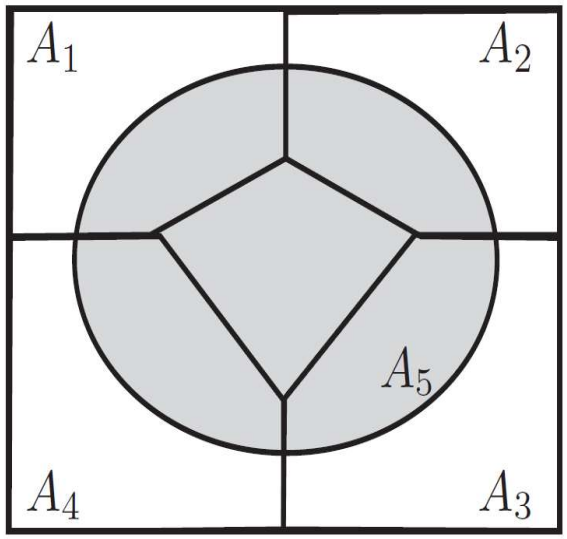
$x \in A^c$ if and only if $x \notin A$.



A collection of sets $A_1, \ldots, A_n$ is ***mutually exclusive*** if and only if

$$A_i \cap A_j = \emptyset, \; i \neq j$$

The word ***disjoint*** is sometimes used as a synonym for ***mutually exclusive***.



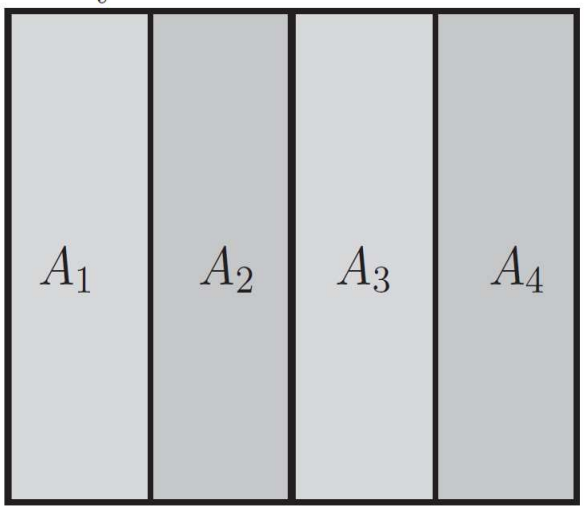
A collection of sets $A_1, \ldots, A_n$ is *collectively exhaustive* if and only if

$$A_1 \cup A_2 \cup \cdots \cup A_n = S.$$

In the definition of *collectively exhaustive*, we used the somewhat cumbersome notation $A_1 \cup A_2 \cup \cdots \cup A_n$ for the union of N sets. Just as $\sum_{i=1}^{n} x_i$ is a shorthand for $x_1 + x_2 + \cdots + x_n$, we will use a shorthand for unions and intersections of n sets:

$$\bigcup_{i=1}^{n} A_i = A_1 \cup A_2 \cup \cdots \cup A_n, \qquad \bigcap_{i=1}^{n} A_i = A_1 \cap A_2 \cap \cdots \cap A_n.$$

We will see that collections of sets that are both mutually exclusive and collectively exhaustive are sufficiently useful to merit a definition.



A collection of sets $A_1, \ldots, A_n$ is a *partition* if it is both mutually exclusive and collectively exhaustive.

From the definition of set operations, we can derive many important relationships between sets and other sets derived from them. One example is

$$A \cap B \subset A.$$

To prove that this is true, it is necessary to show that if $x \in A \cap B$, then it is also true that $x \in A$. A proof that two sets are equal, for example, $X = Y$, requires two separate proofs: $X \subset Y$ and $Y \subset X$. As we see in the following theorem, this can be complicated to show.

Theorem 1.1

De Morgan's law relates all three basic operations:

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

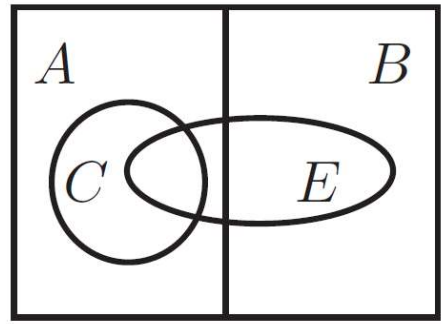
Proof There are two parts to the proof:

- To show $(A \cup B)^c \subset A^c \cap B^c$, suppose $x \in (A \cup B)^c$. That implies $x \notin A \cup B$. Hence, $x \notin A$ and $x \notin B$, which together imply $x \in A^c$ and $x \in B^c$. That is, $x \in A^c \cap B^c$.

Example 1.1

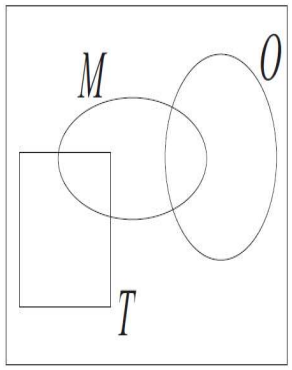
Phonesmart offers customers two kinds of smart phones, Apricot (A) and Banana (B). It is possible to buy a Banana phone with an optional external battery E. Apricot customers can buy a phone with an external battery (E) or an extra memory card (C) or both. Draw a Venn diagram that shows the relationship among the items A,B,C and E available to Phonesmart customers.

Since each phone is either Apricot or Banana, A and B form a partition. Since the external battery E is available for both kinds of phones, E intersects both A and B. However, since the memory card C is available only to Apricot customers, $C \subset A$. A Venn diagram representing these facts is shown on the right.



Quiz 1.1

Gerlandas offers customers two kinds of pizza crust, Tuscan (T) and Neapolitan (N). In addition, each pizza may have mushrooms (M) or onions (O) as described by the Venn diagram at right. For the sets specified below, shade the corresponding region of the Venn diagram.



(a) N

(b) $N \cup M$

(c) $N \cap M$

(d) $T^c \cap M^c$

1.1.1 For Gerlanda's pizza in Quiz 1.1, answer these questions:

(a) Are T and M mutually exclusive?

(b) Are R, T, and M collectively exhaustive?

(c) Are T and O mutually exclusive? State this condition in words.

(d) Does Gerlanda's make Tuscan pizzas with mushrooms and onions?

(e) Does Gerlanda's make regular pizzas that have neither mushrooms nor onions?

1.1.1 For Gerlanda's pizza in Quiz 1.1, answer these questions:

(a) Are T and M mutually exclusive?

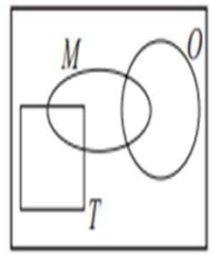
(b) Are R, T, and M collectively exhaustive?

(c) Are T and O mutually exclusive? State this condition in words.

(d) Does Gerlanda's make Tuscan pizzas with mushrooms and onions?

(e) Does Gerlanda's make regular pizzas that have neither mushrooms nor onions?

Problem 1.1.1 Solution Based on the Venn diagram



the answers are fairly straightforward:

(a) Since $T \cap M \neq \phi$, T and M are not mutually exclusive.

(b) Every pizza is either Regular (R), or Tuscan (T). Hence $R \cup T = S$ so that R and T are collectively exhaustive. Thus its also (trivially) true that $R \cup T \cup M = S$. That is, R, T and M are also collectively exhaustive.

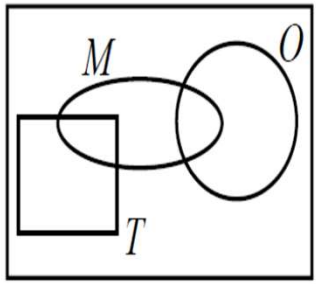
(c) From the Venn diagram, T and O are mutually exclusive. In words, this means that Tuscan pizzas never have onions or pizzas with onions are never Tuscan. As an aside, "Tuscan" is a fake pizza designation; one shouldn't conclude that people from Tuscany actually dislike onions.

(d) From the Venn diagram, $M \cap T$ and O are mutually exclusive. Thus Gerlanda's doesn't make Tuscan pizza with mushrooms and onions.

(e) Yes. In terms of the Venn diagram, these pizzas are in the set $(T \cup M \cup O)^c$.

Quiz 1.1.

A slice of pizza sold by Gerlanda's Pizza is either regular (R) or Sicilian (T as in Thick) In addition, each slice may have mushrooms (M) or onions (O) as described by the Venn diagram at right. For the sets specified in parts (a)–(g) below, shade the corresponding region of the Venn diagram.



(a) R

(b) $M \cup O$

(c) $M \cap O$

(d) $R \cup M$

(e) $R \cap M$

(f) $T^c - M$

(g) $M - T^c$

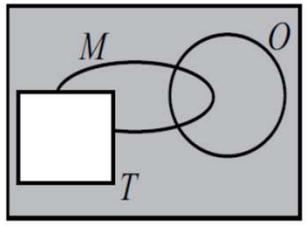
(h) Are T and M mutually exclusive?

(i) Are R, T, and M collectively exhaustive?

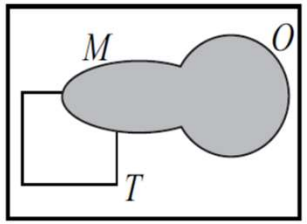
(j) Are T and O mutually exclusive? State this condition in words.

Quiz 1.1

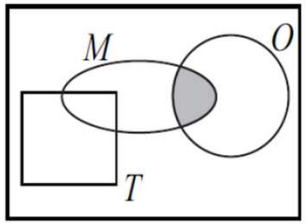
In the Venn diagrams for parts (a)-(g) below, the shaded area represents the indicated set.



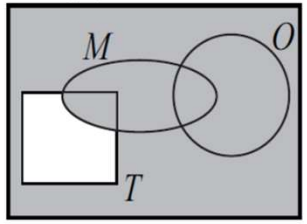
(a) $R = T^c$



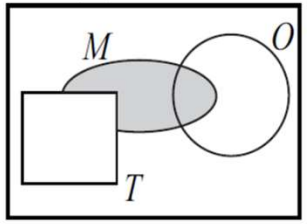
(b) $M \cup O$



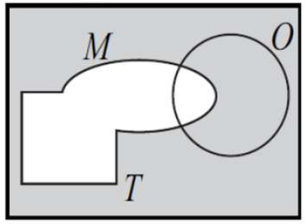
(c) $M \cap O$



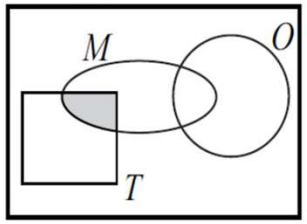
(d) $R \cup M$



(e) $R \cap M$



(f) $T^c - M$



(g) $M - T^c$

The remaining parts are

(h) T and M are not mutually exclusive since $T \cap M \neq \phi$.

(i) Since $R = T^c$, $R \cup T \cup M = T^c \cup T \cup M = S$. Thus R, T and M are collectively exhaustive.

(j) From the Venn diagram, $T \cap O = \phi$ so T and O are mutually exclusive.

1.2 Applying Set Theory to Probability

Probability is based on a repeatable experiment that consists of a procedure and observations. An *outcome* is an observation. An *event* is a set of outcomes.

Probability is a number that describes a set. The higher the number, the more probability there is.

Probability is like a quantity that measures a physical phenomenon; for example, a weight or a temperature. However, it is not necessary to think about probability in physical terms.

The basic model is a repeatable *experiment*. An experiment consists of a *procedure* and *observations*.

There is uncertainty in what will be observed; otherwise, performing the experiment would be unnecessary.

Some examples of experiments include

- ❑ Flip a coin. Did it land with heads or tails facing up?
- ❑ Walk to a bus stop. How long do you wait for the arrival of a bus?
- ❑ Give a lecture. How many students are seated in the fourth row?

- ❑ Transmit one of a collection of waveforms over a channel. What waveform arrives at the receiver?
- ❑ Transmit one of a collection of waveforms over a channel. Which waveform does the receiver identify as the transmitted waveform?

For the most part, we will analyze *models* of actual physical experiments. We create models because real experiments generally are too complicated to analyze.

For example, to describe *all* of the factors affecting your waiting time at a bus stop, you may consider

- ❖ The time of day. (Is it rush hour?)
- ❖ The speed of each car that passed by while you waited.
- ❖ The weight, horsepower, and gear ratios of each kind of bus used by the bus company.
- ❖ The psychological profile and work schedule of each bus driver. (Some drivers drive faster than others.)
- ❖ The status of all road construction within 100 miles of the bus stop.

It should be apparent that it would be difficult to analyze the effect of each of these factors on the likelihood that you will wait less than five minutes for a bus.

Consequently, it is necessary to study a *model* of the experiment that captures the important part of the actual physical experiment.

Since we will focus on the model of the experiment almost exclusively, we often will use the word *experiment* to refer to the model of an experiment.

Example 1.2

An experiment consists of the following procedure, observation, and model:

- **Procedure:** Monitor activity at a Phonesmart store.
- **Observation:** Observe which type of phone (Apricot or Banana) the next customer purchases.
- **Model:** Apricots and Bananas are equally likely. The result of each purchase is unrelated to the results of previous purchases.

An experiment consists of both a ***procedure*** and ***observations***.

It is important to understand that two experiments with the same procedure but with different observations are different experiments.

For example, consider these two experiments:

Example 1.3
Monitor the Teknosa store until three customers purchase phones. Observe the sequence of Iphones and Samsung.

Example 1.4
Monitor the Teknosa store until three customers purchase phones. Observe the number of Iphones.

These two experiments have the same procedure: monitor the Teknosa store until three customers purchase phones. They are different experiments because they require different observations.
We will describe models of experiments in terms of a set of possible experimental outcomes.
In the context of probability, we give precise meaning to the word ***outcome***.

Definition 1.1 Outcome

*An **outcome** of an experiment is any possible observation of that experiment.*

- Implicit in the definition of an outcome is the notion that each outcome is distinguishable from every other outcome.
- As a result, we define the universal set of all possible outcomes. In probability terms, we call this universal set the *sample space*.

Definition 1.2 Sample Space

*The **sample space** of an experiment is the finest-grain, mutually exclusive, collectively exhaustive set of all possible outcomes.*

- The *finest-grain* property simply means that all possible distinguishable outcomes are identified separately.
- The requirement that outcomes be mutually exclusive says that if one outcome occurs, then no other outcome also occurs.
- For the set of outcomes to be collectively exhaustive, every outcome of the experiment must be in the sample space.

Example 1.5

- The sample space in Example 1.2 is $S = \{a, b\}$ where a is the outcome "Apricot sold," and b is the outcome "Banana sold."
- The sample space in Example 1.3 is

$$S = \{aaa, aab, aba, abb, baa, bab, bba, bbb\}$$

- The sample space in Example 1.4 is $S = \{0, 1, 2, 3\}$.

Example 1.6

Manufacture an integrated circuit and test it to determine whether it meets quality objectives. The possible outcomes are "accepted" (a) and "rejected" (r). The sample space is $S = \{a, r\}$.

Set Algebra	Probability
Set	Event
Universal set	Sample space
Element	Outcome

The terminology of set theory and probability.

In common speech, an event is something that occurs.

In an experiment, we may say that an event occurs when a certain phenomenon is observed.

> To define an event mathematically, we must identify *all* outcomes for which the phenomenon is observed. That is, for each outcome, either the particular event occurs or it does not.

In probability terms, we define an event in terms of the outcomes in the sample space.

Definition 1.3 Event

*An **event** is a set of outcomes of an experiment.*

Set Algebra	Probability
Set	Event
Universal set	Sample space
Element	Outcome

Defining the sample space and its outcomes are key elements of the solution of any probability problem.

A probability problem arises from some practical situation that can be modelled as an experiment.

It is necessary to define the experiment carefully and then derive the sample space.

Example 1.7

Suppose we roll a six-sided die and observe the number of dots on the side facing upwards. We can label these outcomes $i = 1, \ldots, 6$ where i denotes the outcome that i dots appear on the up face. The sample space is $S = \{1, 2, \ldots, 6\}$. Each subset of S is an event. Examples of events are

- The event $E_1 = \{\text{Roll 4 or higher}\} = \{4, 5, 6\}$.
- The event $E_2 = \{\text{The roll is even}\} = \{2, 4, 6\}$.
- $E_3 = \{\text{The roll is the square of an integer}\} = \{1, 4\}$.

Example 1.8

Observe the number of minutes a customer spends in the Teknosa store. An outcome T is a nonnegative real number. The sample space is $S = \{T|T \geq 0\}$. The event "the customer stays longer than five minutes is $\{T|T > 5\}$.

Example 1.9

Monitor three customers in the Teknosa store. Classify the behavior as buying (b) if a customer purchases a smartphone. Otherwise the behavior is no purchase (n). An outcome of the experiment is a sequence of three customer decisions. We can denote each outcome by a three-letter word such as bnb indicating that the first and third customers buy a phone and the second customer does not. We denote the event that customer i buys a phone by B_i and the event customer i does not buy a phone by N_i. The event $B_2 = \{nbn, nbb, bbn, bbb\}$. We can also express an outcome as an intersection of events B_i and N_j. For example the outcome $bnb = B_1 N_2 B_3$.

Quiz 1.2

Monitor three consecutive packets going through a Internet router. Based on the packet header, each packet can be classified as either video (v) if it was sent from a Youtube server or as ordinary data (d). Your observation is a sequence of three letters (each letter is either v or d). For example, two video packets followed by one data packet corresponds to vvd. Write the elements of the following sets:

$A_1 = \{\text{second packet is video}\}$, $B_1 = \{\text{second packet is data}\}$,

$A_2 = \{\text{all packets are the same}\}$, $B_2 = \{\text{video and data alternate}\}$,

$A_3 = \{\text{one or more video packets}\}$, $B_3 = \{\text{two or more data packets}\}$.

For each pair of events A_1 and B_1, A_2 and B_2, and so on, identify whether the pair of events is either mutually exclusive or collectively exhaustive or both.

1.3 Probability Axioms

A probability model assigns a number between 0 and 1 to every event.

The probability of the ***union*** of *mutually exclusive* events is the sum of the probabilities of the events in the union.

Our model of an experiment consists of a procedure and observations.

- This leads to a set-theory representation with a **sample space** (universal set *S*), **outcomes** (*s that are elements of S*), and **events** (*A that are sets of elements*).
- To complete the model, we assign a probability P[*A*] to every event, *A*, in sample space.
- **The *relative frequency* notion of probability** : the probability of an event is the proportion of the time that event is observed in a large number of runs of the experiment.

Definition 1.4 Axioms of Probability

A probability measure $P[\cdot]$ *is a function that maps events in the sample space to real numbers such that*

Axiom 1 *For any event A,* $P[A] \geq 0$.
Axiom 2 $P[S] = 1$.
Axiom 3 *For any countable collection* $A_1, A_2, \ldots$ *of mutually exclusive events*

$$P[A_1 \cup A_2 \cup \cdots] = P[A_1] + P[A_2] + \cdots .$$

Axioms 1 and 2 simply establish a probability as a number between 0 and 1.

Axiom 3 states that the probability of the union of mutually exclusive events is the sum of the individual probabilities.

- To use Axiom 3 to solve a practical problem, we will learn to analyze a complicated event as the union of mutually exclusive events whose probabilities we can calculate.
- Then, we will add the probabilities of the mutually exclusive events to find the probability of the complicated event we are interested in.
- A useful extension of Axiom 3 applies to the union of two mutually exclusive events.

Theorem 1.2

For mutually exclusive events A_1 and A_2,

$$\mathrm{P}[A_1 \cup A_2] = \mathrm{P}[A_1] + \mathrm{P}[A_2].$$

Although it may appear that Theorem 1.2 is a trivial special case of Axiom 3, this is not so. In fact, a simple proof of Theorem 1.2 may also use Axiom 2! If you are curious, Problem 1.3.13 gives the first steps toward a proof. It is a simple matter to extend Theorem 1.2 to any finite union of mutually exclusive sets.

Theorem 1.3

If $A = A_1 \cup A_2 \cup \cdots \cup A_m$ and $A_i \cap A_j = \varnothing$ for $i \neq j$, then

$$\mathrm{P}[A] = \sum_{i=1}^{m} \mathrm{P}[A_i].$$

The probability measure established by the axioms corresponds to the idea of relative frequency. The correspondence refers to a sequential experiment consisting of *n* repetitions of the basic experiment.

We refer to each repetition of the experiment as a *trial* . In these *n* trials, $N_A(n)$ is the number of times that event *A* occurs. The relative frequency of *A* is the fraction $N_A(n)/n$.

Theorem 10.7 proves that $\lim_{n\to\infty} NA(n)/n = \mathrm{P}[A]$.

Theorem 1.4

The probability measure $\mathrm{P}[\cdot]$ *satisfies*

(a) $\mathrm{P}[\varnothing] = 0.$

(b) $\mathrm{P}[A^c] = 1 - \mathrm{P}[A].$

(c) For any A *and* B *(not necessarily mutually exclusive),*

$$\mathrm{P}[A \cup B] = \mathrm{P}[A] + \mathrm{P}[B] - \mathrm{P}[A \cap B].$$

(d) If $A \subset B$*, then* $\mathrm{P}[A] \leq \mathrm{P}[B].$

Another consequence of the axioms can be expressed as the following theorem:

Theorem 1.5

The probability of an event $B = \{s_1, s_2, \ldots, s_m\}$ is the sum of the probabilities of the outcomes contained in the event:

$$\mathrm{P}[B] = \sum_{i=1}^{m} \mathrm{P}[\{s_i\}].$$

Proof Each outcome s_i is an event (a set) with the single element s_i. Since outcomes by definition are mutually exclusive, B can be expressed as the union of m mutually exclusive sets:

$$B = \{s_1\} \cup \{s_2\} \cup \cdots \cup \{s_m\}$$

with $\{s_i\} \cap \{s_j\} = \varnothing$ for $i \neq j$.

Applying Theorem 1.3 with $B_i = \{s_i\}$ yields $\mathrm{P}[B] = \sum_{i=1}^{m} \mathrm{P}[\{s_i\}]$

Comments on Notation

We use the notation P[·] to indicate the probability of an event. The expression in the square brackets is an event.

P[*A*] can be viewed as a function that transforms event *A* to a number between 0 and 1.

Note that $\{s_i\}$ is the formal notation for a set with the single element s_i.

Write P[s_i] rather than the more complete P[$\{s_i\}$] to denote the probability of this outcome.

The probability of intersection of two events by definition: **P[$A \cap B$] = P[*A,B*] = P[*AB*].**

Equally Likely Outcomes

A large number of experiments have a sample space $S = \{s_1, \ldots, s_n\}$ in which our knowledge of the practical situation leads us to believe that no one outcome is any more likely than any other.

In these experiments we say that the *n* outcomes are *equally likely*.

In such a case, the axioms of probability imply that every outcome has probability **1/*n*.**

Theorem 1.6

For an experiment with sample space $S = \{s_1, \ldots, s_n\}$ in which each outcome s_i is equally likely,

$$\mathrm{P}[s_i] = 1/n \qquad 1 \le i \le n.$$

Proof Since all outcomes have equal probability, there exists p such that $\mathrm{P}[s_i] = p$ for $i = 1, \ldots, n$. Theorem 1.5 implies

$$\mathrm{P}[S] = \mathrm{P}[s_1] + \cdots + \mathrm{P}[s_n] = np.$$

Since Axiom 2 says $\mathrm{P}[S] = 1$, $p = 1/n$.

Example 1.10

As in Example 1.7, roll a six-sided die in which all faces are equally likely. What is the probability of each outcome? Find the probabilities of the events: "Roll 4 or higher," "Roll an even number," and "Roll the square of an integer."

The probability of each outcome is $\mathrm{P}[i] = 1/6$ for $i = 1, 2, \ldots, 6$. The probabilities of the three events are

- $\mathrm{P}[\text{Roll 4 or higher}] = \mathrm{P}[4] + \mathrm{P}[5] + \mathrm{P}[6] = 1/2.$
- $\mathrm{P}[\text{Roll an even number}] = \mathrm{P}[2] + \mathrm{P}[4] + \mathrm{P}[6] = 1/2.$
- $\mathrm{P}[\text{Roll the square of an integer}] = \mathrm{P}[1] + \mathrm{P}[4] = 1/3.$

Quiz 1.3

A student's test score T is an integer between 0 and 100 corresponding to the experimental outcomes $s_0, \ldots, s_{100}$. A score of 90 to 100 is an A, 80 to 89 is a B, 70 to 79 is a C, 60 to 69 is a D, and below 60 is a failing grade of F. If all scores between 51 and 100 are equally likely and a score of 50 or less never occurs, find the following probabilities:

(a) $\mathrm{P}[\{s_{100}\}]$

(b) $\mathrm{P}[A]$

(c) $\mathrm{P}[F]$

(d) $\mathrm{P}[T < 90]$

(e) P[a C grade or better]

(f) P[student passes]

1.4 Conditional Probability

Conditional probabilities correspond to a modified probability model that reflects partial information about the outcome of an experiment.

The modified model has a smaller sample space than the original model.

It is sometimes useful to interpret P[*A*] as our knowledge of the occurrence of event *A* before an experiment takes place.

- If P[*A*] ≈ 1, we have advance knowledge that *A* will almost certainly occur.
- P[*A*] ≈ 0 reflects strong knowledge that *A* is unlikely to occur when the experiment takes place.
- With P[*A*] ≈ 1/2, we have little knowledge about whether or not *A* will occur.

Thus P[*A*] reflects our knowledge of the occurrence of *A prior* to performing an experiment.

Sometimes, we refer to P[*A*] as the *a priori probability*, or the *prior probability*, of *A*.

1.4 Conditional Probability (cont.)

- In many practical situations, it is not possible to find out the precise outcome of an experiment.
- Rather than the outcome s_i, itself, we obtain information that the outcome is in the set B.
- That is, we learn that some event B has occurred, where B consists of several outcomes.
- Conditional probability describes our knowledge of A when we know that B has occurred but we still don't know the precise outcome.
- The notation for this new probability is P[A|B] : "***the probability of A given B***."
- Before going to the mathematical definition of conditional probability, we provide an example that gives an indication of how conditional probabilities can be used.

Example 1.11

Consider an experiment that consists of testing two integrated circuits (IC chips) that come from the same silicon wafer and observing in each case whether a chip is accepted (*a*) or rejected (*r*).

The sample space of the experiment is $S = \{rr, ra, ar, aa\}$.

- Let B denote the event that the first chip tested is rejected. Mathematically, $B = \{rr, ra\}$.
- Similarly, let $A = \{rr, ar\}$ denote the event that the second chip is a failure.

The chips come from a high-quality production line. So, the prior probability $P[A]$ is very low.

In advance, we are pretty certain that the second circuit will be accepted. However, some wafers become contaminated by dust, and these wafers have a high proportion of defective chips.

When the first chip is a reject, the outcome of the experiment is in event B and $P[A|B]$, the probability that the second chip will also be rejected, is higher than the *a priori* probability $P[A]$ because of the likelihood that dust contaminated the entire wafer.

Definition 1.5 Conditional Probability

The conditional probability of the event A given the occurrence of the event B is

$$\mathrm{P}[A|B] = \frac{\mathrm{P}[AB]}{\mathrm{P}[B]}.$$

Conditional probability is defined only when P[*B*] > 0.

P[*B*] = 0 means that it is certain that *B* never occurs. It is illogical to speak of the probability of *A* given that *B* occurs.

P[*A*|*B*] is a respectable probability measure relative to a sample space that consists of all the outcomes in *B*.

This means that P[*A*|*B*] has properties corresponding to the three axioms of probability.

Theorem 1.7

A conditional probability measure $\mathrm{P}[A|B]$ *has the following properties that correspond to the axioms of probability.*

Axiom 1: $\mathrm{P}[A|B] \geq 0.$

Axiom 2: $\mathrm{P}[B|B] = 1.$

Axiom 3: If $A = A_1 \cup A_2 \cup \cdots$ *with* $A_i \cap A_j = \varnothing$ *for* $i \neq j$, *then*

$$\mathrm{P}[A|B] = \mathrm{P}[A_1|B] + \mathrm{P}[A_2|B] + \cdots$$

Example 1.12

With respect to Example 1.11, consider the *a priori* probability model

$$\mathrm{P}\,[rr] = 0.01, \quad \mathrm{P}\,[ra] = 0.01, \quad \mathrm{P}\,[ar] = 0.01, \quad \mathrm{P}\,[aa] = 0.97.$$

Find the probability of $A =$ "second chip rejected" and $B =$ "first chip rejected." Also find the conditional probability that the second chip is a reject given that the first chip is a reject.

We saw in Example 1.11 that A is the union of two mutually exclusive events (outcomes) rr and ar. Therefore, the a priori probability that the second chip is rejected is

$$\mathrm{P}\,[A] = \mathrm{P}\,[rr] + \mathrm{P}\,[ar] = 0.02$$

This is also the a priori probability that the first chip is rejected:

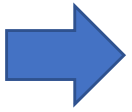
$$\mathrm{P}\,[B] = \mathrm{P}\,[rr] + \mathrm{P}\,[ra] = 0.02.$$

The conditional probability of the second chip being rejected given that the first chip is rejected is, by definition, the ratio of $\mathrm{P}[AB]$ to $\mathrm{P}[B]$, where, in this example,

$$\mathrm{P}\,[AB] = \mathrm{P}\,[\text{both rejected}] = \mathrm{P}\,[rr] = 0.01$$

Thus

$$\mathrm{P}\,[A|B] = \frac{\mathrm{P}\,[AB]}{\mathrm{P}\,[B]} = 0.01/0.02 = 0.5.$$



The information that the first chip is a reject drastically changes our state of knowledge about the second chip. We started with near certainty, $\mathrm{P}[A] = 0.02$, that the second chip would not fail and ended with complete uncertainty about the quality of the second chip, $\mathrm{P}[A|B] = 0.5$.

Example 1.13

Shuffle a deck of cards and observe the bottom card. What is the conditional probability that the bottom card is the ace of clubs given that the bottom card is a black card?

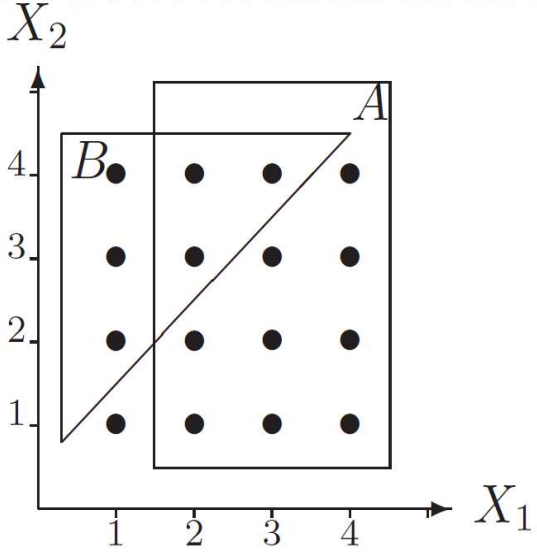
The sample space consists of the 52 cards that can appear on the bottom of the deck. Let A denote the event that the bottom card is the ace of clubs. Since all cards are equally likely to be at the bottom, the probability that a particular card, such as the ace of clubs, is at the bottom is $\mathrm{P}[A] = 1/52$. Let B be the event that the bottom card is a black card. The event B occurs if the bottom card is one of the 26 clubs or spades, so that $\mathrm{P}[B] = 26/52$. Given B, the conditional probability of A is

$$\mathrm{P}[A|B] = \frac{\mathrm{P}[AB]}{\mathrm{P}[B]} = \frac{\mathrm{P}[A]}{\mathrm{P}[B]} = \frac{1/52}{26/52} = \frac{1}{26}.$$

The key step was observing that $AB = A$, because if the bottom card is the ace of clubs, then the bottom card must be a black card. Mathematically, this is an example of the fact that $A \subset B$ implies that $AB = A$.

Example 1.14

Roll two fair four-sided dice. Let X_1 and X_2 denote the number of dots that appear on die 1 and die 2, respectively. Let A be the event $X_1 \geq 2$. What is $\mathrm{P}[A]$? Let B denote the event $X_2 > X_1$. What is $\mathrm{P}[B]$? What is $\mathrm{P}[A|B]$?



We begin by observing that the sample space has 16 elements corresponding to the four possible values of X_1 and the same four values of X_2. Since the dice are fair, the outcomes are equally likely, each with probability $1/16$. We draw the sample space as a set of black circles in a two-dimensional diagram, in which the axes represent the events X_1 and X_2. Each outcome is a pair of values (X_1, X_2). The rectangle represents A. It contains 12 outcomes, each with probability $1/16$.

To find $\mathrm{P}[A]$, we add up the probabilities of outcomes in A, so $\mathrm{P}[A] = 12/16 = 3/4$. The triangle represents B. It contains six outcomes. Therefore $\mathrm{P}[B] = 6/16 = 3/8$. The event AB has three outcomes, $(2,3), (2,4), (3,4)$, so $\mathrm{P}[AB] = 3/16$. From the definition of conditional probability, we write

$$\mathrm{P}[A|B] = \frac{\mathrm{P}[AB]}{\mathrm{P}[B]} = \frac{1}{2}.$$

We can also derive this fact from the diagram by restricting our attention to the six outcomes in B (the conditioning event) and noting that three of the six outcomes in B (one-half of the total) are also in A.

Quiz 1.4

Monitor three consecutive packets going through an Internet router. Classify each one as either video (v) or data (d). Your observation is a sequence of three letters (each one is either v or d). For example, three video packets corresponds to vvv. The outcomes vvv and ddd each have probability 0.2 whereas each of the other outcomes vvd, vdv, vdd, dvv, dvd, and ddv has probability 0.1. Count the number of video packets N_V in the three packets you have observed. Describe in words and also calculate the following probabilities:

(a) $\mathrm{P}[N_V = 2]$

(b) $\mathrm{P}[N_V \geq 1]$

(c) $\mathrm{P}[\{vvd\}|N_V = 2]$

(d) $\mathrm{P}[\{ddv\}|N_V = 2]$

(e) $\mathrm{P}[N_V = 2|N_V \geq 1]$

(f) $\mathrm{P}[N_V \geq 1|N_V = 2]$

1.5 Partitions and the Law of Total Probability

A partition divides the sample space into mutually exclusive sets.

The law of total probability expresses the probability of an event as the sum of the probabilities of outcomes that are in the separate sets of a partition.

Example 1.15: Flip four coins, a penny, a nickel, a dime, and a quarter. Examine the coins in order (penny, then nickel, then dime, then quarter) and observe whether each coin shows a head (*h*) or a tail (*t*). What is the sample space? How many elements are in the sample space?

. . . . . .

The sample space consists of 16 four-letter words, with each letter either *h* or *t*. For example, the outcome *tthh* refers to the penny and the nickel showing tails and the dime and quarter showing heads. There are 16 members of the sample space.

Example 1.16

Continuing Example 1.15, let $B_i = \{\text{outcomes with } i \text{ heads}\}$. Each B_i is an event containing one or more outcomes. For example, $B_1 = \{ttth, ttht, thtt, httt\}$ contains four outcomes. The set $B = \{B_0, B_1, B_2, B_3, B_4\}$ is a partition. Its members are mutually exclusive and collectively exhaustive. It is not a sample space because it lacks the finest-grain property. Learning that an experiment produces an event B_1 tells you that one coin came up heads, but it doesn't tell you which coin it was.

Mathematically, the experiment is equivalent to many real engineering problems.

For example, observe a pair of modems transmitting four bits from one computer to another. For each bit, observe whether the receiving modem detects the bit correctly (*c*) or makes an error (*e*). Or test four integrated circuits.

For each one, observe whether the circuit is acceptable (*a*) or a reject (*r*).

- In all of these examples, the sample space contains 16 four-letter words formed with an alphabet containing two letters.
- If we are interested only in the number of times one of the letters occurs, it is sufficient to refer only to the partition *B*, which does not contain all of the information about the experiment but does contain all of the information we need.
- The partition is simpler to deal with than the sample space because it has fewer members (there are five events in the partition and 16 outcomes in the sample space).
- The simplification is much more significant when the complexity of the experiment is higher.
- **For example,** in testing 20 circuits the sample space has 2^{20} = 1,048,576 members, while the corresponding partition has only 21 members.

Theorem 1.8

For a partition $B = \{B_1, B_2, \ldots\}$ *and any event* A *in the sample space, let* $C_i = A \cap B_i$. *For* $i \neq j$, *the events* C_i *and* C_j *are mutually exclusive and*

$$A = C_1 \cup C_2 \cup \cdots .$$

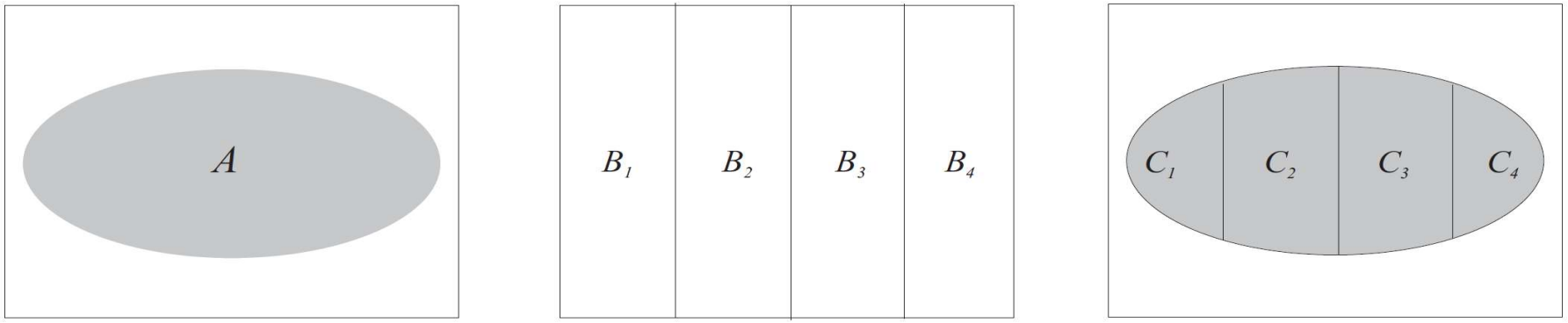


Figure 1.1 In this example of Theorem 1.8, the partition is $B = \{B_1, B_2, B_3, B_4\}$ and $C_i = A \cap B_i$ for $i = 1, \ldots, 4$. It should be apparent that $A = C_1 \cup C_2 \cup C_3 \cup C_4$.

Example 1.17

In the coin-tossing experiment of Example 1.15, let A equal the set of outcomes with less than three heads:

$$A = \{tttt, httt, thtt, ttht, ttth, hhtt, htht, htth, tthh, thth, thht\}.$$

From Example 1.16, let $B_i = \{\text{outcomes with } i \text{ heads}\}$. Since $\{B_0, \ldots, B_4\}$ is a partition, Theorem 1.8 states that

$$A = (A \cap B_0) \cup (A \cap B_1) \cup (A \cap B_2) \cup (A \cap B_3) \cup (A \cap B_4)$$

In this example, $B_i \subset A$, for $i = 0, 1, 2$. Therefore $A \cap B_i = B_i$ for $i = 0, 1, 2$. Also, for $i = 3$ and $i = 4$, $A \cap B_i = \varnothing$ so that $A = B_0 \cup B_1 \cup B_2$, a union of mutually exclusive sets. In words, this example states that the event "less than three heads" is the union of events "zero heads," "one head," and "two heads."

For example, find the probability that there are three or more bad circuits in a batch that comes from a fabrication machine.

The following theorem refers to a partition $\{B_1, B_2, \ldots, B_m\}$ and any event, A.

It states that we can find the probability of A by adding the probabilities of the parts of A that are in the separate components of the event space.

Theorem 1.9

For any event A, and partition $\{B_1, B_2, \ldots, B_m\}$,

$$\mathrm{P}[A] = \sum_{i=1}^{m} \mathrm{P}[A \cap B_i].$$

Proof The proof follows directly from Theorem 1.8 and Theorem 1.3. In this case, the mutually exclusive sets are $C_i = \{A \cap B_i\}$.

This theorem is often used when the sample space can be written in the form of a table.

In this table, the rows and columns each represent a partition.

This method is shown in the following example.

Example 1.18

A company has a model of email use. It classifies all emails as either long (l), if they are over 10 MB in size, or brief (b). It also observes whether the email is just text (t), has attached images (i), or has an attached video (v). This model implies an experiment in which the procedure is to monitor an email and the observation consists of the type of email, t, i, or v, and the length, l or b. The sample space has six outcomes: $S = \{lt, bt, li, bi, lv, bv\}$. In this problem, each email is classifed in two ways: by length and by type. Using L for the event that an email is long and B for the event that a email is brief, $\{L, B\}$ is a partition. Similarly, the text (T), image (I), and video (V) classification is a partition $\{T, I, V\}$. The sample space can be represented by a table in which the rows and columns are labeled by events and the intersection of each row and column event contains a single outcome. The corresponding table entry is the probability of that outcome. In this case, the table is

	T	I	V
L	0.3	0.12	0.15
B	0.2	0.08	0.15

For example, from the table we can read that the probability of a brief image email is $\mathrm{P}[bi] = \mathrm{P}[BI] = 0.08$. Note that $\{T, I, V\}$ is a partition corresponding to $\{B_1, B_2, B_3\}$ in Theorem 1.9. Thus we can apply Theorem 1.9 to find the probability of a long email:

$$\mathrm{P}[L] = \mathrm{P}[LT] + \mathrm{P}[LI] + \mathrm{P}[LV] = 0.57.$$

Example 1.18

A company has a model of email use. It classifies all emails as either long (l), if they are over 10 MB in size, or brief (b). It also observes whether the email is just text (t), has attached images (i), or has an attached video (v). This model implies an experiment in which the procedure is to monitor an email and the observation consists of the type of email, t, i, or v, and the length, l or b.

The sample space has six outcomes: $S = \{lt, bt, li, bi, lv, bv\}$.

In this problem, each email is classifed in two ways: by length and by type.

Using L for the event that an email is long and B for the event that a email is brief, $\{L, B\}$ is a partition. Similarly, the text (T), image (I), and video (V) classification is a partition $\{T, I, V\}$. The sample space can be represented by a table in which the rows and columns are labeled by events and the intersection of each row and column event contains a single outcome. The corresponding table entry is the probability of that outcome. In this case, the table is

	T	I	V
L	0.3	0.12	0.15
B	0.2	0.08	0.15

For example, from the table we can read that the probability of a brief image email is $\mathrm{P}[bi] = \mathrm{P}[BI] = 0.08$. Note that $\{T, I, V\}$ is a partition corresponding to $\{B_1, B_2, B_3\}$ in Theorem 1.9. Thus we can apply Theorem 1.9 to find the probability of a long email:

$$\mathrm{P}[L] = \mathrm{P}[LT] + \mathrm{P}[LI] + \mathrm{P}[LV] = 0.57.$$

Law of Total Probability

In many applications, we begin with information about conditional probabilities and use the law of total probability to calculate unconditional probabilities.

Theorem 1.10 Law of Total Probability

For a partition $\{B_1, B_2, \ldots, B_m\}$ *with* $\mathrm{P}[B_i] > 0$ *for all* i,

$$\mathrm{P}[A] = \sum_{i=1}^{m} \mathrm{P}[A|B_i]\,\mathrm{P}[B_i].$$

Proof This follows from Theorem 1.9 and the identity $\mathrm{P}[AB_i] = \mathrm{P}[A|B_i]\,\mathrm{P}[B_i]$, which is a direct consequence of the definition of conditional probability.

The usefulness of the result can be seen in the next example.

Example 1.19

A company has three machines B_1, B_2, and B_3 making 1 kΩ resistors. Resistors within 50 Ω of the nominal value are considered acceptable. It has been observed that 80% of the resistors produced by B_1 and 90% of the resistors produced by B_2 are acceptable. The percentage for machine B_3 is 60%. Each hour, machine B_1 produces 3000 resistors, B_2 produces 4000 resistors, and B_3 produces 3000 resistors. All of the resistors are mixed together at random in one bin and packed for shipment. What is the probability that the company ships an acceptable resistor?

Let $A = \{\text{resistor is acceptable}\}$. Using the resistor accuracy information to formulate a probability model, we write

$$\mathrm{P}\left[A|B_1\right] = 0.8, \quad \mathrm{P}\left[A|B_2\right] = 0.9, \quad \mathrm{P}\left[A|B_3\right] = 0.6.$$

The production figures state that $3000 + 4000 + 3000 = 10{,}000$ resistors per hour are produced. The fraction from machine B_1 is $\mathrm{P}[B_1] = 3000/10{,}000 = 0.3$. Similarly, $\mathrm{P}[B_2] = 0.4$ and $\mathrm{P}[B_3] = 0.3$. Now it is a simple matter to apply the law of total probability to find the acceptable probability for all resistors shipped by the company:

$$\begin{aligned}\mathrm{P}\left[A\right] &= \mathrm{P}\left[A|B_1\right]\mathrm{P}\left[B_1\right] + \mathrm{P}\left[A|B_2\right]\mathrm{P}\left[B_2\right] + \mathrm{P}\left[A|B_3\right]\mathrm{P}\left[B_3\right] \\ &= (0.8)(0.3) + (0.9)(0.4) + (0.6)(0.3) = 0.78.\end{aligned}$$

For the whole factory, 78% of resistors are within 50 Ω of the nominal value.

Bayes' Theorem

When we have advance information about P[*A*|*B*] and need to calculate P[*B*|*A*], we refer to the following formula:

Theorem 1.11 Bayes' theorem

$$\mathrm{P}\left[B|A\right] = \frac{\mathrm{P}\left[A|B\right]\mathrm{P}\left[B\right]}{\mathrm{P}\left[A\right]}.$$

Proof

$$\mathrm{P}\left[B|A\right] = \frac{\mathrm{P}\left[AB\right]}{\mathrm{P}\left[A\right]} = \frac{\mathrm{P}\left[A|B\right]\mathrm{P}\left[B\right]}{\mathrm{P}\left[A\right]}.$$

Bayes' theorem is a simple consequence of the definition of ***conditional probability***.

It has a name because it is extremely useful for making inferences about phenomena that cannot be observed directly.

Sometimes these inferences are described as "**reasoning about causes when we observe effects**." Bazen bu çıkarımlar "sonuçları gözlemlediğimizde sebepler hakkında akıl yürütme" olarak tanımlanır.

For example, let $\{B_1, \ldots, B_m\}$ be a partition that includes all possible states of something that interests us but that

- We cannot observe directly (*for example, the machine that made a particular resistor*).
- For each possible state, B_i, we know the prior probability $\mathbf{P[}\boldsymbol{B_i}\mathbf{]}$ and $\mathbf{P[}\boldsymbol{A}\mathbf{|}\boldsymbol{B_i}\mathbf{]}$, the probability that an event *A* occurs (*the resistor meets a quality criterion*) if *Bi* is the actual state.
- Now we observe the actual event (*either the resistor passes or fails a test*), and we ask about the thing we are interested in (*the machines that might have produced the resistor*).
- That is, we use Bayes' theorem to find $\mathbf{P[}\boldsymbol{B_1}\mathbf{|}\boldsymbol{A}\mathbf{]}$, $\mathbf{P[}\boldsymbol{B_2}\mathbf{|}\boldsymbol{A}\mathbf{]}$, . . . , $\mathbf{P[}\boldsymbol{B_m}\mathbf{|}\boldsymbol{A}\mathbf{]}$.
- In performing the calculations, we use the law of total probability to calculate the denominator in Theorem 1.11. Thus for state B_i,

$$\mathrm{P}\left[B_i|A\right] = \frac{\mathrm{P}\left[A|B_i\right]\mathrm{P}\left[B_i\right]}{\sum_{i=1}^{m}\mathrm{P}\left[A|B_i\right]\mathrm{P}\left[B_i\right]}$$

Example 1.20

In Example 1.19 about a shipment of resistors from the factory, we learned that:

- The probability that a resistor is from machine B_3 is $\mathrm{P}[B_3] = 0.3$.
- The probability that a resistor is *acceptable* — i.e., within 50 Ω of the nominal value — is $\mathrm{P}[A] = 0.78$.
- Given that a resistor is from machine B_3, the conditional probability that it is acceptable is $\mathrm{P}[A|B_3] = 0.6$.

What is the probability that an acceptable resistor comes from machine B_3?

Now we are given the event A that a resistor is within 50 Ω of the nominal value, and we need to find $\mathrm{P}[B_3|A]$. Using Bayes' theorem, we have

$$\mathrm{P}\left[B_3|A\right] = \frac{\mathrm{P}\left[A|B_3\right]\mathrm{P}\left[B_3\right]}{\mathrm{P}\left[A\right]}.$$

Since all of the quantities we need are given in the problem description, our answer is

$$\mathrm{P}\left[B_3|A\right] = (0.6)(0.3)/(0.78) = 0.23.$$

Similarly we obtain $\mathrm{P}[B_1|A] = 0.31$ and $\mathrm{P}[B_2|A] = 0.46$. Of all resistors within 50 Ω of the nominal value, only 23% come from machine B_3 (even though this machine produces 30% of all resistors). Machine B_1 produces 31% of the resistors that meet the 50 Ω criterion and machine B_2 produces 46% of them.

Quiz 1.5

Monitor customer behavior in the Phonesmart store. Classify the behavior as buying (B) if a customer purchases a smartphone. Otherwise the behavior is no purchase (N). Classify the time a customer is in the store as long (L) if the customer stays more than three minutes; otherwise classify the amount of time as rapid (R). Based on experience with many customers, we use the probability model $\mathrm{P}[N] = 0.7$, $\mathrm{P}[L] = 0.6$, $\mathrm{P}[NL] = 0.35$. Find the following probabilities:

(a) $\mathrm{P}[B \cup L]$

(b) $\mathrm{P}[N \cup L]$

(c) $\mathrm{P}[N \cup B]$

(d) $\mathrm{P}[LR]$

1.6 Independence

Two events are independent if observing one event does not change the probability of observing the other event.

Definition 1.6 Two Independent Events

Events A and B are ***independent*** *if and only if*

$$\mathrm{P}[AB] = \mathrm{P}[A]\,\mathrm{P}[B].$$

When events *A* and *B* have nonzero probabilities, the following formulas are equivalent to the definition of independent events:

$$\mathrm{P}[A|B] = \mathrm{P}[A], \qquad \mathrm{P}[B|A] = \mathrm{P}[B].$$

To interpret independence, consider probability as a description of our knowledge of the result of the experiment.

P[*A*] describes our prior knowledge (*before the experiment is performed*) that the outcome is included in event *A*. The fact that the outcome is in *B* is partial information about the experiment.

P[*A*|*B*] reflects our knowledge of *A* when we learn that *B* occurs.

P[*A*|*B*]=P[*A*] states that learning that *B* occurs does not change our information about *A*.

It is in this sense that the events are independent.

if A and B are independent, then A and B^c are also independent.

The logic behind this conclusion is that if learning that event B occurs does not alter the probability of event A, then learning that B does not occur also should not alter the probability of A.

Keep in mind that **independent and mutually exclusive are *not* synonyms**.

Mutually exclusive events A and B have no outcomes in common. So, P[AB] = 0.

In most cases independent events are not mutually exclusive! Exceptions occur only when P[A] = 0 or P[B] = 0.

Knowledge that events A and B are *mutually exclusive* is very helpful.

Axiom 3 enables us to *add* their probabilities to obtain the probability of the *union*.

Knowledge that events C and D are *independent* is also very useful.

P [AB] = P[A] P [B] enables us to *multiply* their probabilities to obtain the probability of the *intersection*.

Example 1.21

Suppose that for the experiment monitoring three purchasing decisions in Example 1.9, each outcome (a sequence of three decisions, each either buy or not buy) is equally likely. Are the events B_2 that the second customer purchases a phone and N_2 that the second customer does not purchase a phone independent? Are the events B_1 and B_2 independent?

Each element of the sample space $S = \{bbb, bbn, bnb, bnn, nbb, nbn, nnb, nnn\}$ has probability $1/8$. Each of the events

$$B_2 = \{bbb, bbn, nbb, nbn\} \quad \text{and} \quad N_2 = \{bnb, bnn, nnb, nnn\}$$

contains four outcomes, so $\mathrm{P}[B_2] = \mathrm{P}[N_2] = 4/8$. However, $B_2 \cap N_2 = \varnothing$ and $\mathrm{P}[B_2 N_2] = 0$. That is, B_2 and N_2 are mutually exclusive because the second customer cannot both purchase a phone and not purchase a phone. Since $\mathrm{P}[B_2 N_2] \neq \mathrm{P}[B_2]\,\mathrm{P}[N_2]$, B_2 and N_2 are not independent. Learning whether or not the event B_2 (second customer buys a phone) occurs drastically affects our knowledge of whether or not the event N_2 (second customer does not bu
events $B_1 = \{bnn, bnb, bbn, bbb\}$ and $B_2 = \{bbn,$
so $\mathrm{P}[B_1] = \mathrm{P}[B_2] = 4/8 = 1/2$. In this case, the i
has probability $\mathrm{P}[B_1 B_2] = 2/8 = 1/4$. Since $\mathrm{P}[B_1 B_2] = \mathrm{P}[B_1]\,\mathrm{P}[B_2]$, B_1 and B_2 are independent. Learning whether or not the
a phone) occurs does not affect our knowledge of w
customer buys a phone) occurs.

In this example we have analyzed a probability model to determine whether two events are independent. In many practical applications we reason in the opposite direction. Our knowledge of an experiment leads us to *assume* that certain pairs of events are independent. We then use this knowledge to build a probability model for the experiment.

Example 1.22

Integrated circuits undergo two tests. A mechanical test determines whether pins have the correct spacing, and an electrical test checks the relationship of outputs to inputs. We *assume* that electrical failures and mechanical failures occur independently. Our information about circuit production tells us that mechanical failures occur with probability 0.05 and electrical failures occur with probability 0.2. What is the probability model of an experiment that consists of testing an integrated circuit and observing the results of the mechanical and electrical tests?

To build the probability model, we note that the sample space contains four outcomes:

$$S = \{(ma, ea), (ma, er), (mr, ea), (mr, er)\}$$

where m denotes mechanical, e denotes electrical, a denotes accept, and r denotes reject. Let M and E denote the events that the mechanical and electrical tests are acceptable. Our prior information tells us that $\mathrm{P}[M^c] = 0.05$, and $\mathrm{P}[E^c] = 0.2$. This implies $\mathrm{P}[M] = 0.95$ and $\mathrm{P}[E] = 0.8$. Using the independence assumption and Definition 1.6, we obtain the probabilities of the four outcomes:

$$\mathrm{P}[(ma, ea)] = \mathrm{P}[ME] = \mathrm{P}[M]\,\mathrm{P}[E] = 0.95 \times 0.8 = 0.76,$$
$$\mathrm{P}[(ma, er)] = \mathrm{P}[ME^c] = \mathrm{P}[M]\,\mathrm{P}[E^c] = 0.9$$
$$\mathrm{P}[(mr, ea)] = \mathrm{P}[M^cE] = \mathrm{P}[M^c]\,\mathrm{P}[E] = 0.0$$
$$\mathrm{P}[(mr, er)] = \mathrm{P}[M^cE^c] = \mathrm{P}[M^c]\,\mathrm{P}[E^c] =$$

We have considered independence as a property of a pair of events. Often we consider larger sets of independent events. For more than two events to be *independent*, the probability model has to meet a set of conditions. To define mutual independence, we begin with three sets.

Definition 1.7 Three Independent Events

*A_1, A_2, and A_3 are **mutually independent** if and only if*

(a) A_1 *and* A_2 *are independent,*

(b) A_2 *and* A_3 *are independent,*

(c) A_1 *and* A_3 *are independent,*

(d) $\mathrm{P}[A_1 \cap A_2 \cap A_3] = \mathrm{P}[A_1]\,\mathrm{P}[A_2]\,\mathrm{P}[A_3]$.

The final condition is a simple extension of Definition 1.6. The following example shows why this condition is insufficient to guarantee that "everything is independent of everything else," the idea at the heart of independence.

Example 1.23

In an experiment with equiprobable outcomes, the partition is $S = \{1, 2, 3, 4\}$. $\mathrm{P}[s] = 1/4$ for all $s \in S$. Are the events $A_1 = \{1, 3, 4\}$, $A_2 = \{2, 3, 4\}$, and $A_3 = \varnothing$ mutually independent?

These three sets satisfy the final condition of Definition 1.7 because $A_1 \cap A_2 \cap A_3 = \varnothing$, and

$$\mathrm{P}\left[A_1 \cap A_2 \cap A_3\right] = \mathrm{P}\left[A_1\right]\mathrm{P}\left[A_2\right]\mathrm{P}\left[A_3\right] = 0.$$

However, A_1 and A_2 are not independent because, with all outcomes equiprobable,

$$\mathrm{P}\left[A_1 \cap A_2\right] = \mathrm{P}\left[\{3, 4\}\right] = 1/2 \neq \mathrm{P}\left[A_1\right]\mathrm{P}\left[A_2\right] = 3/4 \times 3/4.$$

Hence the three events are not mutually independent.

Definition 1.8 More than Two Independent Events

If $n \geq 3$, the events $A_1, A_2, \ldots, A_n$ are mutually independent if and only if

(a) all collections of $n-1$ events chosen from $A_1, A_2, \ldots A_n$ are mutually independent,

(b) $\mathrm{P}[A_1 \cap A_2 \cap \cdots \cap A_n] = \mathrm{P}[A_1]\,\mathrm{P}[A_2] \cdots \mathrm{P}[A_n]$.

This definition and Example 1.23 show us that when n > 2 it is a complex matter to determine whether or not n events are mutually independent.

On the other hand, if we know that n events are mutually independent, it is a simple matter to determine the probability of the intersection of any subset of the n events.

Just multiply the probabilities of the events in the subset.

Quiz 1.6

Monitor two consecutive packets going through a router. Classify each one as video (v) if it was sent from a Youtube server or as ordinary data (d) otherwise. Your observation is a sequence of two letters (either v or d). For example, two video packets corresponds to vv. The two packets are independent and the probability that any one of them is a video packet is 0.8. Denote the identity of packet i by C_i. If packet i is a video packet, then $C_i = v$; otherwise, $C_i = d$. Count the number N_V of video packets in the two packets you have observed. Determine whether the following pairs of events are independent:

(a) $\{N_V = 2\}$ and $\{N_V \geq 1\}$

(b) $\{N_V \geq 1\}$ and $\{C_1 = v\}$

(c) $\{C_2 = v\}$ and $\{C_1 = d\}$

(d) $\{C_2 = v\}$ and $\{N_V$ is even$\}$

1.2.1 A fax transmission can take place at any of three speeds depending on the condition of the phone connection between the two fax machines. The speeds are high (*h*) at 14, 400 b/s, medium (*m*) at 9600 b/s, and low (*l*) at 4800 b/s. In response to requests for information, a company sends either short faxes of two (*t*) pages, or long faxes of four (*f*) pages. Consider the experiment of monitoring a fax transmission and observing the transmission speed and length. An observation is a two-letter word, for example, a high-speed, two-page fax is *ht*.

(a) What is the sample space of the experiment?

(b) Let A_1 be the event "medium-speed fax." What are the outcomes in A_1?

(c) Let A_2 be the event "short (two-page) fax." What are the outcomes in A_2?

(d) Let A_3 be the event "high-speed fax or low-speed fax." What are the outcomes in A_3?

(a) An outcome specifies whether the connection speed is high (h), medium (m), or low (l) speed, and whether the signal is a mouse click (c) or a tweet (t). The sample space is

$$S = \{ht, hc, mt, mc, lt, lc\}.$$

(b) The event that the wi-fi connection is medium speed is $A_1 = \{mt, mc\}$.

(c) The event that a signal is a mouse click is $A_2 = \{hc, mc, lc\}$.

(d) The event that a connection is either high speed or low speed is $A_3 = \{ht, hc, lt, lc\}$.

(e) Are A_1, A_2, and A_3 mutually exclusive?

(e) Since $A_1 \cap A_2 = \{mc\}$ and is not empty, A_1, A_2, and A_3 are not mutually exclusive.

(f) Are A_1, A_2, and A_3 collectively exhaustive?

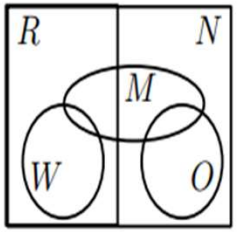
(f) Since

$$A_1 \cup A_2 \cup A_3 = \{ht, hc, mt, mc, lt, lc\} = S,$$

the collection A_1, A_2, A_3 is collectively exhaustive.

1.1.3 Ricardo's offers customers two kinds of pizza crust, Roman (R) and Neapolitan (N). All pizzas have cheese but not all pizzas have tomato sauce. Roman pizzas can have tomato sauce or they can be white (W); Neapolitan pizzas always have tomato sauce. It is possible to order a Roman pizza with mushrooms (M) added. A Neapolitan pizza can contain mushrooms or onions (O) or both, in addition to the tomato sauce and cheese. Draw a Venn diagram that shows the relationship among the ingredients N, M, O, T, and W in the menu of Ricardo's pizzeria.

Problem 1.1.3 Solution



At Ricardo's, the pizza crust is either Roman (R) or Neapolitan (N). To draw the Venn diagram on the right, we make the following observations:

- The set $\{R, N\}$ is a partition so we can draw the Venn diagram with this partition.
- Only Roman pizzas can be white. Hence $W \subset R$.
- Only a Neapolitan pizza can have onions. Hence $O \subset N$.
- Both Neapolitan and Roman pizzas can have mushrooms so that event M straddles the $\{R, N\}$ partition.
- The Neapolitan pizza can have both mushrooms and onions so $M \cap O$ cannot be empty.
- The problem statement does not preclude putting mushrooms on a white Roman pizza. Hence the intersection $W \cap M$ should not be empty.

1.2.1 A hypothetical wi-fi transmission can take place at any of three speeds depending on the condition of the radio channel between a laptop and an access point. The speeds are high (h) at 54 Mb/s, medium (m) at 11 Mb/s, and low (l) at 1 Mb/s. A user of the wi-fi connection can transmit a short signal corresponding to a mouse click (c), or a long signal corresponding to a tweet (t). Consider the experiment of monitoring wi-fi signals and observing the transmission speed and the length. An observation is a two-letter word, for example, a high-speed, mouse-click transmission is hm.

(a) What is the sample space of the experiment?

(b) Let A_1 be the event "medium speed connection." What are the outcomes in A_1?

(c) Let A_2 be the event "mouse click." What are the outcomes in A_2?

(d) Let A_3 be the event "high speed connection or low speed connection." What are the outcomes in A_3?

(e) Are A_1, A_2, and A_3 mutually exclusive?

(f) Are A_1, A_2, and A_3 collectively exhaustive?

Problem 1.2.1 Solution

(a) An outcome specifies whether the connection speed is high (h), medium (m), or low (l) speed, and whether the signal is a mouse click (c) or a tweet (t). The sample space is

$$S = \{ht, hc, mt, mc, lt, lc\}. \quad (1)$$

(b) The event that the wi-fi connection is medium speed is $A_1 = \{mt, mc\}$.

(c) The event that a signal is a mouse click is $A_2 = \{hc, mc, lc\}$.

(d) The event that a connection is either high speed or low speed is $A_3 = \{ht, hc, lt, lc\}$.

(e) Since $A_1 \cap A_2 = \{mc\}$ and is not empty, A_1, A_2, and A_3 are not mutually exclusive.

(f) Since

$$A_1 \cup A_2 \cup A_3 = \{ht, hc, mt, mc, lt, lc\} = S, \quad (2)$$

the collection A_1, A_2, A_3 is collectively exhaustive.

1.2.3 Shuffle a deck of cards and turn over the first card. What is the sample space of this experiment? How many outcomes are in the event that the first card is a heart?

Problem 1.2.3 Solution

The sample space is

$$S = \{A\clubsuit, \ldots, K\clubsuit, A\diamondsuit, \ldots, K\diamondsuit, A\heartsuit, \ldots, K\heartsuit, A\spadesuit, \ldots, K\spadesuit\}.$$

The event H that the first card is a heart is the set

$$H = \{A\heartsuit, \ldots, K\heartsuit\}.$$

The event H has 13 outcomes, corresponding to the 13 hearts in a deck.

Problem 1.2.5 Solution

Of course, there are many answers to this problem. Here are four partitions.

1.2.5 The sample space of an experiment consists of all undergraduates at a university. Give four examples of partitions.

1. We can divide students into engineers or non-engineers. Let A_1 equal the set of engineering students and A_2 the non-engineers. The pair $\{A_1, A_2\}$ is a partition.

2. We can also separate students by GPA. Let B_i denote the subset of students with GPAs G satisfying $i-1 \leq G < i$. At Rutgers, $\{B_1, B_2, \ldots, B_5\}$ is a partition. Note that B_5 is the set of all students with perfect 4.0 GPAs. Of course, other schools use different scales for GPA.

3. We can also divide the students by age. Let C_i denote the subset of students of age i in years. At most universities, $\{C_{10}, C_{11}, \ldots, C_{100}\}$ would be an event space. Since a university may have prodigies either under 10 or over 100, we note that $\{C_0, C_1, \ldots\}$ is always a partition.

4. Lastly, we can categorize students by attendance. Let D_0 denote the number of students who have missed zero lectures and let D_1 denote all other students. Although it is likely that D_0 is an empty set, $\{D_0, D_1\}$ is a well defined partition.

1.3.1 Find $P[B]$ in each case:

(a) Events A and B are a partition and $P[A] = 3P[B]$.

(b) For events A and B, $P[A \cup B] = P[A]$ and $P[A \cap B] = 0$.

(c) For events A and B, $P[A \cup B] = P[A] - P[B]$.

Problem 1.3.1 Solution

(a) A and B mutually exclusive and collectively exhaustive imply $P[A] + P[B] = 1$. Since $P[A] = 3P[B]$, we have $P[B] = 1/4$.

(b) Since $P[A \cup B] = P[A]$, we see that $B \subseteq A$. This implies $P[A \cap B] = P[B]$. Since $P[A \cap B] = 0$, then $P[B] = 0$.

(c) Since it's always true that $P[A \cup B] = P[A] + P[B] - P[AB]$, we have that

$$P[A] + P[B] - P[AB] = P[A] - P[B]. \quad (1)$$

This implies $2P[B] = P[AB]$. However, since $AB \subset B$, we can conclude that $2P[B] = P[AB] \leq P[B]$. This implies $P[B] = 0$.

1.3.3 You roll two fair six-sided dice. Find the probability $\mathrm{P}[D_3]$ that the absolute value of the difference of the dice is 3.

Problem 1.3.3 Solution

An outcome is a pair (i, j) where i is the value of the first die and j is the value of the second die. The sample space is the set

$$S = \{(1,1),(1,2),\ldots,(6,5),(6,6)\}. \tag{1}$$

with 36 outcomes, each with probability 1/36 Note that the event that the absolute value of the difference of the two rolls equals 3 is

$$D_3 = \{(1,4),(2,5),(3,6),(4,1),(5,2),(6,3)\}. \tag{2}$$

Since there are 6 outcomes in D_3, $\mathrm{P}[D_3] = 6/36 = 1/6$.

1.3.5 Computer programs are classified by the length of the source code and by the execution time. Programs with more than 150 lines in the source code are big (B). Programs with ≤ 150 lines are little (L). Fast programs (F) run in less than 0.1 seconds. Slow programs (W) require at least 0.1 seconds. Monitor a program executed by a computer. Observe the length of the source code and the run time. The probability model for this experiment contains the following information: $\mathrm{P}[LF] = 0.5$, $\mathrm{P}[BF] = 0.2$, and $\mathrm{P}[BW] = 0.2$. What is the sample space of the experiment? Calculate the following probabilities: $\mathrm{P}[W]$, $\mathrm{P}[B]$, and $\mathrm{P}[W \cup B]$.

Problem 1.3.5 Solution

The sample space of the experiment is

$$S = \{LF, BF, LW, BW\}. \tag{1}$$

From the problem statement, we know that $\mathrm{P}[LF] = 0.5$, $\mathrm{P}[BF] = 0.2$ and $\mathrm{P}[BW] = 0.2$. This implies $\mathrm{P}[LW] = 1 - 0.5 - 0.2 - 0.2 = 0.1$. The questions can be answered using Theorem 1.5.

(a) The probability that a program is slow is

$$\mathrm{P}[W] = \mathrm{P}[LW] + \mathrm{P}[BW] = 0.1 + 0.2 = 0.3. \tag{2}$$

(b) The probability that a program is big is

$$\mathrm{P}[B] = \mathrm{P}[BF] + \mathrm{P}[BW] = 0.2 + 0.2 = 0.4. \tag{3}$$

(c) The probability that a program is slow or big is

$$\mathrm{P}[W \cup B] = \mathrm{P}[W] + \mathrm{P}[B] - \mathrm{P}[BW] = 0.3 + 0.4 - 0.2 = 0.5. \tag{4}$$

1.3.7 Shuffle a deck of cards and turn over the first card. What is the probability that the first card is a heart?

Problem 1.3.7 Solution

A reasonable probability model that is consistent with the notion of a shuffled deck is that each card in the deck is equally likely to be the first card. Let H_i denote the event that the first card drawn is the ith heart where the first heart is the ace, the second heart is the deuce and so on. In that case, $\mathrm{P}[H_i] = 1/52$ for $1 \le i \le 13$. The event H that the first card is a heart can be written as the mutually exclusive union

$$H = H_1 \cup H_2 \cup \cdots \cup H_{13}. \tag{1}$$

Using Theorem 1.1, we have

$$\mathrm{P}[H] = \sum_{i=1}^{13} \mathrm{P}[H_i] = 13/52. \tag{2}$$

This is the answer you would expect since 13 out of 52 cards are hearts. The point to keep in mind is that this is not just the common sense answer but is the result of a probability model for a shuffled deck and the axioms of probability.

1.3.9 A student's score on a 10-point quiz is equally likely to be any integer between 0 and 10. What is the probability of an A, which requires the student to get a score of 9 or more? What is the probability the student gets an F by getting less than 4?

Problem 1.3.9 Solution

Let s_i equal the outcome of the student's quiz. The sample space is then composed of all the possible grades that she can receive.

$$S = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}. \quad (1)$$

Since each of the 11 possible outcomes is equally likely, the probability of receiving a grade of i, for each $i = 0, 1, \ldots, 10$ is $\mathrm{P}[s_i] = 1/11$. The probability that the student gets an A is the probability that she gets a score of 9 or higher. That is

$$\mathrm{P}[\text{Grade of A}] = \mathrm{P}[9] + \mathrm{P}[10] = 1/11 + 1/11 = 2/11. \quad (2)$$

The probability of failing requires the student to get a grade less than 4.

$$\begin{aligned} \mathrm{P}[\text{Failing}] &= \mathrm{P}[3] + \mathrm{P}[2] + \mathrm{P}[1] + \mathrm{P}[0] \\ &= 1/11 + 1/11 + 1/11 + 1/11 = 4/11. \end{aligned} \quad (3)$$

1.3.11 Use Theorem 1.4 to prove by induction the *union bound*: For any collection of events $A_1, \ldots, A_n$,

$$\mathrm{P}\left[A_1 \cup A_2 \cup \cdots \cup A_n\right] \leq \sum_{i=1}^{n} \mathrm{P}\left[A_i\right].$$

Problem 1.3.11 Solution

Specifically, we will use Theorem 1.4(c) which states that for any events A and B,

$$\mathrm{P}\left[A \cup B\right] = \mathrm{P}\left[A\right] + \mathrm{P}\left[B\right] - \mathrm{P}\left[A \cap B\right]. \tag{1}$$

To prove the union bound by induction, we first prove the theorem for the case of $n = 2$ events. In this case, by Theorem 1.4(c),

$$\mathrm{P}\left[A_1 \cup A_2\right] = \mathrm{P}\left[A_1\right] + \mathrm{P}\left[A_2\right] - \mathrm{P}\left[A_1 \cap A_2\right]. \tag{2}$$

By the first axiom of probability, $\mathrm{P}[A_1 \cap A_2] \geq 0$. Thus,

$$\mathrm{P}\left[A_1 \cup A_2\right] \leq \mathrm{P}\left[A_1\right] + \mathrm{P}\left[A_2\right]. \tag{3}$$

which proves the union bound for the case $n = 2$.Now we make our induction hypothesis that the union-bound holds for any collection of $n-1$ subsets. In this case, given subsets $A_1, \ldots, A_n$, we define

$$A = A_1 \cup A_2 \cup \cdots \cup A_{n-1}, \qquad B = A_n. \tag{4}$$

By our induction hypothesis,

$$\mathrm{P}\left[A\right] = \mathrm{P}\left[A_1 \cup A_2 \cup \cdots \cup A_{n-1}\right] \leq \mathrm{P}\left[A_1\right] + \cdots + \mathrm{P}\left[A_{n-1}\right]. \tag{5}$$

This permits us to write

$$\begin{aligned}
\mathrm{P}\left[A_1 \cup \cdots \cup A_n\right] &= \mathrm{P}\left[A \cup B\right] \\
&\leq \mathrm{P}\left[A\right] + \mathrm{P}\left[B\right] \qquad \text{(by the union bound for } n = 2\text{)} \\
&= \mathrm{P}\left[A_1 \cup \cdots \cup A_{n-1}\right] + \mathrm{P}\left[A_n\right] \\
&\leq \mathrm{P}\left[A_1\right] + \cdots \mathrm{P}\left[A_{n-1}\right] + \mathrm{P}\left[A_n\right]
\end{aligned} \tag{6}$$

which completes the inductive proof.

1.3.13 Using the three axioms of probability and the fact that $\mathrm{P}[\varnothing] = 0$, prove Theorem 1.3. Hint: Define $A_i = B_i$ for $i = 1, \ldots, m$ and $A_i = \varnothing$ for $i > m$.

.3.13 Solution

Following the hint, we define the set of events $\{A_i | i = 1, 2, \ldots\}$ such that $i = 1, \ldots, m$, $A_i = B_i$ and for $i > m$, $A_i = \phi$. By construction, $\cup_{i=1}^{m} B_i = \cup_{i=1}^{\infty} A_i$. Axiom 3 then implies

$$\mathrm{P}\left[\cup_{i=1}^{m} B_i\right] = \mathrm{P}\left[\cup_{i=1}^{\infty} A_i\right] = \sum_{i=1}^{\infty} \mathrm{P}\left[A_i\right]. \tag{1}$$

For $i > m$, $\mathrm{P}[A_i] = \mathrm{P}[\phi] = 0$, yielding the claim $\mathrm{P}[\cup_{i=1}^{m} B_i] = \sum_{i=1}^{m} \mathrm{P}[A_i] = \sum_{i=1}^{m} \mathrm{P}[B_i]$.

Note that the fact that $\mathrm{P}[\phi] = 0$ follows from Axioms 1 and 2. This problem is more challenging if you just use Axiom 3. We start by observing

$$\mathrm{P}\left[\cup_{i=1}^{m} B_i\right] = \sum_{i=1}^{m-1} \mathrm{P}\left[B_i\right] + \sum_{i=m}^{\infty} \mathrm{P}\left[A_i\right]. \tag{2}$$

Now, we use Axiom 3 again on the countably infinite sequence $A_m, A_{m+1}, \ldots$ to write

$$\sum_{i=m}^{\infty} \mathrm{P}\left[A_i\right] = \mathrm{P}\left[A_m \cup A_{m+1} \cup \cdots\right] = \mathrm{P}\left[B_m\right]. \tag{3}$$

Thus, we have used just Axiom 3 to prove Theorem 1.3:

$$\mathrm{P}\left[\cup_{i=1}^{m} B_i\right] = \sum_{i=1}^{m} \mathrm{P}\left[B_i\right]. \tag{4}$$

1.4.1 Mobile telephones perform *handoffs* as they move from cell to cell. During a call, a telephone either performs zero handoffs (H_0), one handoff (H_1), or more than one handoff (H_2). In addition, each call is either long (L), if it lasts more than three minutes, or brief (B). The following table describes the probabilities of the possible types of calls.

	H_0	H_1	H_2
L	0.1	0.1	0.2
B	0.4	0.1	0.1

(a) What is the probability that a brief call will have no handoffs?

(b) What is the probability that a call with one handoff will be long?

(c) What is the probability that a long call will have one or more handoffs?

Problem 1.4.1 Solution

Each question requests a conditional probability.

(a) Note that the probability a call is brief is

$$\mathrm{P}[B] = \mathrm{P}[H_0 B] + \mathrm{P}[H_1 B] + \mathrm{P}[H_2 B] = 0.6. \quad (1)$$

The probability a brief call will have no handoffs is

$$\mathrm{P}[H_0|B] = \frac{\mathrm{P}[H_0 B]}{\mathrm{P}[B]} = \frac{0.4}{0.6} = \frac{2}{3}. \quad (2)$$

(b) The probability of one handoff is $\mathrm{P}[H_1] = \mathrm{P}[H_1 B] + \mathrm{P}[H_1 L] = 0.2$. The probability that a call with one handoff will be long is

$$\mathrm{P}[L|H_1] = \frac{\mathrm{P}[H_1 L]}{\mathrm{P}[H_1]} = \frac{0.1}{0.2} = \frac{1}{2}. \quad (3)$$

(c) The probability a call is long is $\mathrm{P}[L] = 1 - \mathrm{P}[B] = 0.4$. The probability that a long call will have one or more handoffs is

$$\mathrm{P}[H_1 \cup H_2|L] = \frac{\mathrm{P}[H_1 L \cup H_2 L]}{\mathrm{P}[L]} = \frac{\mathrm{P}[H_1 L] + \mathrm{P}[H_2 L]}{\mathrm{P}[L]} = \frac{0.1 + 0.2}{0.4} = \frac{3}{4}. \quad (4)$$

1.4.3 You have a shuffled deck of three cards: 2, 3, and 4. You draw one card. Let C_i denote the event that card i is picked. Let E denote the event that the card chosen is a even-numbered card.

(a) What is $\mathrm{P}[C_2|E]$, the probability that the 2 is picked given that an even-numbered card is chosen?

(b) What is the conditional probability that an even-numbered card is picked given that the 2 is picked?

Problem 1.4.3 Solution

Since the 2 of clubs is an even numbered card, $C_2 \subset E$ so that $\mathrm{P}[C_2E] = \mathrm{P}[C_2] = 1/3$. Since $\mathrm{P}[E] = 2/3$,

$$\mathrm{P}[C_2|E] = \frac{\mathrm{P}[C_2E]}{\mathrm{P}[E]} = \frac{1/3}{2/3} = 1/2. \quad (1)$$

The probability that an even numbered card is picked given that the 2 is picked is

$$\mathrm{P}[E|C_2] = \frac{\mathrm{P}[C_2E]}{\mathrm{P}[C_2]} = \frac{1/3}{1/3} = 1. \quad (2)$$

1.4.5 The basic rules of genetics were discovered in mid-1800s by Mendel, who found that each characteristic of a pea plant, such as whether the seeds were green or yellow, is determined by two genes, one from each parent. In his pea plants, Mendel found that yellow seeds were a dominant trait over green seeds. A yy pea with two yellow genes has yellow seeds; a gg pea with two recessive genes has green seeds; a hybrid gy or yg pea has yellow seeds. In one of Mendel's experiments, he started with a parental generation in which half the pea plants were yy and half the plants were gg. The two groups were crossbred so that each pea plant in the first generation was gy. In the second generation, each pea plant was equally likely to inherit a y or a g gene from each first-generation parent. What is the probability $\mathrm{P}[Y]$ that a randomly chosen pea plant in the second generation has yellow seeds?

Problem 1.4.5 Solution

The first generation consists of two plants each with genotype yg or gy. They are crossed to produce the following second generation genotypes, $S = \{yy, yg, gy, gg\}$. Each genotype is just as likely as any other so the probability of each genotype is consequently $1/4$. A pea plant has yellow seeds if it possesses at least one dominant y gene. The set of pea plants with yellow seeds is

$$Y = \{yy, yg, gy\}. \tag{1}$$

So the probability of a pea plant with yellow seeds is

$$\mathrm{P}[Y] = \mathrm{P}[yy] + \mathrm{P}[yg] + \mathrm{P}[gy] = 3/4. \tag{2}$$

1.4.8 Deer ticks can carry both Lyme disease and human granulocytic ehrlichiosis (HGE). In a study of ticks in the Midwest, it was found that 16% carried Lyme disease, 10% had HGE, and that 10% of the ticks that had either Lyme disease or HGE carried both diseases.

(a) What is the probability P$[LH]$ that a tick carries both Lyme disease (L) and HGE (H)?

(b) What is the conditional probability that a tick has HGE given that it has Lyme disease?

Problem 1.5.6 Solution

The problem statement yields the obvious facts that $P[L] = 0.16$ and $P[H] = 0.10$. The words "10% of the ticks that had either Lyme disease or HGE carried both diseases" can be written as

$$P[LH|L \cup H] = 0.10. \tag{1}$$

(a) Since $LH \subset L \cup H$,

$$P[LH|L \cup H] = \frac{P[LH \cap (L \cup H)]}{P[L \cup H]} = \frac{P[LH]}{P[L \cup H]} = 0.10. \tag{2}$$

Thus,

$$P[LH] = 0.10P[L \cup H] = 0.10\,(P[L] + P[H] - P[LH])\,. \tag{3}$$

Since $P[L] = 0.16$ and $P[H] = 0.10$,

$$P[LH] = \frac{0.10\,(0.16 + 0.10)}{1.1} = 0.0236. \tag{4}$$

(b) The conditional probability that a tick has HGE given that it has Lyme disease is

$$P[H|L] = \frac{P[LH]}{P[L]} = \frac{0.0236}{0.16} = 0.1475. \tag{5}$$

1.5.1 Given the model of handoffs and call lengths in Problem 1.4.1,

(a) What is the probability $P[H_0]$ that a phone makes no handoffs?

(b) What is the probability a call is brief?

(c) What is the probability a call is long or there are at least two handoffs?

Problem 1.5.1 Solution

From the table we look to add all the mutually exclusive events to find each probability.

(a) The probability that a caller makes no hand-offs is

$$P[H_0] = P[LH_0] + P[BH_0] = 0.1 + 0.4 = 0.5. \quad (1)$$

(b) The probability that a call is brief is

$$P[B] = P[BH_0] + P[BH_1] + P[BH_2] = 0.4 + 0.1 + 0.1 = 0.6. \quad (2)$$

(c) The probability that a call is long or makes at least two hand-offs is

$$\begin{aligned} P[L \cup H_2] &= P[LH_0] + P[LH_1] + P[LH_2] + P[BH_2] \\ &= 0.1 + 0.1 + 0.2 + 0.1 = 0.5. \end{aligned} \quad (3)$$

1.6.1 Is it possible for A and B to be independent events yet satisfy $A = B$?

Problem 1.6.1 Solution

This problems asks whether A and B can be independent events yet satisfy $A = B$? By definition, events A and B are independent if and only if $\mathrm{P}[AB] = \mathrm{P}[A]\,\mathrm{P}[B]$. We can see that if $A = B$, that is they are the same set, then

$$\mathrm{P}[AB] = \mathrm{P}[AA] = \mathrm{P}[A] = \mathrm{P}[B]. \tag{1}$$

Thus, for A and B to be the same set and also independent,

$$\mathrm{P}[A] = \mathrm{P}[AB] = \mathrm{P}[A]\,\mathrm{P}[B] = (\mathrm{P}[A])^2. \tag{2}$$

There are two ways that this requirement can be satisfied:

- $\mathrm{P}[A] = 1$ implying $A = B = S$.
- $\mathrm{P}[A] = 0$ implying $A = B = \phi$.

1.6.5 In an experiment, A and B are mutually exclusive events with probabilities $P[A] = 1/4$ and $P[B] = 1/8$.

(a) Find $P[A \cap B]$, $P[A \cup B]$, $P[A \cap B^c]$, and $P[A \cup B^c]$.

(b) Are A and B independent?

1.6.5 Solution

(a) Since A and B are mutually exclusive, $P[A \cap B] = 0$. Since $P[A \cap B] = 0$,

$$P[A \cup B] = P[A] + P[B] - P[A \cap B] = 3/8. \quad (1)$$

A Venn diagram should convince you that $A \subset B^c$ so that $A \cap B^c = A$. This implies

$$P[A \cap B^c] = P[A] = 1/4. \quad (2)$$

It also follows that $P[A \cup B^c] = P[B^c] = 1 - 1/8 = 7/8$.

(b) Events A and B are dependent since $P[AB] \neq P[A]P[B]$.

1.6.7 In an experiment, A and B are mutually exclusive events with probabilities $P[A \cup B] = 5/8$ and $P[A] = 3/8$.

(a) Find $P[B]$, $P[A \cap B^c]$, and $P[A \cup B^c]$.

(b) Are A and B independent?

Problem 1.6.7 Solution

(a) Since $A \cap B = \emptyset$, $P[A \cap B] = 0$. To find $P[B]$, we can write

$$P[A \cup B] = P[A] + P[B] - P[A \cap B] \quad (1)$$

or

$$5/8 = 3/8 + P[B] - 0. \quad (2)$$

Thus, $P[B] = 1/4$. Since A is a subset of B^c, $P[A \cap B^c] = P[A] = 3/8$. Furthermore, since A is a subset of B^c, $P[A \cup B^c] = P[B^c] = 3/4$.

(b) The events A and B are dependent because

$$P[AB] = 0 \neq 3/32 = P[A]P[B]. \quad (3)$$

1.6.11 For independent events A and B, prove that

(a) A and B^c are independent.

(b) A^c and B are independent.

(c) A^c and B^c are independent.

Problem 1.6.11 Solution

(a) For any events A and B, we can write the law of total probability in the form of

$$\mathrm{P}[A] = \mathrm{P}[AB] + \mathrm{P}[AB^c]. \quad (1)$$

Since A and B are independent, $\mathrm{P}[AB] = \mathrm{P}[A]\,\mathrm{P}[B]$. This implies

$$\mathrm{P}[AB^c] = \mathrm{P}[A] - \mathrm{P}[A]\,\mathrm{P}[B] = \mathrm{P}[A](1 - \mathrm{P}[B]) = \mathrm{P}[A]\,\mathrm{P}[B^c]. \quad (2)$$

Thus A and B^c are independent.

(b) Proving that A^c and B are independent is not really necessary. Since A and B are arbitrary labels, it is really the same claim as in part (a). That is, simply reversing the labels of A and B proves the claim. Alternatively, one can construct exactly the same proof as in part (a) with the labels A and B reversed.

(c) To prove that A^c and B^c are independent, we apply the result of part (a) to the sets A and B^c. Since we know from part (a) that A and B^c are independent, part (b) says that A^c and B^c are independent.