Mathematical Statistics

Gregg Waterman
Oregon Institute of Technology

Contents

0 Introduction to This Book

0.1 Goals and Essential Questions

The title of this book is perhaps misleading, as there is no statistics within. It is instead a fairly straightforward introduction to mathematical probability, which is the foundation of mathematical statistics. One could follow this course with a rigorous treatment of statistics, beyond that usually seen in most introductory statistics courses.

Our study of the subject of probability will be guided by some overarching goals, and essential questions related to those goals.

Goals

Upon completion of his/her study, the student should

- understand the basic properties of sets, functions, infinite sums and integrals as they apply to the study of probability,
- be able to solve problems using principles of counting and classical probability,
- understand probability functions and cumulative probability functions for discrete, continuous and joint distributions,
- be able to apply commonly used distributions to solve problems,
- be able to express understanding and methodology of problem solving using correct and precise notation.

Our pursuit of these goals will take place through the consideration of some related *essential questions.*

Essential Questions:

- What is/are a sample space, events, random variables and probability functions?
- What are probability distribution/density functions, and how do they differ in the discrete and continuous cases?
- What are cumulative probability distribution functions, and how do they differ in the discrete and continuous cases?
- What are the expected value and variance of a distribution?
- What are some commonly used distributions and how are they used to solve real problems?
- What are joint probability distributions?

0.2 To The Student

This textbook is designed to provide you with a basic reference for the topics within. That said, it cannot learn for you, nor can your instructor; ultimately, the responsibility for learning the material lies with you. Before beginning the mathematics, I would like to tell you a little about what research tells us are the best strategies for learning. Here are some of the principles you should adhere to for the greatest success:

- **It's better to recall than to review.** It has been found that re-reading information and examples does little to promote learning. Probably the single most effective activity for learning is attempting to recall information and procedures yourself, rather than reading them or watching someone else do them. The process of trying to recall things you have seen is called *retrieval*.

- **Spaced practice is better than massed practice.** Practicing the same thing over and over (called *massed practice*) is effective for learning very quickly, but it also leads to rapid forgetting as well. It is best to space out, over a period of days and even weeks, your practice of one kind of problem. Doing so will lead to a bit of forgetting that promotes retrieval practice, resulting in more lasting learning. And it has been determined that your brain makes many of its new connections while you sleep!

- **Interleave while spacing.** *Interleaving* refers to mixing up your practice so that you're attempting to recall a variety of information or procedures. Interleaving naturally supports spaced practice.

- **Attempt problems that you have not been shown how to solve.** It is beneficial to attempt things you don't know how to do *if you attempt long enough to struggle a bit.* You will then be more receptive to the correct method of solution when it is presented, or you may discover it yourself!

- **Difficult is better.** You will not strengthen the connections in your brain by going over things that are easy for you. Although your brain is not a muscle, it benefits from being "worked" in a challenging way, just like your body.

- **Connect with what you already know, and try to see the "big picture."** It is rare that you will encounter an idea or a method that is completely unrelated to anything you have already learned. New things are learned better when you see similarities and differences between them and what you already know. Attempting to "see how the pieces fit together" can help strengthen what you learn.

- **Quiz yourself to find out what you *really* do (and don't) know.** Understanding examples done in the book, in class, or on videos can lead to the illusion of knowing a concept or procedure when you really don't. Testing yourself frequently by attempting a variety of exercises without referring to examples is a more accurate indication of the state of your knowledge and understanding. This also provides the added benefit of interleaved retrieval practice.

- **Seek and utilize immediate feedback.** The answers to all of the exercises in the book are in the back. Upon completing any exercise, check your answer right away and correct any misunderstandings you might have. Many of our in-class activities will have answers provided, in one way or another, shortly after doing them.

0.3 Additional Comments

I had no formal education in this subject myself during my years of schooling. Writing this book was, in part, a means for me to develop an understanding of the subject. My lack of expertise in the area on the one hand results in the possible omission of insightful details but, on the other hand, hopefully also results in a treatment of the subject that is understandable to the learner.

My introduction to the subject came from sitting through the course as taught by my colleague Tim Thompson, prior to my first time teaching the course myself. He did a masterful job of weaving the parts of the subject together as an unfolding story, and it is my hope that at least a bit of that is conveyed here. Thanks are also due to a number of students, Jeremiah Lipp and Alex Huettis in particular, for pointing out a number of errors that have since been fixed.

It is somewhat inevitable that there will be some errors in this text that I have not caught. As soon as errors are brought to my attention, I will update the online version of the text to reflect those changes. If you are using a hard copy (paper) version of the text, you can look online if you suspect an error. If it appears that there is an uncorrected error, please send me an e-mail at *gregg.waterman@oit.edu* indicating where to find the error.

Gregg Waterman
Oregon Institute of Technology
October 2015

1 Probability Basics

Performance Criteria:

1. (a) Given a verbal description of an experiment and an event, give sets representing the sample space of the experiment and the event. Be able to do this for experiments involving repeated trials, with selection either with or without replacement.
   (b) Determine whether two events are mutually exclusive or complementary.
   (c) Determine the probability of an event using the classical definition of probability.
   (d) Solve counting problems using combinatorial methods.
   (e) Use the addition rule to determine a probability.
   (f) Use a Venn diagram to model a probability problem.
   (g) Compute a conditional probability.
   (h) Apply the multiplication rule to determine probabilities.
   (i) Determine whether two events are independent.
   (j) Use Bayes' Theorem to determine probabilities.

In this chapter we will introduce the methods of "classical" probability that have been in use for hundreds of years, along with some useful combinatorial principles. These things will provide the underpinnings for our study of more "modern" (1900s) probability theory in the rest of the book.

1.1 Experiments, Outcomes and Events

Performance Criteria:

1. (a) Given a verbal description of an experiment and an event, give sets representing the sample space of the experiment and the event. Be able to do this for experiments involving repeated trials, with selection either with or without replacement.

   (b) Determine whether two events are mutually exclusive or complementary.

The main objective of this course is to understand the concept of probability. The idea of probability is to attach numbers to things that could happen that indicate their likelihoods of happening. A common example is when a weatherman (or woman) says there is a 30% chance of snow they are giving a probability that it will snow. We might be interested in other probabilities, like the probability of getting more than 100 hits on a web page in a 5 minute period, or the probability that a part for something we are making is within some tolerance of a desired value.

In order to study probability, we need some common language that we all use so that we can communicate our ideas clearly and precisely. We will be making use of the following definitions.

- An **experiment** is an act whose result can be summarized by some sort of observation.
- When an experiment is conducted, the results that are observed are called **outcomes** of the experiment.
- The *set* of all possible outcomes of an experiment is called the **sample space** of the experiment, denoted by S. This set will usually play the role of the universal set defined in Appendix C.4.
- Any *subset* of the sample space of an experiment is called an **event**. Note that a subset could contain just one outcome, so in a sense an outcome is also an event; the converse is not necessarily true.

◇ **Example 1.1(a):** An experiment consists of flipping a coin and rolling a die. (A **die** is a single small cube with one to six dots on each face. **Dice** are a pair of die.) The outcomes could be denoted by giving the result on the coin followed by the result on the die: H1, H2, ... The sample space is then $S = \{$H1,H2,H3,H4,H5,H6,T1,T2,T3,T4,T5,T6$\}$. One possible event would be tails on the coin and an odd number on the die. In set notation, this event is the set $\{$T1, T3, T5$\}$

◇ **Example 1.1(b):** The number of hits on a website during a 24 hour period is observed (the experiment). *Theoretically*, the sample space is $S = \{0, 1, 2, 3, ...\}$. The event of at least 500 hundred hits is the set $\{500, 501, 502, ...\}$.

◇ **Example 1.1(c):** The lengths of machine bolts are recorded. Assuming that we could measure them with any degree of accuracy we wished, the sample space would be all real numbers greater than zero. Using interval notation, $S = (0, \infty)$. (In reality, the lengths would fall in some shorter interval, but the interval given is certainly the safest one to give.)

A sample space containing a finite or countable number of outcomes is called a **discrete sample space**; the sample spaces in Examples 1.1(a) and 1.1(b) above are discrete sample spaces. A sample space that is an interval of the real line, or a region in the plane or in three-dimensional space (or any higher dimensional space) is called a **continuous sample space**. The sample space for Example 1.1(c) is continuous.

1. For each of the following exercises, an experiment is given, followed by an event. Give the sets that represent the sample space of the experiment and the event.

   (a) *Experiment:* A coin is flipped once. *Event:* Heads is observed.

   (b) *Experiment:* A coin is flipped three times. *Event:* At least two of the flips result in heads.

   (c) *Experiment:* A coin is flipped repeatedly until a head is obtained. *Event:* A head is obtained in *less than* five flips.

   (d) *Experiment:* The three letters A, B and C are arranged in all possible ways, using each letter exactly once. *Event:* The first letter is B.

2. Consider again the experiment of of flipping a coin and rolling a die.

   (a) Give the event of a tail on the coin and a number less than three on the die.

   (b) Give the event of a tail on the coin or a number less than three on the die.

   (c) Give the event of four or more on the die. (Since there is no mention of the coin, we assume that it can be either heads or tails.)

3. A pair of dice is rolled. Assume that the two die can be distinguished from one another; perhaps one is red and the other green.

   (a) How many outcomes will there be for this experiment?

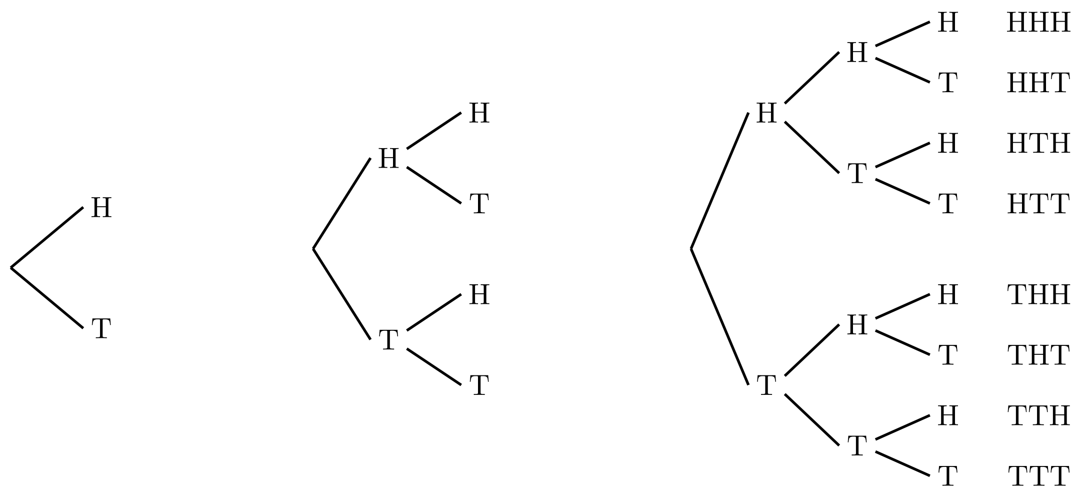
   (b) Each outcome can be denoted by an ordered pair of numbers. Give the event that the sum of the numbers on the dice is eight *as a set of ordered pairs*. (This event has more than three outcomes!)

4. The actual resistance R of a 10 ohm resistor is measured, to the nearest hundredth of an ohm. Give the event that the resistance is greater than 10.12 ohms, using set notation.

5. The amount of time t (in years, allowing parts of years, like 13.4915 years) from when an electrical component starts to be used until it fails is observed. If we assume that the component is functional to begin with, we can describe the sample set as $\{t \in \mathbb{R} \mid t > 0\}$. Give the set notation for the event that the component lasts at least 10 years but no more than 15 years.

There is a tool called a **tree diagram** that will be very useful when considering experiments consisting of a sequence of steps. Consider the experiment consisting of flipping a coin three times in a row, from Exercise 1(b). At the first step, one of two things can happen: we can get heads, or tails. So our tree starts with two branches, on for getting heads on the first flip, one for tails.

See the leftmost picture at the top of the next page. At the second step we can get heads or tails regardless of whether we got heads or tails on the first flip. So each of the original two branches has two branches coming off of it, as shown in the middle picture on the next page. The tree is then finished by giving the branches for the third flip and writing the individual outcomes at the ends of the branches, as shown.



6. (a) Draw a tree diagram for the experiment from Exercise 1(d). *Remember that each letter is only used once.*
   (b) Draw a tree diagram for the experiment from Exercise 2.
   (c) It should be clear that the tree diagram for the experiment from Exercise 1(c) is infinite. Draw the diagram for the first three "branchings", putting $\cdots$ at the ends of any branches that continue on.

Consider again the experiment of flipping a coin until a head is obtained. It is very likely that you represented the sample space of that experiment with the set $\{\text{H, TH, TTH, TTTH, ...}\}$. Note that instead we could represent the sample space with the set $\{1, 2, 3, 4, ...\}$, with each number representing the number of the flip on which the first head is obtained. This set could also be represented by $\{n \in \mathbb{N} \mid n > 0\}$. *Note that this situation is completely analogous to the situation in Exercise 5*, except that the sample space from Exercise 1(c) is a countable set, and the sample space from Exercise 5 is an uncountable set.

When the same process is repeated over and over, like repeatedly flipping a coin, we call each time it is done a **trial**. When a process has two possible outcomes we often refer to one of the outcomes as **success** and the other outcome as **failure**. *We will attach no judgement to these words; we might call something a success when in fact it is a failure in the "real world" sense.* Note that if we were to consider getting a head on the coin to be a failure, then each element of the sample space $\{1, 2, 3, ...\}$ for Exercise 1(c) represents the number of trials to failure. Similarly, each real number in the set $\{t \in \mathbb{R} \mid t > 0\}$ of possible times to failure of the electrical component represents the amount of time to failure.

Now look at your sample space from Exercise 1(d), where you found the possible arrangements of a set of objects. Such arrangements are called **permutations** of the objects. Note that this concept comes into play when considering situations like the following.

7. The four members of a famous rock band are Trent, Chance, John and Colin. They are trying to decide what order to come out of the dressing room for a concert in front of thousands of adoring fans. How many ways can they do this? If you feel ambitious, use the letters T, Ch, J and Co to represent the band members and give all possible orders in which they can go onstage.

Two events are called **mutually exclusive** if they are disjoint sets. (Remember that this means they have no elements in common.)

8. For each of the following, an experiment and two events are given. Determine whether the events are mutually exclusive; if they are not, give an outcome that is in both events.

   (a) A coin is flipped three times. The events are (i) a head is obtained on at least two of the tosses and (ii) a tail is obtained on exactly one of the tosses.

   (b) A coin is flipped three times. The events are (i) a head is obtained on at least two of the tosses and (ii) a tail is obtained on more than one of the tosses.

   (c) A pair of dice is rolled. The events are (i) the sum of the numbers on the dice is at least ten and (ii) the number on at least one of the die is a two.

9. (a) An experiment consists of selecting two of the four letters A, B, C and D. A letter cannot be used more than once and different orders of the letters are not considered distinct from each other. Give the sample space of this experiment.

   (b) Consider the same experiment, but with the first letter replaced before selecting the second, so that the same letter can be selected twice. Give the sample space.

It is important that we make a distinction between the two selection processes from the last exercise. When we select items from a group and no item can be selected more than once we say we are selecting **without replacement**. When an item is replaced in the group after it has been selected, we say we are selecting **with replacement**.

10. (a) For the experiment from Exercise 9(a), give the event that at least one of the letters selected is an A.

    (b) For the experiment from Exercise 9(b), give the event that both letters selected are the same.

    (c) For the experiment from Exercise 9(b), give the event that the two letters selected *are not* the same.

Note that the two events from Exercises 10(b) and (c) are mutually exclusive, and that their union is the entire sample space. Such events are called **complementary events**. As sets, they are complementary sets.

11. Five men, Adam, Bill, Clint, Derek and Ed, are broken into two groups, one with two men and one with three. Different orders within each group are not considered distinct. One possible outcome is that Adam and Derek are grouped together, and Bill, Clint and Ed are grouped together. We will denote this outcome by (AD, BCE). Using this notation, give the sample space for this experiment.

NOTE: The act of breaking a set up into smaller sets is called **partitioning** the set. (See Appendix C.8.) Note that

- each of the smaller sets is a subset of the original set,
- no object is in more than one of the subsets,
- the *union* of all of the subsets is the original set.

12. Consider the experiment consisting of partitioning the same set of five men from Exercise 11 into two sets of two and a set of one. Give the sample space of this experiment.

1.2 Probability

Performance Criteria:

1. (c) Determine the probability of an event using the classical definition of probability.

As stated previously, the general idea of **probability** is to assign to an event A a number, denoted by $P(A)$, that measures the "likelihood" that the event will happen. (Note the use of function notation here; probability is a function that takes an event and gives out a number.) To do this we will need to devise some scheme for assigning such numbers, called probabilities. Of course if there are two events A and B with A more likely to happen than B, we would probably like to have $P(A) > P(B)$. Therefore we need some range of values for probabilities, and it would be logical to make the smallest possible value be zero, for events that can't happen. We will take one to be the upper limit of probabilities, the probability of an event that is sure to happen. Since any event of a sample space S is a subset of S and S is the largest subset of itself, we then should have $P(S) = 1$.

Now consider the experiment of flipping a coin twice in a row, with event A denoting the event of obtaining heads on at least one of the flips. Let A_1 be the event of obtaining heads on *exactly one* flip and let A_2 be the event of obtaining heads on *both* flips. Then A_1 and A_2 are mutually exclusive events with $A = A_1 \cup A_2$. Intuitively, we would all probably agree that it should be the case that $P(A) = P(A_1) + P(A_2)$. This should extend to more mutually exclusive sets, so that if $A = A_1 \cup A_2 \cup \cdots \cup A_n$, with all of the sets A_i mutually exclusive with all the others, then $P(A) = P(A_1) + P(A_2) + \cdots + P(A_n)$. This can even be extended to a *countable* sequence of mutually exclusive sets $A_1, A_2, A_3, ...$ (Note that we are then dealing with an infinite sum, requiring some knowledge of the theory of infinite series.)
We will then build our theory of probability on a set of basic rules, called **postulates** or **axioms**, which we take to be true without proof. These axioms are a precise summary of the previous discussion.

Axiom 1: For any event A, $0 \leq P(A) \leq 1$.

Axiom 2: For an experiment with sample space S, $P(S) = 1$.

Axiom 3: If $A_1, A_2, A_3, ..., A_n$ is a finite sequence of mutually exclusive events whose union is A, then

$$P(A) = P(A_1) + P(A_2) + \cdots + P(A_n)$$

and if $A_1, A_2, A_3, ...$ is a countably infinite sequence of mutually exclusive events whose union is A, then

$$P(A) = P(A_1) + P(A_2) + P(A_3) + \cdots$$

Mathematicians generally prefer to build a theory from the fewest axioms necessary, and there is a little bit of redundancy in the first two items above. We will not fuss about that, but will just go on our merry way from here! Once the axioms are established we then attempt to build a structure of consequences that logically follow from the axioms. The first consequence of our axioms can be seen as follows. Each outcome of an experiment is itself an event (well, the singleton set containing that outcome is an event), and those outcomes are obviously mutually exclusive. If we then have a discrete sample space, those outcomes constitute a countable set of events as well, so we can apply Axiom 3 to get the following.

> **Theorem 1.1:** If an experiment has discrete sample space S, then the probability of any event A is the sum of the probabilities of the individual outcomes in A.

Any "rule" that can be deduced logically from axioms or other previously established "rules" is called a **theorem**. (We will number theorems *.*, with the ones digit indicating the chapter in which the theorem can be found, and the tenth's digit indicating which theorem in that chapter it is.) A very useful application of Theorem 1.1 occurs when we have a finite sample space with each outcome having equal probability.

◇ **Example 1.2(a):** If a coin is flipped three times in a row, what is the probability of getting heads exactly once?

The sample space for this experiment is

$$S = \{HHH, HHT, HTH, THH, TTH, THT, HTT, TTT\}$$

and it would seem that all outcomes should be equally likely. Furthermore, each outcome alone can be considered an event as well, and all such events are mutually exclusive. Thus the probability of each outcome must be $\frac{1}{8}$. (Do you see why that is, based on the axioms?) The event $A = \{HTT, THT, TTH\}$ of getting heads exactly once then has probability $\frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{3}{8}$.

Note that in this example $|A| = 3$, $|S| = 8$ and $P(A) = \frac{3}{8}$. Although this does not constitute proof of the following, it does seem to indicate its truth.

> **Theorem 1.2:** Suppose that all outcomes of an experiment with *finite* sample space S are equally likely (have equal probability as events themselves). Then the probability of an event A is
>
> $$P(A) = \frac{|A|}{|S|}\,.$$

NOTE: For the time being every probability should be given in *exact form*. This means fractions or decimals that *have not been rounded*. Unlike in your past mathematical experience, it is not necessary to reduce fractions; leaving them unreduced will often be more revealing about how they were obtained.

1. A coin is flipped three times. Find the probability that a head is obtained at least twice.

2. A pair of dice is rolled.

   (a) How many outcomes does the sample space contain?

(b) Suppose that we want to know the probability that the sum of the numbers on the dice is four. Make a table with the outcomes from one die along the top and the outcomes from the other along the left side. Where a row and column meet, put the sum of the two outcomes for that row and column.

(c) From your table you can now determine the probability that the sum of the numbers on the dice is four. Do this.

(d) What is the probability that the sum of the numbers on the dice is four or less?

(e) *In words*, what is the complement of the event that the sum of the numbers on the dice is four or less? What is the probability of that event?

In answering 2(e) you should have seen (or maybe even used unknowingly) the following.

> **Theorem 1.3:** For any event A with complement A',
>
> $$P(A') = 1 - P(A)\,.$$

As a final note, let's re-examine what probability is in a slightly more formal way. First, recall that the set of all subsets of a set U is called the **power set** of U. Probability is a function defined on the power set of the sample space, that assigns to any subset of the sample space a real number between zero and one. If we let $\mathcal{P}(S)$ be the power set of the sample space, we can denote this idea symbolically by

$$P : \mathcal{P}(S) \to [0, 1]$$

1.3 Counting

> **Performance Criteria:**
>
> 1. (d) Solve counting problems using combinatorial methods.

When computing probabilities using Theorem 1.2 it is necessary to know how many elements are in a sample space, and how many elements are in an event. This is often a challenging problem. The area of mathematics that deals with how many ways things can be done is called **combinatorics** or, more simply, **counting**. In this section you will learn the fundamental operations of counting.

1. Suppose that Mr. Waterman has two pairs of pants, a blue pair and a khaki pair. Suppose also that he has three shirts, one blue, one green and one red. Suppose also that he has no qualms about wearing any color with any other, and that he puts on his pants before his shirt in the mornings.

    (a) How many ways can Mr. W dress?

    (b) Mr. Waterman's wife tells him that he looks ridiculous when he wears the blue shirt with the blue pants. If Mr. W always selects his shirt and pants randomly, what it the probability that he will dress in a way that his wife does not approve of?

The above exercise illustrates the following.

> **Theorem 1.4:** Suppose that a sequential process consists of selecting one of n_1 choices, then selecting one of n_2 choices. Then there are $n_1 \cdot n_2$ ways to complete the process.

Note that it would not have made any difference whether Mr. W put on his shirt before his pants, there would have been the same number of ways for him to dress. This is, of course, because multiplication is commutative! The same idea can, of course, be extended to situations where there are more choices to be made.

> **Theorem 1.5:** Suppose that a sequential process with k steps consists of selecting one of n_1 choices, one of n_2 choices, ... , and one of n_k choices. Then there are $n_1 \cdot n_2 \cdot n_3 \cdots n_k$ ways to complete the process.

2. A state's license plate numbers consist of two letters, followed by three digits. (A letter or digit may be re-used.) For whatever reason, they do not want the last digit to be zero.

    (a) How many different license plates can they make with this scheme?

    (b) How many plates begin with AB and end with the digit 7?

    (c) What is the probability that a randomly selected plate begins with a vowel and ends with an even number? (Remember that the vowels are *a, e, i, o* and *u*.)

3. The **permutations** of a set of distinct objects are all the ways that those objects can be arranged (from a first object to a last object). When thinking about the following, it might be helpful to think of tree diagrams and how many branches there are at each selection step.

   (a) How many permutations of two objects are there?

   (b) How many permutations of three objects are there?

   (c) How many permutations of four objects are there?

   (d) How many permutations of n objects are there?

For $n \geq 1$, the quantity $n \cdot (n-1) \cdot (n-2) \cdots 2 \cdot 1$ is denoted by $n!$, spoken as n **factorial**. The factorial operation is extremely important in combinatorics, and proves useful in other areas of mathematics as well. Because it makes things work out well, we *define* $0!$ to have value one. Because this is a definition, it is not up for debate - that's just the way it is! You will soon see that all the formulas involving factorials make sense with this definition, and would not with any other. Using this notation, the result of the previous exercise is as follows.

Theorem 1.6: The number of permutations of n distinct objects is $n!$.

Often we are interested in situations where we want to know the number of permutations of a few objects selected from a larger group of objects. The following exercise examines such situations.

4. (a) Four contestants, Adam, Bill, Clint and Derek run a race. The top two finishers are recorded, *in order*. How many possible finishes are there?

   (b) A horse race has eight contestants, and only the first three spots are recorded (win, place or show). How many possible top three finishes are possible in a horse race?

   (c) The language we use for the values you found in this exercise is "permutations of four things taken two at a time" and "permutations of eight things taken three at a time". Can you see how to find the number of permutations of 18 things taken 5 at a time?

   (d) Can you see, in general, how to find the number of permutations of n things taken r at a time? Try to express your answer as a factorial divided by another factorial.

In part (d) above you should have come up with this:

Theorem 1.7: The number of permutations of n distinct objects taken r at a time $(0 \leq r \leq n)$ is

$$_nP_r = n \cdot (n-1) \cdots (n-r+1) = \frac{n!}{(n-r)!}\,.$$

If $r = n$ then we just have the number of permutations of n objects.

We will now look at similar sorts of situations, but for which *the order in which the individuals chosen doesn't matter.*

5. (a) Two of the four men from Exercise 4(a) are chosen randomly to help carry a heavy object. Clearly it makes no difference who is chosen first and who is chosen second, they both have to carry the load! How many ways can two of the four be chosen?

(b) Suppose that instead of a horse race, we are selecting three students from a group of eight to give presentations on a given day. It is of no concern who will go first, just who the three are. How many ways can three be chosen? (**Hint:** How many ways could three be chosen *if the order was significant*? Note that this gives the same groups of people several times over. How many times?)

We call the answer to (b) the number of **combinations** of eight things taken three at a time. Another way you might hear this stated is "eight choose three". The hint should have suggested the following to you:

- The number of permutations of eight objects taken three at a time would tell us how many groups of three could be chosen *if order did matter.*
- Since any group of three individuals can be arranged in 3! ways, the same three individuals will show up together in the permutations six times, so we need to divide the number of permutations by 3! to get the number of combinations of eight people taken three at a time.

In general, we have the following theorem.

Theorem 1.8: The number of combinations of n distinct objects taken r at a time ($0 \leq r \leq n$) is

$$ {}_nC_r = \binom{n}{r} = \frac{n!}{r!\,(n-r)!}. $$

There are situations in which we wish to compute n choose n or n choose 0. Since we have defined $0! = 1$, this causes no problems. You will note that r is allowed to be zero in both of the above theorems.

Theorem 1.9: Suppose that there are n objects of k different types; n_1 of the objects are of one type, n_2 of another, and so on. Then there are

$$ \frac{n!}{n_1!\,n_2!\cdots n_k!} $$

distinct permutations of the objects. (Here two arrangements in which two objects of the same type are exchanged are *not* considered distinct.)

6. An urn containing five blue marbles, two yellow marbles and nine red marbles. How many different arrangements of the the marbles are there, if we can't distinguish marbles of the same color from each other?

7. A variation of this that will be important later is this: Suppose that an urn contains red and blue marbles, say at least ten of each. How many ways can we draw three red marbles and five blue marbles, assuming that we can't distinguish between marbles of the same color?

1.4 The Addition Rule

Performance Criteria:

1. (e) Use the addition rule to determine a probability.
   (f) Use a Venn diagram to model a probability problem.

1. Consider the experiment of drawing a card from a deck of cards, with Event F being that a face card (Jack, Queen or King) is drawn and Event E being that an even numbered card is drawn.

   (a) What is the probability of Event F?

   (b) What is the probability of Event E?

   (c) What is the probability of the event $F \cup E$?

   (d) Let A and B be two events for the same experiment. Based on what you have just seen, how do you think $P(A \cup B)$ relates to $P(A)$ and $P(B)$?

2. Again consider the experiment of drawing a random card from a deck of cards. Let Event C be the event that a club is drawn, with Event F again being that a face card is drawn.

   (a) Are Events C and F mutually exclusive? If not, you know what you must do!

   (b) Find the probabilities of each of the two events.

   (c) What is the probability of event $C \cup F$?

   (d) Go back to your answer to 1(d). Do you think it is correct, in general? If not, amend it so that it is. Does the new version hold for the situations of both this exercise *AND* Exercise 1?

In 2(d) you should have arrived at the following, which we sometimes call the **addition rule**.

Theorem 1.11: For events A and B from the same sample space S,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)\,.$$

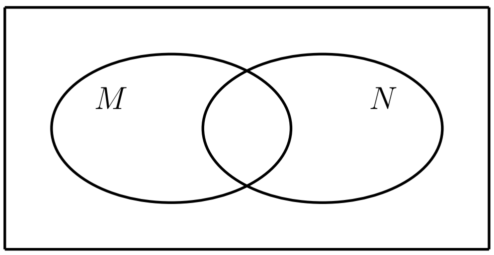
3. The table below shows the distribution of 27 students in one of Mr. Waterman's statistics classes relative to gender and smoking habits. (For example, the number 2 means that two of the students are males who smoke.)

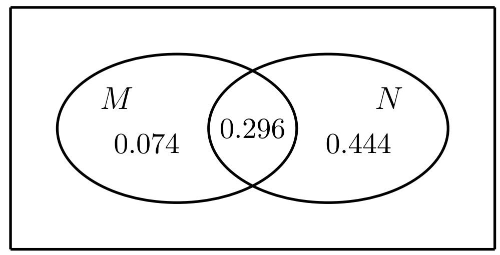
	Male	**Female**	
Smoker	2	5	7
Non-Smoker	8	12	20
	10	17	27

(a) Find the probability of randomly selecting a male or a non-smoker, showing clearly how you obtained your answer.

(b) Can you find other ways to obtain the same answer? There are several more...

Something called a **Venn diagram** can be a useful tool for solving probability problems. (See additional discussion of Venn diagrams in Appendix C.) Consider the data from Exercise 3, with the experiment of randomly selecting a student from the class. If we let M be the event of selecting a male and N be the event of selecting a non-smoker, we can draw the Venn diagram shown below and to the left.The rectangle represents the sample space (the entire class) and the two ovals represent the males and non-smokers. The region where the two ovals overlap represents the individuals who are both male and non-smokers, and the region outside both ovals represents the individuals who are not male and who smoke.





Above and to the right the Venn diagram has been modified by putting in each region the probability of the event it represents. Note that the probability of 0.074 is for male smokers. That is, *it is the probability of being in the oval representing the smokers and NOT in the oval representing the non-smokers*. Similarly the probability of 0.444 is for the non-smoking females, not all of the non-smokers. The probability of 0.296 is for the male non-smokers.

The remaining probability outside both ovals can easily be found and written into the appropriate part of the Venn diagram - this is left to the reader. Putting the probabilities in using this method has the advantage that, once labelled, we can obtain any probability relating to the Venn diagram by simply adding some of the numbers.

4. Use the Venn diagram to find the probability of selecting an individual who

   (a) is a female non-smoker;

   (b) who is female;

   (c) who is a male or a non-smoker, but not both.

5. Draw a Venn diagram for a situation in which $P(A \cup B) = P(A)$. Explain what is happening here in terms of Theorem 1.11.

6. Draw a Venn diagram for a situation in which $P(A \cup B) = P(A) + P(B)$. Explain what is happening here in terms of Theorem 1.11.

7. For an experiment with Events A and B, $P(A) = 0.42$, $P(B) = 0.17$ and $P(A - B) = 0.35$. Find each of the following.

   (a) $P(A \cup B)$

   (b) $P(A \cap B)$

   (c) $P[(A \cup B)']$

8. On a Friday night vehicles are stopped at a roadblock and a police officer has a brief conversation with the driver to try to determine whether they are drunk. For the experiment of randomly selecting a driver, let Event D be that the driver is drunk and Event B be that the

officer believes the driver is drunk and chooses to administer a breathalizer test. Clearly we can have any combination of these events. We are given that

$$P(D) = 0.14, \qquad P(B) = 0.227, \qquad P(D \cap B) = 0.098.$$

Find the probability of each of the following:

(a) The driver is not drunk.

(b) The driver is drunk *OR* the officer administers a breathalizer test.

(c) The driver is not drunk *AND* a breathalizer test is not administered.

1.5 Conditional Probability and Independence

Performance Criteria:

1. (g) Compute a conditional probability.
   (h) Determine whether two events are independent.

1. Consider again the data from Exercise 3 of the last section, shown below, and assume that one student will be drawn at random.

	Male	**Female**	
Smoker	2	5	7
Non-Smoker	8	12	20
	10	17	27

   (a) What is the probability that the student drawn is a smoker?
   (b) Suppose that a student is drawn at random, and we know they are male. What is the probability that the student drawn is a smoker?

The probability that you computed in 1(b) is called a **conditional probability** (because you found the probability of getting a smoker under the condition that the person selected was a male). The language we often use for talking about conditional probabilities is a little bit special; the question of 1(b) will be phrased "What is the probability that a randomly selected student is a smoker, *given that they are a male*?"

2. A die is rolled.
   (a) What is the probability of getting a two, given that a number less than or equal to four was obtained?
   (b) What is the probability of getting a two, given that an even number was obtained?
   (c) What is the probability of getting an even number, given that a two was obtained?

3. Use the data from Exercise 1 above to answer the following.
   (a) What is the probability of selecting a smoker, given that you have selected a female?
   (b) What is the probability of selecting a male, given that you have selected a non-smoker?
   (c) What is the probability of selecting a non-smoker, given that you have selected a male?

Suppose that an experiment has events A and B. We denote the probability of B given A by $P(B|A)$. Note that the last two parts of each of the above exercises illustrate that $P(B|A)$ and $P(A|B)$ are not necessarily the same. (Can you think of an experiment with two events A and B where they *ARE* the same?) We will now consider the question of how $P(B|A)$ is related to $P(A)$, $P(B)$, or other probabilities.

Think about how you computed your answer to 1(b). You probably considered that there are ten males and two of those are smokers, so the probability of selecting a smoker given that you have selected a male is $\frac{2}{10} = \frac{1}{5}$. Suppose that Sm is the event of selecting a smoker (I'm not using S here because it is reserved for the sample space) and M is the event of selecting a male. With just a little thought we can see that the probability of selecting a smoker, given that we've selected a male is

$$\frac{2}{10} = \frac{2/27}{10/27} = \frac{P(Sm \cap M)}{P(M)}\,.$$

We denote this probability by $P(Sm|M)$. In general, we have this definition.

> Let A and B be events with the same sample space, and with $P(A) \neq 0$. Then the **conditional probability of B given A** is
>
> $$P(B|A) = \frac{P(B \cap A)}{P(A)}\,.$$

Clearly there are times, like in Exercises 1-3, when conditional probabilities can be determined without this definition. Sometimes the definition is necessary, however.

4. Recall the information from Exercise 8 of the previous section: On a Friday night vehicles are stopped at a roadblock and a police officer has a brief conversation with the driver to try to determine whether they are drunk. For the experiment of randomly selecting a driver, let Event D be that the driver is drunk and Event B be that the officer believes the driver is drunk and chooses to administer a breathalizer test. Clearly we can have any combination of these events. We are given that

$$P(D) = 0.14, \qquad P(B) = 0.227, \qquad P(D \cap B) = 0.098.$$

   (a) What is the probability that a person is given the breathalizer test, given that they are drunk?
   (b) What is the probability that a person is not drunk, given that they are given the breathalizer test?
   (c) Which of the above probabilities would the police want to be high, and why? Which would the police want to be low, and why?

5. Find $P(A|B)$ and $P(B|A)$ for the situation of Exercise 7 of the previous section.

6. A card is drawn from a standard deck of cards.

   (a) What is the probability that the card is a king?
   (b) What is the probability of selecting a king, given that you have selected a face card?
   (c) What is the probability of selecting a king given that you have selected a heart?

Note that the probabilities for Exercises 6(a) and (c) are the same. When the probability of an event is the same as the probability of that event given a second event, we say that the two events are **independent**. We might then define two events A and B to be independent if $P(B|A) = P(B)$, but this presents a problem with the above definition if A happens to be the null set. We get around this as follows. If two events A and B are independent we then have $P(B|A) = P(B)$, so from the definition of conditional probability we have

$$P(B) = \frac{P(B \cap A)}{P(A)} \qquad \Longrightarrow \qquad P(A \cap B) = P(A)P(B)\,.$$

This leads us to make the following definition.

Two events A and B are **independent events** if

$$P(A \cap B) = P(A)P(B)\,.$$

We can sometimes sense intuitively whether or not two events are independent, by simply asking whether one event occurring should influence the probability that the other will occur. We must still use the above definition to show formally that the two events are independent, however.

7. Consider the experiment of flipping a coin and rolling a die.

   (a) Let H be the event of getting a head on the coin and A be the event of getting a two on the die. Do you think these events are independent? Find $P(H \cap A)$, $P(H)$ and $P(A)$. Are the events of getting a head on the coin and a two on the die independent?

   (b) Let A still be the event of getting a two on the die and let B be the event of getting an even on the die. Do you think these events are independent? Use the definition to determine whether they actually are.

8. A pair of dice are rolled. Let A be the event of getting a two on either or both die. Let B be the event that the sum of the numbers on the die is 7. Do you think these events are independent? Use the definition to find out whether they are.

9. Suppose that an urn contains two blue marbles and three yellow marbles, and two marbles are to be drawn, without replacement. Find the probability of the event of drawing two yellow marbles as follows.

   (a) Let B_1 and B_2 stand for the two blue marbles and Y_1, Y_2 and Y_3 stand for the three yellow marbles. Draw a tree diagram of all possible outcomes for drawing two marbles without replacement.

   (b) Of course each possible outcome has the same probability, so you can determine the desired probability from your tree diagram. Do so.

1.6 The Multiplication Rule

Performance Criteria:

1. (i) Apply the multiplication rule to determine probabilities.

Recall the last exercise from the previous section, in which two marbles are drawn, without replacement, from an urn containing two blue marbles and three yellow marbles. You were asked to find the probability of the event of drawing two yellow marbles. The method used there for finding the desired probability was not difficult, but it would be a bit cumbersome if there have been 17 yellow marbles and 12 blue marbles! In this section we will see a method for working with such probabilities in a simpler way.

1. Suppose again that we have an urn with two blue marbles and three yellow marbles, and that we are again going to select two marbles without replacement.

   (a) Let A be the event of getting a yellow marble on the first selection, and B be the event of getting a yellow marble on the second selection. Do you think the events are independent? Verify your answer using the definition of independent events.

   (b) Find $P(A)$. You should also be able to find the probability $P(B|A)$ without using the formula in the definition. Do so.

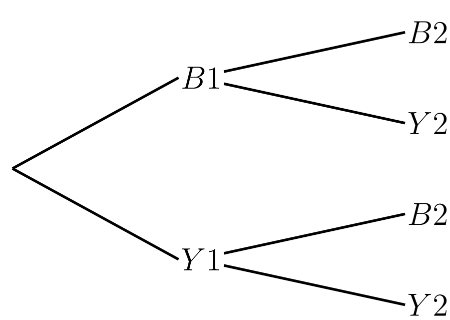
   (c) Note that the event of getting two yellow marbles is the event $A \cap B$. Thus the probability of getting two yellow marbles is $P(A \cap B)$. Insert your answers to (b) into the formula for the definition of conditional probability and solve for $P(A \cap B)$.

The previous exercise illustrates the following, which we call the **multiplication rule**.

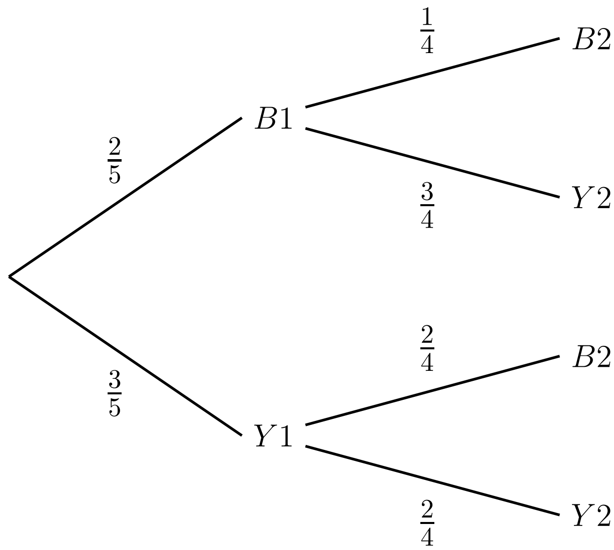
Theorem 1.12: For any two events A and B from the same sample space and with $P(A) \neq 0$,

$$P(A \cap B) = P(A)P(B|A)\,.$$

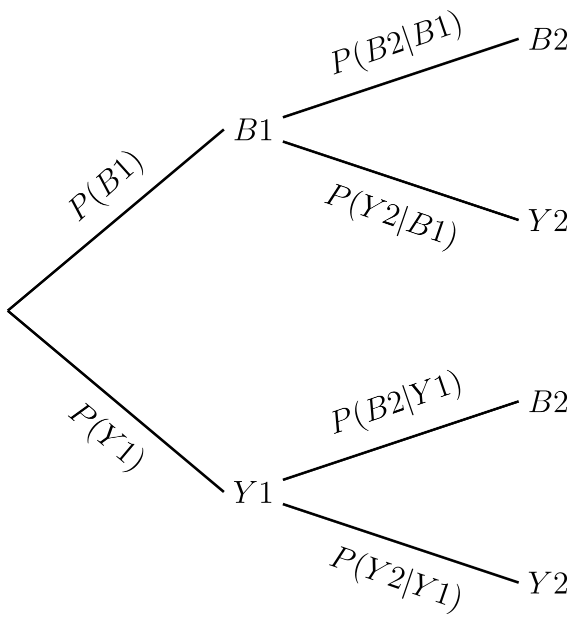
We can use the multiplication rule to make a type of tree diagram that is more efficient for finding probabilities like the one desired in Exercise 1. Such a diagram can be used in any situation where experiments are performed in sequence, or can be thought of as being performed in sequence. For our situation we begin with the following tree diagram that shows what can happen at each step, selecting the first marble and selecting the second. Here $B1$ means getting a blue marble on the first selection, $Y2$ means yellow on the second selection, and so on.



We then put the probability of following each branch of the tree *given that the previous branch has already been followed* along each branch, as shown here:



By the multiplication rule we can then obtain the probability of arriving at the end of any final branch of the tree by multiplying the probabilities along all of the branches that lead to that point:



$$P(B1 \cap B2) = P(B1)P(B2|B1) = \tfrac{2}{20}$$

$$P(B1 \cap Y2) = P(B1)P(Y2|B1) = \tfrac{6}{20}$$

$$P(Y1 \cap B2) = P(Y1)P(B2|Y1) = \tfrac{6}{20}$$

$$P(Y1 \cap Y2) = P(Y1)P(Y2|Y1) = \tfrac{6}{20}$$

2. Use the type of tree diagram just demonstrated to find the probability of selecting two yellow marbles when selecting two marbles without replacement from an urn containing 17 yellow marbles and 12 blue marbles.

1.7 Bayes' Theorem

> **Performance Criteria:**
>
> 1. (j) Use Bayes' Theorem to determine probabilities.

Let's begin this section with a classical and interesting problem: Suppose that it is known that 2 out of every 1000 adult Americans is afflicted with a particular disease. There is a test to determine whether a person has the disease, but it is not perfect. In 99% of the cases where a person has the disease, the test will say that they have it. (We will say the person "tests positive.") It is also known to give a positive result for 3% of the people who do not have the disease. Suppose now that you are diagnosed as having the disease. Due to the imperfections of the test, it is possible that you really do have the disease, but it is also possible that you do not. (When a test indicated that a person has a disease when in fact they don't, it is called a **false positive**.) Our eventual goal is to compute the probability that you really do have the disease.

1. Consider the experiment of randomly selecting an adult American. Let A be the event that a person has the disease and let B be the event that a person tests positive for the disease.

   (a) There are three probabilities given above. Give each of them in terms of the events A and B.

   (b) In terms of the events A and B, what probability is it that we wish to compute? Give the correct "formula" for computing that probability.

   (c) Think of this experiment as a two-step process: At the first step the randomly selected person either has the disease or they don't, and at the second step they either test positive for the disease or they don't. Draw a tree diagram for this situation and label each of the branches with the probability of "travelling" that branch, as we have already done. (For an example, see the tree diagrams in the previous section.)

   (d) Find $P(B)$, the probability that a person tests positive for the disease. Don't round your answer.

   (e) Find the desired probability to three places past the decimal, which is the probability that you really have the disease, given that you tested positive. Do you have cause for immediate concern?

2. The objective of this exercise is to obtain a formula for determining the probability you just found, in terms of the known probabilities.

   (a) Redraw your tree diagram, but instead of labelling the branches whose probabilities were originally known with the probabilities themselves, label them with the symbolic representations of those probabilities, like the second tree diagram in the previous section.

   (b) Use the tree diagram to give an equation for $P(B)$ in terms of the known probabilities. (Note that even though the probability of *not* having the disease was not explicitly given, it is easily found, so consider it given.)

   (c) We were given $P(B|A)$. Give the formula for this probability, and solve it for $P(A \cap B)$.

   (d) The probability that you were looking for was $P(A|B)$, which by definition is $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$. Substitute your results from parts (b) and (c) into this to get the desired formula for $P(A|B)$.

(e) Use your formula from (D) to compute the desired probability, showing a step where each of the known probabilities is substituted into the formula. You should, of course, get the same result as you did for exercise 1!

Your final result from the previous exercise should have been the following.

Theorem 1.13 (Bayes' Theorem): For any two events A and B from the same sample space with $P(A) \neq 0$ and $P(B) \neq 0$,

$$P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A) + P(A')P(B|A')}.$$

Note that if $P(A) = 0$ then $P(A|B) = 0$, and if $P(B) = 0$ then $P(A|B)$ makes no sense.

3. It is known that about 73% of all messages sent to a person's e-mail address are "spam." (I made this number up, so don't repeat it! If anyone knows the correct figure, please let me know.) A particular spam filter detects and removes 97% of all spam messages, but it also removes 5% of the non-spam messages. Determine the probability that a removed message is spam. Do this by either using a tree, or by defining events and applying Bayes' Theorem. If you choose the second method, you might try to define events A and B in such a way that your events correspond to those of the theorem.

4. A small manufacturing operation employs three people Ann, Bob and Cathy for assembly of Widgets. The three work at different speeds, so out of every 1000 Widgets assembled, Ann assembles 355, Bob assembles 314 and Cathy assembles 331. Based on past data, it is known that each of the three assembles the following percentage incorrectly: Ann, 4.7%, Bob 3.1% and Cathy 2.5%.

   (a) What is the probability that a Widget is assembled incorrectly? (Use a tree diagram?)

   (b) Let A, B and C be the events that a randomly selected Widget is assembled is assembled by Ann, Bob or Cathy, respectively, and let I be the event that a Widget is assembled incorrectly. Label each of the given probabilities in terms of this notation. For example, $P(A) = \frac{355}{1000}$.

   (c) Using the notation from (b), give a formula for how you computed $P(I)$ in terms of the given probabilities.

The last exercise illustrated a simple version of the following theorem, which is sometimes called the **total probability rule**.

Theorem 1.14: Let $A_1, A_2, ..., A_n$ be a partition of the sample space of an experiment. Then for any event B for the same experiment,

$$P(B) = P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \cdots + P(A_n)P(B|A_n).$$

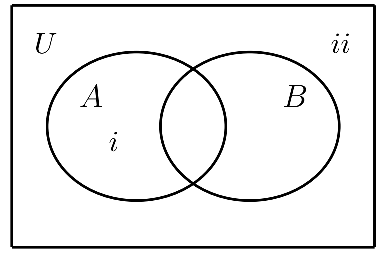
5. Draw a tree for this theorem, indicating missing branches with $\vdots$. Label the branches that are relevant to the theorem with their probabilities, and put the probabilities at the ends of the relevant branches.

6. Consider the situation of Exercise 4. Suppose that a randomly selected Widget is assembled incorrectly. What is the probability that it was assembled by Cathy?

Note that the sets A and A' of Bayes' Theorem partition the sample space S. In the previous exercise the events A, B and C also partition the sample space, so whether you knew it or not you were actually using Bayes' Theorem to do the last exercise, but you were using a "bigger" version. I won't give it here because it is very confusing notationally, but it can be found in most books on probability or mathematical statistics.

1.8 Chapter 1 Exercises

1. The Venn diagram below and to the right is for the experiment of rolling a pair of dice. Event A is the event that at least one of the die is a three and Event B is that the sum of the numbers on the dice is at least seven.



   (a) It should be clear that the two events are not mutually exclusive. Give the event indicated with the letter i on the diagram, *as a set.*
   (b) Give the event indicated by ii on the diagram, *as a set.*
   (c) Give the event from (a) as a union or intersection of two of A, B, A' or B'.
   (d) Give the event from (b) as a union or intersection of A and B or their complements.

2. Suppose that three letters are to be selected from the letters A, B, C, D and E, without regard to order and without replacement.
   (a) Determine how many different ways this can be done by using one of the counting formulas from the previous section. Which theorem did you use?
   (b) Which other theorem also applies?
   (c) Draw a tree diagram illustrating all possible outcomes of this experiment. Write each outcome at the end of the branch of the tree resulting in that outcome.

3. Suppose that we were to instead select three letters from the same set as in the previous exercise (still without replacement), but with different orders of the same three letters being distinct. We wish to determine how many outcomes this experiment has.
   (a) Which theorem from the previous section applies? Use that formula to determine the number of outcomes.
   (b) If you were to draw a tree showing all the outcomes of this experiment it would be rather large. How many branches would it have at the first level? At the second level? at the third?
   (c) There are actually two theorems from the previous section that apply to this scenario. Find the other one and apply it to find the number of outcomes.

4. Determine how many arrangements of three letters from the same list there are if we allow each letter to be used more than once. This is equivalent to asking how many ways can we draw three of the five letters with replacement, if different orders of the same letters are considered distinct. This exercise can be answered using one of the counting formulas.

5. Determine how many arrangements of three letters from the same list there are if we allow each letter to be used more than once, but we don't distinguish between different arrangements of the same three letters?. I don't think that this question can easily be answered using the counting formulas. Try using a tree diagram to answer this question.

6. A class has 31 students. Suppose that I decide to give away \$120 to three students in the class.

   (a) Suppose that I decide to give \$40 to each of three students. How many ways can this be done?

   (b) Suppose that I decide to give \$60 to one student, \$40 to another, and \$20 to a third. How many ways can this be done?

7. In a particular state, licence plates are to consist of a three digit number, followed by a letter. *The first digit of the number is not allowed to be zero.*

   (a) How many different number/letter arrangements are there? (Different orders of the same digits are considered distinct, since they can be distinguished from each other.)

   (b) How many of number/letter arrangements consist of an odd number followed by a consonant?

   (c) A witness to a bank robbery gets a glimpse of the getaway car as it speeds off. They remember that the letter is an L and that the digits of the number are 0, 5 and 8, but they can't remember what order they were in. How many plates do the police need to trace?

8. An urn contains three blue marbles and two yellow marbles. A single marble is drawn repeatedly, *with replacement*, until yellow has been obtained twice.

   (a) This experiment has an infinite, but discrete, sample space. Write it by listing the "first" six events in the sample space, using the notation BYBBY, YBY, etc.

   (b) Find the probability of obtaining a yellow marble twice within the first three draws of a marble. (There are a couple of ways that I can think of to do this. One is to use a probability tree of the sort that we have been working with this week.)

   (c) Suppose now that there is a new urn, containing 30 blue marbles and 20 yellow marbles, and that marbles are drawn *without replacement* until two yellow marbles have been drawn. Find the probability of obtaining a yellow marble twice within the first three draws of a marble.

9. Consider the experiment that a coin is flipped and a four-sided die is rolled.

   (a) Give the sample space for this experiment.

   (b) Suppose that we assign the value one to a flip of the coin resulting in a head, and zero to a flip resulting in a tail. Give the event that the sum of the numbers from the coin and the die is even.

   (c) Determine the probability that the sum of the numbers from the coin and the die is less than three.

   (d) Determine the probability that the sum of the numbers of the coin and the die is greater than three.

10. (a) Give two events A and B from the experiment of the previous exercise for which $P(A \cup B) = P(A) + P(B)$ is true. You may describe the events in words OR with set notation, but you need NOT do both.

    (b) Give two events C and D from the experiment of the previous exercise for which $P(C \cup D) = P(C) + P(D)$ is *NOT* true.

11. Consider the following experiment: A coin is flipped once. If it comes up heads, then a four-sided die (with equal probability of getting any of the values from 1 to 4) is rolled and the experiment is over. If the coin comes up tails, then it is flipped two more times and the experiment is over. So the possible outcomes are

$$H1, H2, H3, H4, TTT, TTH, THT, THH,$$

and *all eight outcomes are equally likely.* Consider also the following events:

Event A: An even number is obtained on the die.

Event B: A tail is obtained on the second flip of the coin.

Event C: A head is obtained on the first flip of the coin.

(a) Draw a probability tree for the diagram with all probabilities labelled appropriately. It should show that all eight outcomes do in fact have equal probabilities.

(b) Find the probability of each of the following.

(i) $P(B) =$ __________ (ii) $P(B \cup C) =$ __________

(iii) $P(B \cap C) =$ __________ (iv) $P(A|C) =$ __________

(v) $P(C|A) =$ __________

(c) Show/explain *mathematically* why Events A and C are not independent.

12. An experiment consists of flipping a coin repeatedly until a head is obtained; The possible outcomes are

$$H, TH, TTH, TTTH, TTTTH, TTTTTH, ...$$

(a) What is the probability of obtaining a head on the first flip?

(b) What is the probability of obtaining a head on the second flip?

(c) What is the probability of obtaining a head on the third flip?

(d) What is the probability that it will take at least three flips to obtain the first head?

13. A coin is flipped once.

(a) Give the sample space for the experiment.

(b) List all possible outcomes. Remember that an outcome is simply a subset of the sample space.

14. A coin is flipped twice.

(a) Give the sample space for the experiment.

(b) List all possible outcomes. Remember that an outcome is simply a subset of the sample space.

15. A coin is flipped three times.

(a) Give the sample space for the experiment.

(b) Give the event that exactly one head is obtained, as a set.

(c) Give the event that at least one head is obtained.

16. The digits 3, 4 and 7 are used to create three digit numbers, using each of those three digits exactly once.

    (a) Give the sample space.
    (b) Give the event that the number is less than 350.
    (c) Give the event that the number is even.

17. Two marbles are drawn randomly, with replacement, from an urn containing five blue marbles and seven red marbles. In this exercise you will determine the probability that exactly one of the marbles is red by using a Venn diagram.

    (a) What is the probability of selecting a red marble on the first draw?
    (b) What is the probability of selecting a red marble on the second draw?
    (c) What is the probability of selecting a red marble on both draws, and why?
    (d) Let Event A be that a red marble is selected on the first draw (but not necessarily on the first draw only), and let Event B be that a red marble is selected on the second draw.
    (e) Sketch a Venn diagram illustrating the sample space and events given. Fill each region with its probability as an event.
    (f) Use your Venn diagram to determine the probability of getting exactly one red marble.
    (g) Use your Venn diagram to determine the probability of getting no red marbles.

18. Two marbles are drawn randomly, with replacement, from an urn containing five blue marbles and seven red marbles. In this exercise you will determine the probability that exactly one of the marbles is red by using a tree diagram.

    (a) Draw a tree diagram for this experiment, labeling each branch with its probability.
    (b) Give the outcome at the end of each branch (like RB) and give its probability.
    (c) Use your tree diagram to determine the probability of getting exactly one red marble.
    (d) Use your tree diagram to determine the probability of getting no red marbles.

19. Two marbles are drawn randomly, with replacement, from an urn containing five blue marbles and seven red marbles. In this exercise you will determine the probability that exactly one of the marbles is red by using Theorems 1.4 and 1.2.

    (a) Assuming that the 12 marbles being used can all be distinguished from each other, how many different ways can two marbles be selected? *Remember that marbles are being drawn WITH replacement.* Use Theorem 1.4.
    (b) How many of those ways consist of a red marble followed by a blue marble?
    (c) How many of those ways consist of a blue marble followed by a red marble?
    (d) Use Theorem 1.2 to determine the probability of selecting exactly one red marble.

20. An urn contains three blue marbles and two red marbles. An experiment consists of drawing two marbles *without replacement.* What is the probability of having gotten blue for the first marble, given that you got blue for the second?

2 Probability Distributions

Performance Criteria:

2. Understand random variables and basic principles of discrete and continuous probability distributions.

   (a) Give the range of a random variable; determine the value of a random variable for a given outcome.

   (b) Describe an event in terms of a random variable.

   (c) Find the probability of an event that is described in terms of a random variable.

   (d) Give values of a discrete probability distribution function and a discrete cumulative discrete probability function associated with a random variable.

   (e) Find probabilities from either a discrete probability function or a discrete cumulative probability distribution.

   (f) Determine the value of a constant that makes a function defined on a discrete set a probability distribution function.

   (g) Given a discrete probability distribution function, find the associated cumulative distribution, or vice-versa.

   (h) Determine probabilities from a histogram.

   (i) Find probabilities from either a continuous probability density function or a continuous cumulative probability distribution.

   (j) Determine the value of a constant that makes a function defined on a continuous set a probability distribution function.

   (k) Given a continuous probability density function, find the associated cumulative distribution, or vice-versa.

   (l) Determine the mean, variance or standard deviation of a random variable.

We will now begin our study of random variables and probability functions, which are the basis for "modern" probability theory.

2.1 Random Variables

Performance Criteria:

2. (a) Give the range of a random variable; determine the value of a random variable for a given outcome.
   (b) Describe an event in terms of a random variable.
   (c) Find the probability of an event that is described in terms of a random variable.

For many experiments there is some logical way to assign a number to each outcome.

◇ **Example 2.1(a):** For the experiment of flipping a coin three times in a row we could assign to each outcome the number of heads obtained.

◇ **Example 2.1(b):** For the experiment of rolling a pair of dice, one red and one green, we could assign to each outcome the number showing on the red die.

◇ **Example 2.1(c):** For the experiment of flipping a coin repeatedly until the first head is obtained, we could assign to each outcome the number of the flip on which the head is obtained.

◇ **Example 2.1(d):** For the experiment of randomly selecting a resistor from a batch of resistors from the manufacturer, we assign to each outcome (resistor) its resistance, in ohms.

A function that assigns a real number to each outcome in a sample space is called a **random variable**. Note that the domain of such a function is the sample space of the experiment. The range of the random variable is a set of real numbers; in the case of the first example above, the range is the set $\{0, 1, 2, 3\}$.

1. (a) Give the range of the random variable in Example 2.1(b).
   (b) Give the range of the random variable in Example 2.1(c).
   (c) Give the range of the random variable in Example 2.1(d). (For those who are not electronics folks, the resistance must be greater than zero.)

If the range of a random variable is a finite or countably infinite set of numbers, we call it a **discrete random variable**. If it is some interval of the real line, we call it a **continuous random variable**. The random variables from Example 2.1(a), (b) and (c) are discrete, the random variable from Example 2.1(d) is continuous.

NOTE: Due to the fact that we can only measure with so much accuracy, *practically speaking* all random variables are discrete. We will act as if we could measure to any degree of accuracy when determining whether a random variable is discrete or continuous.

In an algebra or calculus class the most common letter for denoting a function is f. It is common practice to denote random variables by the *upper case* letter X. When we need more letters, we use Y and Z; for the first example above we would write $X(HTH) = 2$.

2. For Example 2.1(b), use the notation (a, b) for the outcomes, where a is the value rolled on the red die and b is the value rolled on the green. Let X be the random variable described in the example, and find $X((2, 5))$.

3. For the same experiment we can define different random variables. For the experiment from Example 2.1(b), define the random variable Y to be the sum of the numbers on the two die.

   (a) What is $Y((2, 5))$?

   (b) Are there other outcomes for which the random variable Y has the same value as you obtained in (a)? If so, list them all.

   (c) List all outcomes for which the random variable Y has value 12.

   (d) List all outcomes for which $Y = 3$.

4. Consider the random variable X for Example 2.1(c), where a coin is flipped repeatedly until the first head is obtained.

   (a) Find $X(TTTTTH)$.

   (b) Are there any two outcomes for which the random variable takes the same value?

We can see from Exercise 3(b) that a random variable, like functions that you are used to (think $f(x) = x^2$) can assign the same value to more than one outcome of an experiment. In some cases they don't, as shown in Exercise 4(b).

At this point you are probably wondering what purpose random variables have! Their purpose will become more apparent in the next section, but for now we make the following observation: *Many events for a given experiment can be described very precisely in terms of a well defined random variable.* Consider the experiment and random variable of Example 2.1(a). The event of getting exactly two heads is described by the statement $X = 2$, and the event of getting at least two heads can be described by $X \geq 2$. For the experiment and random variable of Example 2.1(c), the event of getting the first head somewhere between the fifth and tenth flip (inclusive) is $5 \leq X \leq 10$.

5. For the experiment and random variable of Example 2.1(a), write each of the following events in terms of the random variable.

   (a) Getting one or two heads.

   (b) Getting all tails.

6. For the experiment and random variable of Example 2.1(c), write each of the following events in terms of the random variable.

   (a) Getting the first heads in less than six flips.

   (b) Getting the first heads in exactly three flips.

With the convenience of describing events using random variables, we can also concisely state certain probabilities. Again using Example 2.1(a), the probability of obtaining *exactly* two heads in the three flips of the coin is denoted by $P(X = 2)$. Since there are eight outcomes to the experiment and three of them have exactly two heads, $P(X = 2) = \frac{3}{8}$.

7. Consider again the experiment of rolling a pair of dice, one red and one green, and again let Y be the random variable that assigns to any outcome the sum of the numbers on the two die. Make a six by six table with the numbers 1, 2, .., 6 across the top and down the left side. In each cell, record the value of the random variable Y for the outcome of getting the row number on the red die and the column number on the green die.

8. Use your table from the previous exercise to find each of the following probabilities.

(a) $P(Y = 4)$ (b) $P(Y = 7)$ (c) $P(Y \leq 4)$

(d) $P(Y > 4)$ (e) $P(Y = 15)$ (f) $P(Y \neq 10)$

9. An urn contains three red marbles and seven blue marbles. An experiment consists of randomly selecting four marbles from the urn, with replacement. Let X be the random variable that assigns to each outcome the number of red marbles selected.

(a) What is $\text{Ran}(X)$? (Recall that this means "range of X.")

(b) Find $P(X = 3)$, $P(X \geq 1)$.

10. Consider the same urn as used in the previous exercise, but with the experiment being that four marbles are selected randomly without replacement. Let X again be the random variable that assigns to each outcome the number of red marbles selected.

(a) What is $\text{Ran}(X)$?

(b) Find $P(X = 3)$, $P(X \geq 1)$.

2.2 Discrete Probability Distributions

Performance Criteria:

2. (d) Give values of a discrete probability distribution function and a discrete cumulative discrete probability function associated with a random variable.

   (e) Find probabilities from either a discrete probability function or a discrete cumulative probability distribution.

   (f) Determine the value of a constant that makes a function defined on a discrete set a probability distribution function.

   (g) Given a discrete probability distribution function, find the associated cumulative distribution, or vice-versa.

In this section we introduce two functions associated with a random variable; these functions give us any probability of interest for that random variable. The reason for developing such functions is that they then allow us to use tools from algebra and calculus in dealing with probabilities.

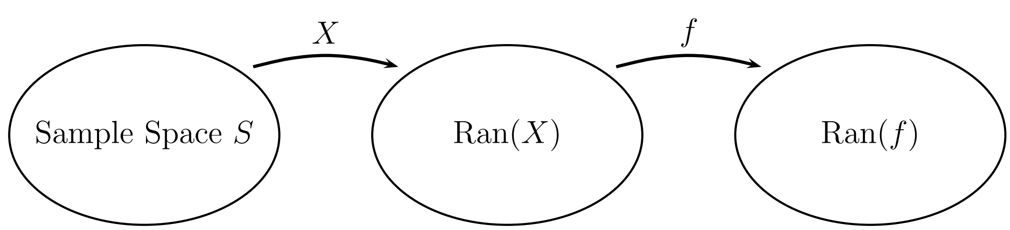
Let X be a discrete random variable defined on the sample space S of an experiment. We define the **probability distribution function** f of X by

$$f(x) = P(X = x)$$

for each x in the range of X.

We will often just say "probability distribution" when we mean "probability distribution function".

The notation here is a bit confusing when you are first learning the concepts of a random variable and probability distribution function. I'll try to make some sense of it, but you should expect to have to give it some deep thought as well! Let's try starting with a picture:



So X is a function that assigns a number to each element of the sample space, and the numbers assigned by X are referred to with the variable x. f is then another function that assigns to each possible value of x another number, which is the probability of getting that value x for the random variable.

◇ **Example 2.2(a):** Suppose that an experiment consists of flipping a coin twice in a row, and let X be the random variable that assigns to each outcome the number of heads. Give the sample space, the range of the random variable X, and the value of f for each value in the range of X.

The sample space is $S = \{TT, TH, HT, HH\}$ and the range of X is $\text{Ran}(X) = \{0, 1, 2\}$. Since the probability of no heads is $\frac{1}{4}$, the above definition gives us that $f(0) = \frac{1}{4}$. Similarly, $f(2) = \frac{1}{4}$ also and $f(1) = \frac{1}{2}$. We will usually summarize a discrete probability function with a table of values like this:

x	0	1	2
$f(x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

(We could write such a table vertically, like you did in an algebra class, but I'll write them horizontally to save space.) Once we have such a table, we can use f to find various probabilities. For example. $P(X = 1) = f(1) = \frac{1}{2}$ and $P(X \geq 1) = f(1) + f(2) = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}$.

There is a very useful way of thinking of probability distribution functions. With the previous example in mind, consider the number line as an "infinite ruler" that has no mass of its own. (To make sure that we are all thinking of the same picture, the positive end of our number line is to the right.) We then place a mass of $\frac{1}{4}$ unit at zero on the number line, a mass of $\frac{1}{2}$ at one, and a mass of $\frac{1}{4}$ at two. Note that the total mass is one unit. (People will sometimes talk about a probability mass function, which just means a probability distribution.)

Using this idea, let's define another function that we will name F. (*Note that our f's are now case sensitive!*) F will have a value for any real number x. To find $F(x)$ for some particular x, we go to the point x on our infinite number line and total up the mass at, and to the left of, x. So, for the above example, $F(-1) = 0$ because there is no mass at $x = -1$ or to its left. $F(\frac{3}{2}) = \frac{1}{4} + \frac{1}{2} = \frac{3}{4}$ because there is no mass at $x = \frac{3}{2}$, but there are masses of $\frac{1}{4}$ and $\frac{1}{2}$ at $x = 0$ and $x = 1$, both of which are to the left of $\frac{3}{2}$.

NOTE: f is only defined for values in the range of the random variable X, which might just be a few numbers, like in the example above. F, on the other hand, *is defined for all real numbers!*

1. For the above example, find each of the following.

   (a) $F(1)$ (b) $F(.83)$ (c) $F(1.99999)$ (d) $F(2)$ (e) $F(73)$

The usefulness of the function F is not necessarily apparent with finite random variables like the one in the above example, but it can be handy for discrete random variables with infinite ranges. The real value of it will be most apparent in the next section, when we develop analogous ideas for continuous random variables.

We will now formalize the previous discussion. Note that the probability distribution function f is defined only on the range of the random variable X. That is, the domain of f is the range of X. We could extend the domain of f to all real numbers by simply defining $f(x) = 0$ for all values of x not in the range of X. Assume that we have done that before we make the following definition.

Let X be a discrete random variable with probability distribution f. We define the **cumulative distribution function** of X by

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t)$$

for each $x \in (-\infty, \infty)$.

Again, F is defined for ALL real numbers. In the discrete case F can only be described in a piecewise manner. The method for determining a description of F is given in the next example.

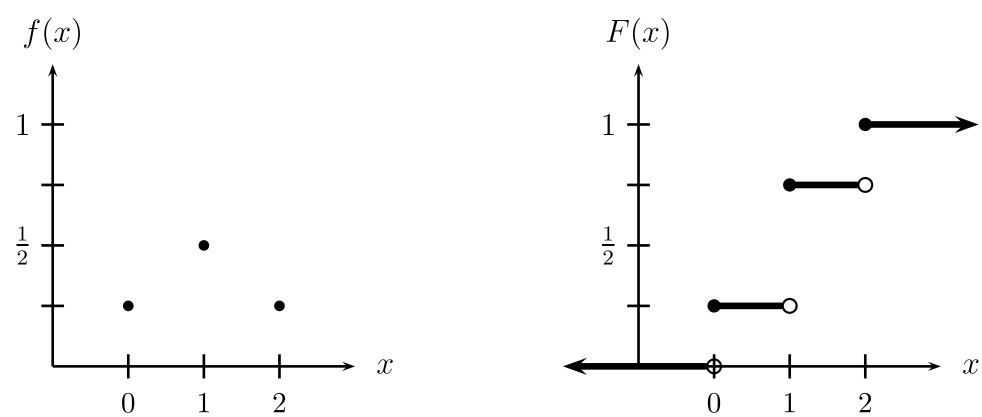
◇ **Example 2.2(b):** Determine the cumulative distribution function F for the random variable and probability distribution from Example 2.2(a).

F must be defined for all real numbers, so let's consider a negative value of x first. To find $F(x)$ we add up all values of $f(t)$ for $t \leq x$. But since zero is the smallest value of t for which $f(t)$, $f(t) = 0$ for all $t \leq x$ and we have $F(x) = 0$ for any negative value of x.

Now if x is greater than or equal to zero, but less than one, $f(t) = 0$ for all $t \leq x$ except $t = 0$, where $f(0) = \frac{1}{4}$. Thus the sum of the values of $f(t)$ for $t \leq x$ is also $\frac{1}{4}$, so $F(x) = \frac{1}{4}$ for $0 \leq x < 1$. Notice that this does not include $x = 1$, where F "picks up" another $\frac{1}{2}$ since $f(1) = \frac{1}{2}$. To summarize, F can be described as follows:

$$F(x) = \begin{cases} 0 & \text{for } x < 0 \\ \frac{1}{4} & \text{for } 0 \leq x < 1 \\ \frac{3}{4} & \text{for } 1 \leq x < 2 \\ 1 & \text{for } x \geq 2 \end{cases}$$

Spend enough time thinking about this example to develop a complete understanding of it. if it is not pretty clear as is, try to work through the details from the the values of f and the definition of F. It is possible to graph f and F, and doing so might illuminate what is going on here. The graphs are shown below, with the graph of f to the left and the graph of F to the right.



2. Consider the experiment of flipping a coin 4 times, with the sample space denoted by

$$\{TTTT, HTTT, THTT, ..., HHHT, HHHH\}.$$

Note that the sample space has $2 \cdot 2 \cdot 2 \cdot 2 = 2^4 = 16$ outcomes. Define the random variable X to be the number of heads obtained in four tosses.

(a) $X(THTH) =$ __________ (b) $X(HHHH) =$ __________

(c) List all possible values of X, from smallest to largest: ________________________

(d) $f(3) = P(X$ ______ $) =$ ______ (e) $F(3) = P(X$ ______ $) =$ ______

(f) $P(X \leq 2) = f(\_\_\_) + f(\_\_\_) + f(\_\_\_) = F(\_\_\_\_\_) = \_\_\_\_\_\_\_$

(g) $P(X \geq 2) = f(\_\_\_) + f(\_\_\_) + f(\_\_\_) = 1 - F(\_\_\_\_\_) = \_\_\_\_\_\_\_$

(h) Write the probability that X is odd in terms of f, and find its value.

(i) Write $P(1 \leq X \leq 3)$ in terms of f, and find its value.

(j) Write $P(1 \leq X \leq 3)$ in terms of F, and find its value.

3. Use the experiment from the previous exercise for the following.

(a) Give a table of values for the probability density function f.

(b) Sketch the graph of f.

(c) Give a piecewise definition of the cumulative distribution function F.

(d) Sketch the graph of F.

The result of Exercise 2(j) illustrates a useful idea.

Theorem 2.1: Let X be a discrete random variable with values $x_0, x_1, x_2, ..., x_n$ and having probability distributions f and F. Then for $1 \leq i < j \leq n$,

$$P(x_i \leq X \leq x_j) = F(x_j) - F(x_{i-1}).$$

In particular,

$$P(X = x_j) = f(x_j) = F(x_j) - F(x_{j-1}).$$

4. A random variable X has the values $x = 0, 1, 2, 3$. The cumulative distribution function $F(x)$ has the following values:

$$F(x) = \begin{cases} 0 & \text{for } x < 0 \\ \frac{3}{15} & \text{for } 0 \leq x < 1 \\ \frac{10}{15} & \text{for } 1 \leq x < 2 \\ \frac{14}{15} & \text{for } 2 \leq x < 3 \\ 1 & \text{for } x \geq 3 \end{cases}$$

Give the probability distribution function f for the random variable X in table form, like done in the first example. The above theorem might be found useful here!

5. Find each of the following for the random variable from Exercise 4.

(a) $P(X \le 2)$ (b) $P(X < 2)$ (c) $P(X = 2)$

(d) $P(X \ge 1)$ (e) $P(X > 1)$ (f) $P(X \le -2)$

(g) $P(X \ge 5)$ (h) $P(X \le 5)$ (i) $P(1 \le X \le 3)$

6. An experiment consists of flipping a coin repeatedly until a head is obtained; The possible outcomes are

$$H, TH, TTH, TTTH, TTTTH, TTTTTH, \ldots$$

The random variable X assigns to each outcome the number of the flip on which the first head occurred so, for example, $X(TTH) = 3$.

(a) Give a table of values for the probability function f for values of x up through 5. Then draw a neat and reasonably sized graph of f for those values.

(b) Give a piecewise definition of the cumulative probability function F for values of x up through 3, then draw its graph.

(c) Give a formula for $f(x)$, $x = 1, 2, 3, 4, \ldots$

(d) Let n be any natural number greater than or equal to one. Give a formula for $F(x)$ when $n \le x < n + 1$.

7. Prove the first part of Theorem 2.1.

It should not be hard to believe the following.

Theorem 2.2: Let X be a discrete random variable having probability distribution f. Then

1) $f(x) \ge 0$ for every $x \in \text{Ran}(X)$

2) $\displaystyle\sum_{\text{Ran}(X)} f(x) = 1$

The above two conditions must be satisfied by the probability distribution function for the discrete random variable X. Conversely, any function satisfying the above two conditions is in fact a probability distribution for the random variable X for which $P(X = x) = f(x)$ for every x in the range of X.

8. Determine the value of c so that $f(x) = cx$, $x = 1, 2, 3, 4$ is a probability distribution function.

9. Determine the value of c so that $f(x) = \dfrac{c}{x}$, $x = 1, 2, 3, 4$ is a probability distribution function.

10. Determine the value of c so that $f(x) = c \begin{pmatrix} 3 \\ x \end{pmatrix} \begin{pmatrix} 5 \\ 2-x \end{pmatrix}, \quad x = 0,\ 1,\ 2$ is a probability distribution function for a discrete random variable X.

The symbols $-\infty$ and ∞ do not represent numbers, so it makes no sense to write something like $f(\infty)$. In the interest of brevity we will define $f(\infty) = \lim_{x\to\infty} f(x)$, and similarly for $-\infty$ and F. Using this notation,

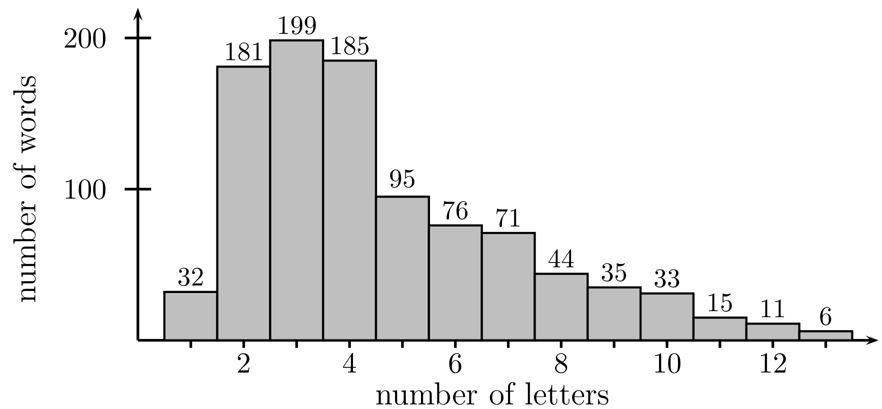
> **Theorem 2.3:** Let X be a discrete random variable with probability distributions f and F. Then
>
> 1) $f(-\infty) = f(\infty) = 0$ and $F(-\infty) = 0,\ F(\infty) = 1$
>
> 2) for any $a < b$, $F(a) \leq F(b)$.

2.3 Data and Histograms

Performance Criteria:

2. (h) Determine probabilities from a histogram.

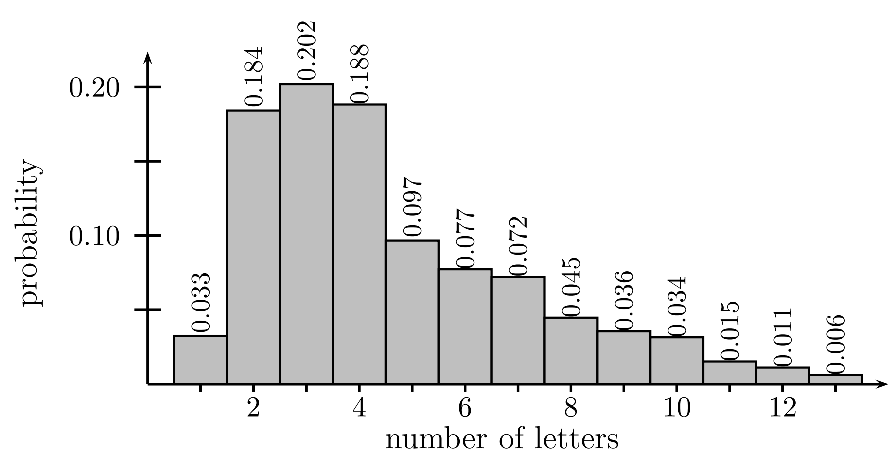
In a display of incredible fortitude, Mr. Waterman counted the number of letters in each of 983 randomly selected words from a statistics book. The graph below shows how many words of each word length there were - such a graph is called a histogram. The number at the top of each column indicates the number of words with that many letters.



1. If we were to randomly select one of the 983 words, what is the probability that it would have

   (a) six letters? (b) more than eight letters? (c) three or more letters?

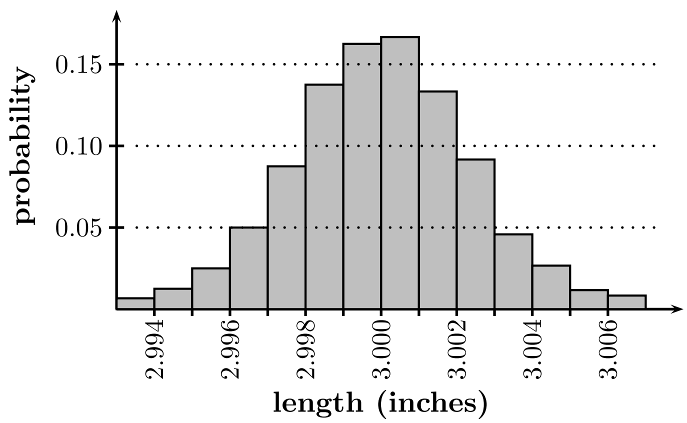
Selecting one of the 983 words randomly is an experiment, with the outcomes being the entire set of words. There is an obvious random variable here; it assigns to each outcome (word) the number of letters in the word. It is a discrete random variable, with values 1, 2, 3, ..., 13. If we were to divide the number of words of each length by the total number of words we would see that $P(X = 1) = \frac{32}{983} \approx 0.0326$, $P(X = 2) = \frac{181}{983} \approx 0.1841$, and so on. (From now on, when stating decimal approximations to probabilities we will simply use =, with the understanding that the value is likely a rounded approximation.) We can create a new histogram, called a probability histogram, where the height of each column indicates the probability of selecting a word with that many letters, rather than the number of words with that many letters:



This, of course, is a graph of the probability distribution function f for the random variable that assigns to a randomly chosen word the number of letters in that word. *Now here is a key idea - rather than thinking of the height of each column indicating the probability of selecting a word with the corresponding number of letters, we want to think of the area of each column as being the probability!*

2. With this interpretation, what is the total shaded area for the graph above?

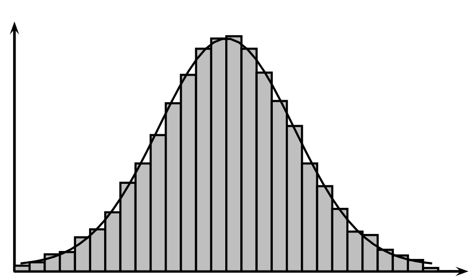
The random variable for the previous situation was a discrete random variable. Now imagine this scenario: We go to a plant where three inch long bolts are made, and we consider the experiment of randomly selecting many bolts, one at a time. We then let X be the random variable that assigns to each bolt its length in inches. Then this is, at least theoretically, a continuous random variable, since its domain is the set of all real numbers greater than zero. Now we certainly can't create a graph with a bar above every real number, like above. Instead we make a probability histogram by designating "classes", or ranges of lengths, and graphing the probabilities of randomly selecting a bolt in each of the classes. Such a histogram is shown above and to the right. *We again think of the area of each column in the histogram as representing the probability of selecting a bolt in that class.*



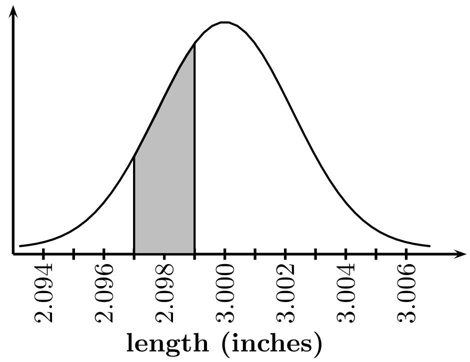
3. *If possible*, determine the approximate probability that the randomly selected bolt has length in the given range. *Assume that all bolts selected are represented in the graph.*

   (a) at least 3.003 inches.
   (b) less than 3.003 inches.
   (c) less than 2.0985 inches.
   (d) between 2.097 and 2.099 inches.
   (e) between 2.093 and 3.007 inches.

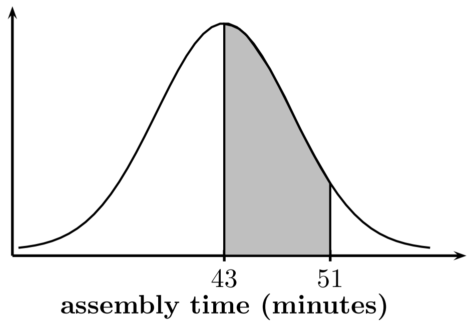
Note that we can draw a continuous curve that approximates the shape of the probability distribution from the above exercise, as shown below and to the left. If we were to gather more data and increase the number of classes so that every bar representing a class was very narrow, we would find that the resulting probability distribution would look even more like the smooth curve, as shown below and to the right.



Recall that the probability of randomly selecting a bolt with length between 2.097 inches and 2.099 inches can be thought of as the sum of the areas of the two bars representing lengths in those ranges. If we were to replace the histogram with a continuous curve, as shown to the right, the shaded area below the continuous curve and between those two values should be a good approximation of the probability of randomly selecting a bolt with a length between 2.097 and 2.099 inches. The continuous curve is an example of what is called a **continuous probability distribution**. The total area under a continuous probability distribution is one, and the probability of randomly obtaining a data value between any two given values is the area under the curve and between those two values.



4. A number of adults have assembled an "easy to assemble" children's toy. The probability distribution for random variable that assigns to a randomly selected adult the length of time it takes them to assemble the toy is shown to the right. *The distribution is symmetric about 43 minutes.* The shaded area is 0.4. Using the distribution, what is the probability that a randomly selected adult will be able to assemble the toy in



(a) between 43 and 51 minutes?

(b) less than 43 minutes?

(c) less than 51 minutes?

(d) more than 51 minutes?

Considering the situation from the above exercise, what would happen to the probabilities you would find for assembling the toy in 43 to 51 minutes, then 43 to 47 minutes, 43 to 44 minutes, 43 to 43.1 minutes, and so on? Clearly they will get smaller and smaller, approaching zero. The probability of randomly selecting an adult who takes *exactly* 43 minutes (or 35 minutes, or 49 minutes, etc.) to assemble the toy is zero.

2.4 Continuous Probability Distributions

Performance Criteria:

2. (i) Find probabilities from either a continuous probability density function or a continuous cumulative probability distribution.

   (j) Determine the value of a constant that makes a function defined on a continuous set a probability distribution function.

   (k) Given a continuous probability density function, find the associated cumulative distribution, or vice-versa.

Based on the discussion at the end of the previous section we see that when working with continuous probability distributions, *we do not want have a distribution function f whose value at any x gives the probability of the random variable taking that value.* Such a function would be zero at all values of x! Instead we define a continuous probability distribution function as follows.

Let X be a continuous random variable defined on the sample space S of an experiment. The **probability density function** of X is a function f for which

$$P(a \leq X \leq b) = \int_a^b f(x)\,dx$$

for all real numbers a and b with $a \leq b$.

Because the probability of a continuous random variable taking an exact value is zero, we have

$$P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b) = P(a < X < b)\,.$$

Recall the physical model of masses on a number line for a discrete probability distribution. The analog for this situation is a number line with a piece of wire, of perhaps varying thickness, laying on it. No single point on the number line has any mass on it; mass only exists in sections of the wire of some length. If the density were constant, then the mass of any piece of the wire would simply be the (linear) density (measured in some units like grams per centimeter) times the length of the piece. Since the density is perhaps variable, we have to compute the mass of any section of the wire using an integral, as defined above. This is the reason for the term probability *density* function.

1. A continuous random variable X has the probability density function

$$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ \frac{1}{2}x & \text{for } 0 \leq x \leq 2 \\ 0 & \text{for } x > 2 \end{cases}$$

   Find each of the following.

   (a) $P(0 \leq X \leq 1)$ (b) $P(\frac{1}{2} \leq X < 2)$ (c) $P(X = 1)$

(d) $P(-1 \le X \le 1)$ Note that since the definition of f changes at $x = 0$, one must compute $\displaystyle\int_{-1}^{0} f(x)\,dx + \int_{0}^{1} f(x)\,dx$ to find this probability.

(e) $P(X \le 3)$ (f) $P(X \le -2)$

2. A continuous random variable X has the probability density function

$$f(x) = \begin{cases} 0 & \text{for } x < 1 \\ \dfrac{1}{x^2} & \text{for } x \ge 1 \end{cases}$$

Find each of the following.

(a) $P(2 \le X \le 4)$ (b) $P(0 \le X < 2)$ (c) $P(X \ge 1)$

(d) $P(X \le 5)$ (e) $P(X > 3)$ (f) $P(X = 7)$

The following theorem is the continuous analog to Theorem 2.2.

> **Theorem 2.4:** Let X be a continuous random variable having probability density f. Then
>
> 1) $f(x) \ge 0$ for every x for all real numbers x,
>
> 2) $\displaystyle\int_{-\infty}^{\infty} f(x)\,dx = 1.$

Like Theorem 2.2, this indicates that any function that satisfies the above two conditions can serve as a probability density function for a continuous random variable.

3. Determine a value of C so that the function is a probability density function.

$$f(x) = \begin{cases} C & \text{for } 0 \le x \le 5 \\ 0 & \text{otherwise} \end{cases}$$

4. Determine a value of C so that the function is a probability density function.

$$f(x) = \begin{cases} Cx & \text{for } 1 \le x \le 4 \\ 0 & \text{otherwise} \end{cases}$$

Consider the probability distribution function from Exercise 2, and let a and b be constants with $1 \le a < b$. Then

$$P(a \le X \le b) = \int_a^b f(x)\,dx = \int_a^b \frac{1}{x^2}\,dx = -\frac{1}{x}\Big]_a^b = \left(-\frac{1}{b}\right) - \left(-\frac{1}{a}\right)$$

This shows that finding the probability $P(a \le X \le b)$ amounts to simply finding the difference between the antiderivative of f at a and at b. That is, if F is such that $F'(x) = f(x)$, then $P(a \le X \le b) = F(b) - F(a)$. This function F is (almost) the cumulative distribution function of the continuous random variable X.

Let X be a continuous random variable with probability density function f. We define the **cumulative distribution function** of X by

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)\,dt$$

for each $x \in (-\infty, \infty)$.

Note that $P(X \ge x) = 1 - P(X < x) = 1 - P(X \le x)$.

◇ **Example 2.4(a):** Find the cumulative probability function F for the density function of Exercise 2.

For $x < 1$ we have $F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_{-\infty}^{x} 0\,dt = 0$.

For $x \ge 1$, $F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_{-\infty}^{1} 0\,dt + \int_{1}^{x} \frac{1}{t^2}\,dt = 0 + \left[-\frac{1}{t}\right]_1^x = 1 - \frac{1}{x}$.

In conclusion then, $F(x) = \begin{cases} 0 & \text{for } x < 1 \\ 1 - \frac{1}{x} & \text{for } x \ge 1 \end{cases}$

---

5. Find the cumulative probability distribution for the probability density function of

(a) $f(x) = \begin{cases} \frac{1}{5} & \text{for } 0 \le x \le 5 \\ 0 & \text{otherwise} \end{cases}$ (b) $f(x) = \begin{cases} 0 & \text{for } x < 0 \\ \frac{1}{2}x & \text{for } 0 \le x \le 2 \\ 0 & \text{for } x > 2 \end{cases}$

The following result follows easily from the definition of the cumulative distribution.

Theorem 2.5: Let X be a continuous random variable having probability density f and cumulative distribution F. Then

$$P(a \le X \le b) = F(b) - F(a).$$

6. A continuous random variable X has the cumulative distribution function

$$F(x) = \begin{cases} 0 & \text{for } x < 0 \\ \frac{\sqrt{x}}{2} & \text{for } 0 \le x \le 4 \\ 1 & \text{for } x > 4 \end{cases}$$

Give each of the following probabilities **in exact form**. (No decimals! Some of your answers will contain square roots.)

(a) $P(X < 3)$ (b) $P(X \geq 1)$ (c) $P(1 < X \leq 3)$

(d) $P(X = 2)$ (e) $P(X \geq -1)$ (f) $P(X \leq -1)$

7. Consider again the probability density function $f(x) = \begin{cases} 0 & \text{for } x < 0 \\ \frac{1}{2}x & \text{for } 0 \leq x \leq 2 \\ 0 & \text{for } x > 2 \end{cases}$

   (a) Graph the density function f, from $x = -2$ to $x = 4$.

   (b) Graph the cumulative distribution function F, from $x = -2$ to $x = 4$.

8. Graph the density function $f(x) = \begin{cases} \frac{1}{5} & \text{for } 0 \leq x \leq 5 \\ 0 & \text{otherwise} \end{cases}$ and its cumulative distribution $F(x)$.

The graphs from the previous exercises can perhaps illuminate what is going on here. The function F evaluated at any point x simply gives the area under f to the left of x. So when x is less than zero, there is no area under the graph of f. As x crosses zero (moving right) it starts picking up area . It continues to pick up area until $x = 2$, after which no new area is accumulated as x continues to move to the right. This illustrates the following theorem.

Theorem 2.6: Let X be a continuous random variable with probability distributions f and F. Then

1) $f(-\infty) = f(\infty) = 0$ and $F(-\infty) = 0$, $F(\infty) = 1$

2) for any $a < b$, $F(a) \leq F(b)$.

9. A continuous random variable X has the cumulative distribution function

$$F(x) = \begin{cases} 0 & \text{for } x \leq 2 \\ c(x-2)^2 & \text{for } 2 < x < 5 \\ 1 & \text{for } x \geq 5 \end{cases}$$

   Find the value of c. *This does not require integration!*

10. Find each of the following for the random variable and cumulative distribution function F from the previous exercise.

   (a) $P(X \leq 4)$ (b) $P(X < 4)$ (c) $P(X = 4)$

   (d) $P(X \geq 3)$ (e) $P(3 \leq X < 4)$ (f) $P(1 < X < 4)$

   (g) $P(X \geq 7)$ (h) $P(X \leq 7)$

The definition of the cumulative distribution function F tells us how to obtain it from the probability density f. Since F is obtained by integrating f, one might suspect that f is obtained from F by differentiating. This is in fact the case.

Theorem 2.7: Let X be a continuous random variable with probability distributions f and F. For all values of x for which F is differentiable we have $f(x) = F'(x)$.

11. Find the probability density function f for the cumulative probability function F from the previous two exercises. Check it by using it to compute $P(3 \leq X < 4)$ and seeing if it agrees with your answer to (e) above.

12. Find the probability density function for the cumulative distribution from Exercise 6.

2.5 Mean and Variance of a Distribution

Performance Criteria:

2. (l) Determine the mean, variance or standard deviation of a random variable.

1. Consider the discrete random variable X with probability distribution

$$\begin{array}{rcccc} x: & 0 & 1 & 2 & 3 \\ f(x): & \frac{1}{16} & \frac{4}{16} & \frac{8}{16} & \frac{3}{16} \end{array}$$

Give the answers to the following in exact (fraction) form.

(a) Find the average of the four numbers 0, 1, 2, 3.

(b) We define the **mean** or **expected value** μ of the discrete random variable X with probability distribution f by

$$\mu = E(X) = \sum_x x\, f(x).$$

Find the expected value of the random variable given.

(c) Sketch the graph of f. Put a vertical arrow $\uparrow$ pointing at the position of μ on the x-axis.

(d) Once we have the expected value of a random variable, we can compute something called the **variance** σ^2 of the random variable, defined by

$$\sigma^2 = E[(X-\mu)^2] = \sum_x (x-\mu)^2\, f(x).$$

Find the variance of the random variable given.

(e) The positive square root of the variance, denoted by σ, is called the **standard deviation** of the random variable X. Find the standard deviation of the given random variable.

So what does all this mean? Suppose that we had an urn with red, blue, green and yellow marbles in it, *equal numbers of each*. You are going to draw one marble at random from the urn; if you draw a red marble you will be given \$0, if you draw a blue you are given \$1, and you are given \$2 for a green and \$3 by a yellow. If you were to draw a whole bunch of times, your average "winnings" would be what you computed in (a) of the above exercise.

Now suppose instead that the urn you were drawing from had one red marble, four blue marbles, eight green marbles and three yellow marbles. (How many marbles are there in all? Note how this situation relates to the probability distribution function from the above exercise.) Then your winnings, on average, would be higher than the previous arrangement, because the probability of winning \$2 or \$3 is higher than the probability of winning \$0 or \$1. The expected value is then the average winnings that you could expect, hence the name "expected value".

Going back to our physical model of a discrete probability distribution as a set of masses located at single points on a number line, the expected value of the distribution is the center of mass, or "balance point" of the number line.

The next exercise should illuminate what the variance is about.

2. Consider the following two different probability distributions g and h for the random variable X that takes the values $x = 0,\ 1,\ 2,\ 3$.

x:	0	1	2	3
$g(x)$:	$\frac{1}{16}$	$\frac{7}{16}$	$\frac{7}{16}$	$\frac{1}{16}$

x:	0	1	2	3
$h(x)$:	$\frac{3}{16}$	$\frac{5}{16}$	$\frac{5}{16}$	$\frac{3}{16}$

   (a) Find the expected value for the first distribution. Does the result make sense? What do you think the expected value for the second distribution will be?

   (b) Find the variances of the two distributions. Which probability distribution has the larger variance? Can you see what it indicates about the distribution?

Hopefully you recognized that the variance indicates how "spread out" the probabilities are over the values of the random variable.

Let's get the definitions of the expected value and variance formalized a bit. Note that we also call the expected value the **mean**.

Let X be a discrete random variable with probability distribution f. Then the **mean** or **expected value** of X is

$$\mu = E(X) = \sum_x x\, f(x)$$

and the **variance** of X is

$$\sigma^2 = E[(X-\mu)^2] = \sum_x (x-\mu)^2\, f(x)\,.$$

The positive square root of σ^2 is called the **standard deviation**, denoted by σ.

In situations where it might not be clear what the random variable is, we sometimes use μ_X and σ_X^2 instead of μ and σ^2.

◇ **Example 2.5(a):** Suppose that for some course an instructor assigns the following *weights* to the various components of students' grades: assignments, 15%; quizzes, 15%; exams, 40%; final exam, 30%. If an individual student has grades in each area of 93%, 81%, 84% and 78%, respectively, what is their overall course grade?

Their grade is $93(0.15) + 81(0.15) + 84(0.40) + 78(0.30) = 83.1\%$.

What has been done here is the students grade has been computed as what we call a **weighted average**. Rather than just average the percentage grades in each area "straight out", we average them in a way that makes some "count" more than others. Note that we could compute an "ordinary average" by replacing the **weights** 0.15, 0.15, 0.40, 0.30 each with 0.25. Notice that either way, *the weights are all positive and their sum is one!* this means that the weights are a discrete probability distribution.

This is exactly what we are doing when we compute an expected value; we are averaging the values of a random variable, weighting each according to its probability. To understand what we are doing when we compute the variance, let's introduce some terminology. We will call the quantity $x - \mu$ the **deviation** of the random variable value x (from the mean). When we compute the variance we are also computing a weighted average, but we are averaging the squares of the deviations.

3. The point of the variance is to give some idea of how the values of a random variable vary. It would make more sense to average the deviations of the values of the random variable rather than their squares. That is, it would seem that we should compute

$$\sum_x (x - u)\, f(x)\,.$$

Do this for the random variable from Exercise 1.

This exercise illustrates why averaging the deviations is not a good idea. One solution to the problem would be to average the absolute values of the deviations. It turns out, however, that averaging the squares of the deviations is more mathematically advantageous. We will not go into why that is here, just take my word for it!

4. Compute the mean and the variance for the random variable with probability distribution given by

x	0	1	2	3
$f(x)$	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$

5. Find the mean and variance for the experiment consisting of flipping a coin four times in a row, with the random variable X that assigns to each outcome the number of heads obtained. How many heads should one expect to get on average when performing this experiment repeatedly?

We now make a very convenient observation:

$$\begin{aligned}
\sum_x (x-\mu)^2 f(x) &= \sum_x (x^2 - 2x\mu + \mu^2)\cdot f(x) \\
&= \sum_x x^2 f(x) - 2\mu \sum_x x f(x) + \mu^2 \sum_x f(x) \\
&= E(X^2) - 2\mu E(X) + \mu^2 \\
&= E(X^2) - 2\mu^2 + \mu^2 \\
&= E(X^2) - \mu^2
\end{aligned}$$

Theorem 2.8: If X is a discrete random variable, then the variance of X can be computed by

$$\sigma^2 = E(X^2) - \mu^2 = E(X^2) - [E(X)]^2\,.$$

6. Use this formula to find the variance for the probability distribution from Exercise 4.

7. We will define a **Bernoulli random variable** to be a random variable X with range 0, 1, with $P(X = 1) = p$.

   (a) What is $P(X = 0)$?

   (b) Find the mean and variance of this random variable.

For continuous random variables we define the expected value and variance in essentially the same way as for discrete random variables, replacing summation with integration.

> Let X be a continuous random variable with probability distribution f. Then the **expected value** of X is
>
> $$\mu = E(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx$$
>
> and the **variance** of X is
>
> $$\sigma^2 = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2\, f(x)\, dx\,.$$

The argument preceding Theorem 2.8 can be repeated for continuous probability functions, replacing summations with integrals.

> **Theorem 2.9:** Let X be a continuous random variable. Then the variance of X can be computed by
>
> $$\sigma^2 = E(X^2) - \mu^2 = E(X^2) - [E(X)]^2\,,$$
>
> where $E(X^2) = \int_{-\infty}^{\infty} x^2\, f(x)\, dx\,.$

8. Find the mean and variance for the probability density function $f(x) = \begin{cases} 0 & \text{for } x < 0 \\ \frac{1}{2}x & \text{for } 0 \le x \le 2 \\ 0 & \text{for } x > 2 \end{cases}$

9. Find the mean and variance for the probability density function $f(x) = \begin{cases} \frac{1}{5} & \text{for } 0 \le x \le 5 \\ 0 & \text{otherwise} \end{cases}$

10. Find the mean and variance for the probability density function $f(x) = \begin{cases} \frac{2}{15}x & \text{for } 1 \le x \le 4 \\ 0 & \text{otherwise} \end{cases}$

11. Try finding the mean for the probability density function $f(x) = \begin{cases} 0 & \text{for } x < 1 \\ \dfrac{1}{x^2} & \text{for } x \ge 1 \end{cases}$

    What goes wrong?

> **Theorem 2.10:** If X is a discrete random variable with expected value $E(X)$ and variance σ, then
>
> $$E(aX+b) = aE(X) + b \qquad \text{and} \qquad \text{var}(aX+b) = a^2\sigma^2$$

2.6 Chapter 2 Exercises

Give all answers in exact form - NO DECIMALS!

1. An experiment consists of first flipping a coin and then, if the coin comes up heads, flipping the coin again. If the coin comes up tails on the first flip a four-sided die is rolled. A random variable X is assigned to the outcomes as follows: A sum is obtained for each outcome, with tails counting as zero and heads as one, and with the numbers on the die counting their value. So, for example, tails followed by a three is assigned $0+3=3$ and heads followed by heads is assigned $1+1=2$. (Of course there are multiple ways to obtain some range values of the random variable.)

   (a) Abbreviating the outcomes as HH, HT, T3, etc., give the value of X for each outcome *using function notation.*

   (b) Give $\text{Ran}(X)$.

   (c) Define $f : \text{Ran}(X) \to \mathbb{R}$ and sketch its graph.

   (d) Define $F : \mathbb{R} \to \mathbb{R}$ and sketch its graph.

2. *If possible*, find values of c so that each of the following is a probability density function for a *continuous* random variable X.

   (a) $f(x) = \begin{cases} 0 & \text{for } x < 1 \\ \dfrac{c}{\sqrt{x}} & \text{for } x \geq 1 \end{cases}$

   (b) $f(x) = \begin{cases} 0 & \text{for } x < 1 \\ \dfrac{c}{x^3} & \text{for } x \geq 1 \end{cases}$

3. Find the mean for the probability function from part (b) of the previous exercise.

4. Find the cumulative probability function F for the function f from Exercise 2(b).

5. A *discrete* random variable X has values $x = 1, 2, 3, 4, 5$. It is known that $F(3) = \frac{8}{17}$ and $F(4) = \frac{14}{17}$. Find any of the following that you can; *a number of them cannot be determined from the information given.*

   (a) $P(X < 3)$ (b) $P(X \leq 3)$ (c) $P(X = 3)$

   (d) $P(X \geq 4)$ (e) $P(3 < X \leq 4)$ (f) $P(X = 4)$

   (g) $P(X > 3)$ (h) $P(X > 4)$ (i) $P(3 \leq X \leq 4)$

6. For a *continuous* random variable X, $F(3) = \frac{8}{17}$ and $F(4) = \frac{14}{17}$. Find any of the values from Exercise 3 that you can.

7. A *continuous* random variable X has the cumulative distribution function

$$F(x) = \begin{cases} 0 & \text{for } x < 2 \\ 1 - \dfrac{2}{x} & \text{for } x \geq 2 \end{cases}$$

   Find any of the values from Exercise 4 that you can for this cumulative distribution function.

8. Find the probability density function f for the cumulative probability function from the previous exercise. *Be sure to define it for all values of x.*

9. A continuous random variable X has the cumulative density function

$$F(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 - e^{-x} & \text{for } x \geq 0 \end{cases}$$

Find each of the following. Again, your answers should be in exact form, so they will contain terms like e^{-3}. *Simplify when possible.*

(a) $P(X \leq -2) =$ ______ (b) $P(X \geq -2) =$ ______

(c) $P(1 < X \leq 5) =$ ______ (d) $P(X < 7) =$ ______

(e) $P(X = 7) =$ ______ (f) $P(X \geq 7) =$ ______

10. Consider the probability density function f for the *continuous* random variable X given by

$$f(x) = \begin{cases} \frac{1}{3} & \text{for } 0 \leq x < 2 \\ \frac{1}{9} & \text{for } 2 \leq x \leq 5 \\ 0 & \text{elsewhere} \end{cases}$$

(a) Find $P(X < 1)$. (b) Find $P(1 \leq X \leq 4)$.

(c) Fill in the blanks:

$$F(x) = \begin{cases} \_\_\_\_\_\_ & \text{for } \_\_\_\_\_\_\_\_\_\_\_\_ \\ \_\_\_\_\_\_ & \text{for } \_\_\_\_\_\_\_\_\_\_\_\_ \\ \frac{1}{9}x + \frac{4}{9} & \text{for } \_\_\_\_\_\_\_\_\_\_\_\_ \\ \_\_\_\_\_\_ & \text{for } \_\_\_\_\_\_\_\_\_\_\_\_ \end{cases}$$

11. Find the mean and variance for the distribution from the previous exercise.

12. A probability distribution is given by

$$F(x) = \begin{cases} 0 & \text{for } x < 0 \\ \frac{1}{5}x & \text{for } 0 < x \leq 5 \\ 1 & \text{for } x > 5 \end{cases}$$

(a) Is this distribution for a discrete random variable, or for a continuous random variable? Is the distribution cumulative?

(b) Give the probability function f.

13. Consider a random variable X.

(a) Under what conditions is $P(a \leq X \leq b) = F(b) - F(a)$?

(b) Under what conditions is $P(a \leq X \leq b) = F(b) - F(a-1)$?

(c) In the context of parts (a) and (b), discuss $P(a \leq X < b)$ and $P(a < X \leq b)$.

14. Find the expected value and variance for each of the following.

(a) The discrete probability distribution function

x:	0	1	2
$f(x)$:	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{6}$

(b) The continuous probability density function $f(x) = \begin{cases} -\frac{1}{2}x + 1 & \text{for } 0 \le x \le 2 \\ 0 & \text{elsewhere} \end{cases}$

15. Consider the continuous probability density function $f(x) = \begin{cases} \frac{1}{15} & \text{for } 0 \le x < 3 \\ \frac{1}{9} & \text{for } 3 \le x \le 7 \\ 0 & \text{elsewhere} \end{cases}$

    (a) Find $P(2 \le X \le 6)$. *You should not need to integrate, since the function is constant on intervals.*

    (b) Sketch the graph of $F(x)$.

16. Which of the graphs below represents

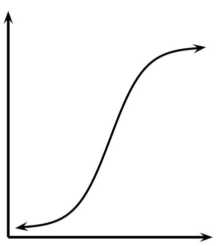
    (a) f for a discrete random variable?

    (b) F for a discrete random variable?

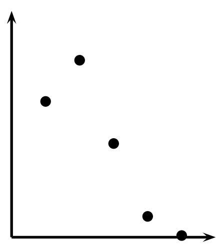
    (c) f for a continuous random variable?

    (d) F for a continuous random variable?

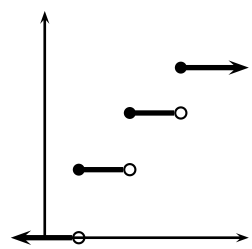
(i)



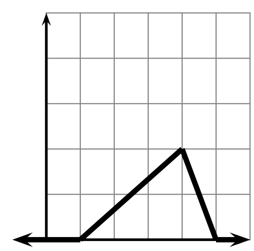
(ii)



(iii)



(iv)



17. Consider the continuous probability distribution function

$$f(x) = \begin{cases} -\frac{1}{2}x + 1 & \text{for } 0 \le x \le 2 \\ 0 & \text{elsewhere} \end{cases}$$

    (a) Why can't $\mu = 2\frac{1}{2}$?

    (b) Is μ less than 1, greater than 1, or is it 1?

18. (a) Is it possible for a distribution to have an expected value of zero? If so, make such a distribution.

    (b) is it possible for a distribution to have a variance of zero? If so, make one.

3 The Binomial and Normal Distributions

Performance Criteria:

3. (a) Determine whether an experiment is a Bernoulli process. If it is not, tell why not. If it is, give the probability of a success, when possible.
   (b) Correctly denote the binomial distribution function and cumulative binomial distribution function associated with a Bernoulli process. Give, and distinguish between, the parameters and the variable for the distribution.
   (c) Compute probabilities using both the binomial probability distribution and cumulative binomial probability distribution.
   (d) Compute the expected value and variance for a binomial distribution.
   (e) Compute probabilities from the standard normal distribution.
   (f) Determine a range of values of the random variable Z of the standard normal distribution having a given probability.
   (g) Compute probabilities from a normal distribution.
   (h) Determine a range of values of a normal random variable having a given probability.
   (i) Use the normal distribution to approximate the binomial probabilities when appropriate.

In the last chapter you studied some simple examples of discrete and continuous probability distributions. Most of those distributions I made up solely for the purposes of learning the concepts of random variables, probability functions, and the expected value and variance of a random variable.

Discrete distributions arise in practice whenever we are counting things, like a the number of heads when flipping a coin a certain number of times, or the number of hits on a web site during a given period of time. Continuous distributions are used when we are dealing with measurements, usually of time or dimensions in one, two or three dimensions.

In this chapter we will look at what are perhaps the most representative examples of discrete and continuous distributions, the binomial distribution and the normal distribution. In reality, each of these is not a single distribution, but a family of distributions that are distinguished from each other by certain **parameters**.

It is important to distinguish between parameters and variables. In some sense, parameters are variables that vary from situation to situation, but once the specific situation has been determined they become constants. We will say that variables are things that can still vary once a specific situation has been determined. For example, suppose we are flipping a coin a fixed number of times, and we want to discuss the probabilities of getting different numbers of heads. The number of times we will flip the coin is a parameter that will be determined ahead of time. Once the number of flips has been determined, the number of heads can still vary, so it is a variable.

3.1 The Binomial Distribution

Performance Criteria:

3. (a) Determine whether an experiment is a Bernoulli process. If it is not, tell why not. If it is, give the probability of a success, when possible.
   (b) Correctly denote the binomial distribution function and cumulative binomial distribution function associated with a Bernoulli process. Give, and distinguish between, the parameters and the variable for the distribution.
   (c) Compute probabilities using both the binomial probability distribution and cumulative binomial probability distribution.
   (d) Compute the expected value and variance for a binomial distribution.

A **Bernoulli process** is an experiment satisfying the following conditions:

- The experiment consists of a fixed number n repeated trials, called **Bernoulli trials**.
- For each trial there are two outcomes, generally referred to as "success" or "failure".
- The probability of success is the same for each trial; we will denote it by p.
- The trials are independent.

1. Determine which of the following experiments are Bernoulli processes; for those that are not, tell why. For those that are, describe what a "success" is, and give the probability of a success if you can.

   (a) An urn contains seven blue marbles and ten red marbles. Three marbles are drawn, *without replacement*. The number of blue marbles drawn is noted.

   (b) A die is rolled 5 times, and each time the number on the die is noted.

   (c) Three marbles are drawn, *with replacement*, from the same urn as used for (a). The number of blue marbles drawn is noted.

   (d) Every day on your way to school you pass though a particular stoplight. Assume that the light is not set to recognize the approach of vehicles; it simply remains green for a period of 60 seconds, yellow for 5 seconds, then red for 75 seconds, over and over again. (This is in the direction you pass through it.)Assume also that the time at which you approach the light each day, relative to its cycle, is random. For one week (assuming also that you go to school all five of those days) the number of times you encounter a green light there is recorded.

   (e) Consider the situation described in (d), but suppose that the light senses the approach of vehicles and changes in some way related to those approaches.

(f) The Acme Company makes Widgets using parts obtained from Fly-By-Night Machining (FBNM). Out of each batch of 1000 parts Acme receives from Fly-By-Night, they test 10 parts (with replacement) and note how many of them are defective.

(g) A coin is flipped repeatedly until heads is obtained. The number of flips it takes to get a head is recorded.

For any Bernoulli experiment there is a random variable X, the number of successes. This is called a **binomial random variable**, and is perhaps the most common discrete random variable for applications. The probability distribution for this random variable is called the **binomial distribution**. It is actually a family of distributions; there is one such distribution for every value 1, 2, 3, ... of n and for each real number p in the interval $[0,1]$. The numbers n and p are the parameters of the distribution. *Note again that for a particular Bernoulli process, n and p are FIXED quantities. x is the only variable.*

Instead of writing $f(x)$ for the binomial distribution, we write it as $b(x;n,p)$. The b is the name of the function, b for binomial. x is the variable, and the two symbols after the semi-colon are the parameters for the particular binomial distribution being used.

◇ **Example 3.1(a):** Consider the experiment consisting of flipping a coin five times in a row, with the random variable X assigning to each outcome of the number of heads obtained. This experiment is a Bernoulli process -

There are $n = 5$ trials, and the probability of success on any trial is $p = \frac{1}{2}$. We would denote the probability distribution by $b(x;5,0.5)$ or $b(x;5,\frac{1}{2})$.

---

2. Write the notation for the probability distribution associated with any of the experiments from Exercise 1 that *ARE* Bernoulli processes. If either n or p cannot be determined from the information given, just use those letters for that parameter.

So now it boils down to this: What are the values of the function $b(x;n,p)$? Those values could be determined on an experiment-by-experiment basis. But that is not necessary. Earlier we actually derived the following.

Binomial Distribution

Suppose that a Bernoulli process has n trials and the probability of success is p. The probability density function for the associated Bernoulli random variable is the **binomial distribution**

$$b\,(x;n,p) = \binom{n}{x} p^x\, q^{n-x}, \qquad x = 0,\ 1,\ 2, ...,\ n\,,$$

where $q = 1-p$.

The right way to think of the binomial distribution is not that it is one distribution, but rather a family of distributions, each with a different combination of the two parameters n and p. Sometimes we will refer to this as a two-parameter family of distributions. Note that p is the probability of a success, and its exponent x in the probability density function is the number of successes. Similarly, q is the number of failures and its exponent $n-x$ is the number of failures.

◇ **Example 3.1(b):** Give the probability distribution for the experiment of flipping a coin five times in a row (see Example 3.1(b)), as both an algebraic expression and as a table of values.

Again, $n = 5$ and $p = \frac{1}{2}$, so the probability distribution is

$$b\left(x; 5, \frac{1}{2}\right) = \binom{5}{x}\left(\frac{1}{2}\right)^x\left(\frac{1}{2}\right)^{5-x} = \frac{5!}{x!(5-x)!}\left(\frac{1}{2}\right)^5, \qquad x = 0, 1, 2, ..., 5\,.$$

Letting x take the integer values from zero to five gives us

x	0	1	2	3	4	5
$b(x; 5, \frac{1}{2})$	$\frac{1}{32}$	$\frac{5}{32}$	$\frac{10}{32}$	$\frac{10}{32}$	$\frac{5}{32}$	$\frac{1}{32}$

3. (a) Three marbles are drawn, *with replacement*, from an urn containing seven blue and ten red marbles. The number of blue marbles drawn is noted. (This is the experiment from Exercise 1(c).) Give the probability distribution for the experiment, *giving all probabilities in fractional form.*

   (b) Give the probability distribution for the experiment from Exercise 1(d), *again giving all probabilities in fractional form.*

4. Refer to Exercise 1(f). If Fly-By-Night Machining claims that 985 out of every 1000 parts that they manufacture are not defective, what is the probability that in a sample of 10 such parts, Acme will find

   (a) exactly one defective part?
   (b) exactly four defective parts?
   (c) at most 1 defective part?
   (d) one, two, three or four defective parts?

5. (a) Write out a table of values for $b(x; 3, \frac{1}{5})$, giving all probabilities in exact form.

   (b) The cumulative binomial distribution with parameters n and p is denoted by $B(x; n, p)$. Give the piecewise definition of $B(x; 3, \frac{1}{5})$.

   (c) Suppose that this experiment had five trials instead of three. How many successes would you then expect? How many successes would you expect if there were 15 trials? How many would you expect if there were three trials?

   (d) Find μ and σ^2 for this probability distribution. Does the value of μ you obtain this way agree with your last answer to (c)?

Theorem 3.1: The mean and variance for the binomial distribution $b(x; n, p)$ are

$$\mu = np\,, \qquad \sigma^2 = npq\,.$$

6. Verify that the equation given above for σ^2 gives the correct value for the distribution from Exercise 5(d).

7. Consider a Bernoulli process with 3 trials.

(a) How many successes are possible; that is, what values does the random variable X have?

(b) Write a simplified expression for $b(x; n, p)$, in terms of p and q, for each possible value of X.

(c) What is the sum of the expressions you found in (b), and why?

(d) Expand the expression $(p+q)^3$. What do you notice?

(e) What is the value of $p+q$, assuming these are the p and q used in the binomial distribution? What must the value of $(p+q)^3$ then be?

The figure shown below is **Pascal's triangle**, which you are probably familiar with. Note how it relates to the above exercise.

$$\begin{array}{ccccccccc} & & & & 1 & & & & \\ & & & 1 & & 1 & & & \\ & & 1 & & 2 & & 1 & & \\ & 1 & & 3 & & 3 & & 1 & \\ 1 & & 4 & & 6 & & 4 & & 1 \end{array}$$

3.2 The Standard Normal Distribution

Performance Criteria:

3. (e) Compute probabilities from the standard normal distribution.
   (f) Determine a range of values of the random variable Z of the standard normal distribution having a given probability.

A very important class of functions in mathematics are those of the form $f(x) = Ce^{-kx^2}$. When $C = \dfrac{1}{\sqrt{2\pi}}$ and $k = \frac{1}{2}$ this function is called the **standard normal distribution**. Traditionally the letter Z has been used for the continuous random variable for this distribution.

Standard Normal Distribution

A random variable Z has the **standard normal distribution** if it has the probability density function

$$f(z) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{z^2}{2}}$$

This distribution doesn't really describe any phenomena that are seen in the "real world", but it is closely related to a family of distributions called normal distributions, which we will look at in the next section.

You will show later that the expected value of the standard normal distribution is zero, and its variance (and standard deviation) is one. The other normal distributions have other means and variances. A normal distribution with mean μ and variance σ^2 will be denoted $n(x\,;\mu,\sigma)$; like the binomial distribution, the normal distributions are a two-parameter family of distributions. So the standard normal distribution is $n(z\,;0,1)$.

By the definition of a continuous probability distribution,

$$P(a \le Z \le b) = \int_a^b \frac{1}{\sqrt{2\pi}}\, e^{-\frac{z^2}{2}}\, dz\,.$$

The problem here is that the function $\dfrac{1}{\sqrt{2\pi}}\, e^{-\frac{z^2}{2}}$ can't be integrated analytically, since it has no anti-derivative. Therefore it must be done numerically. This can be done by many calculators and a variety of computer software. Prior to the advent of this technology, people used (or still use!) a table of values of the cumulative distribution $N(z\,;0,1)$, which were of course computed numerically as well. I have provided you with such a table in Appendix B; it gives values of $N(z\,;0,1)$ in increments of 0.01 for the variable z. By Theorem 3.5 we then have

$$P(a \le Z \le b) = \int_a^b \frac{1}{\sqrt{2\pi}}\, e^{-\frac{z^2}{2}}\, dz = N(b\,;0,1) - N(a\,;0,1)\,.$$

◇ **Example 3.2(a):** Use the standard normal distribution table to find $N(1.32\,;0,1)$.

First we note that $1.32 = 1.3 + 0.02$. Now look at the table that has positive numbers down the left hand side of the table and find the row starting with 1.3. Go across the top of the table to the column that has .02 at the top and find the number in that column that is also in the row that you located with the value 1.3. Try this; you should get $N(1.32\,;0,1) = 0.9066$.

---

1. Find each probability.

   (a) $P(Z < -1.86)$ (b) $P(Z \geq 2.26)$

   (c) $P(-1.54 \leq Z \leq -0.13)$ (d) $P(-2.09 \leq Z \leq 1.27)$

Sometimes we will know a probability and wish to find a value of the random variable Z from the given probability. In a sense we are reversing the process from the previous example .

◇ **Example 3.2(b):** Find a value z such that $P(Z \leq z) = 0.27$.

Recalling that $P(Z \leq z) = N(z\,;0,1)$, we look in the body of the standard normal distribution table for the value that is nearest to 0.27. That number lies between 0.2709 and 0.2676, with 0.2709 being the closer number. We then go to the left end of the row and the top of the column to find z, finding that $z = -0.61$.

---

2. For each of the following, find the value of z such that

   (a) $P(Z \leq z) = 0.58$ (b) $P(Z \geq z) = 0.71$

   (c) $P(-z \leq Z \leq z) = 0.37$ (d) $P(0 \leq Z \leq z) = 0.42$

   Just use the probability in the standard normal distribution table that is closest to the given probability, as in the example above.

3.3 Normal Distributions

Performance Criteria:

3. (g) Compute probabilities from a normal distribution.
   (h) Determine a range of values of a normal random variable having a given probability.

Replacing z with $\dfrac{x-\mu}{\sigma}$ in the standard normal distribution gives us the family of normal distributions, each of which can be shown to have expected value μ and variance σ.

Normal Distribution

A random variable X has the **normal distribution** if it has the probability density function

$$n(x\,;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \tag{1}$$

Here μ and σ are parameters, and the distributions in this two-parameter family model probabilities for many applied situations. Sizes and weights of objects are often normally distributed.

When computing probabilities for these distributions we run into the same problem as we did in the previous section, for the standard normal distribution. Howeve r, if we have tables or software that allows us to evaluate integrals for the standard normal distribution, we can use those tables for other normal distributions as well. This is because if we make the substitution $z = \dfrac{x-\mu}{\sigma}$ we get

$$P(a \le X \le b) = P\left(\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma}\right).$$

◇ **Example 3.3(a):** Suppose that you know that actual resistances of 10 ohm resistors from a particular manufacturer are normally distributed, with mean 9.97 ohms and standard deviation 0.03 ohms. Find the probability that a randomly selected resistor has resistance between 9.95 and 10.05 ohms.

Letting X represent the random variable that assigns to a randomly selected resistor its resistance, its probability distribution is then $n(x\,;9.97, 0.03)$. To find the desired probability, we first compute the values of Z that correspond to $x = 9.95$ and $x = 10.05$:

$$z = \frac{9.95-9.97}{0.03} = -0.67 \qquad \text{and} \qquad z = \frac{10.05-9.97}{0.03} = 2.67\,.$$

(Here we have rounded the Z values to the nearest hundredth because that is the precision of the standard normal distribution tables.) We now have

$$\begin{aligned} P(9.95 \le X \le 10.05) &= P(-0.67 \le Z \le 2.67) \\ &= N(2.67\,;0,1) - N(-0.67\,;0,1) \\ &= 0.9962 - 0.2514\,. \end{aligned}$$

1. Replacement times for CD players are normally distributed, with a mean of 5.3 years and a standard deviation of 1.7 years.

   (a) If you just bought a CD player, what is the probability that you will have to replace it within 6 years?

   (b) Find the probability that a CD player will need replacement sometime between 3 and 7 years after it is purchased.

   (c) Find the probability that a CD player will last at least 8 years before it needs replacement.

2. The lengths of pregnancies are normally distributed, with a mean of 268 days and a standard deviation of 15 days. A baby is considered to be premature if it is born at least three weeks early. What is the probability that a randomly selected baby will have been premature?

3. Suppose that we want to know how long the most reliable 20% of all CD players can be expected to last. In other words, find the length of time such that there is a probability of 0.20 that a CD player will last at least that long. Think about trying to solve this without the following hints, then use them if needed.

   (a) Sketch a normal distribution and shade an area at the right end that appears to have an area of about 0.20.

   (b) Use the Standard Normal Distribution Table to find the z-score that corresponds to that area.

   (c) Solve the equation $z = \dfrac{x-\mu}{\sigma}$ to find the x-value that corresponds to the z that you just found. That is the answer!

4. The most *unreliable* 10% of CD players will need replacement within at most how many years?

5. The middle 60% of all CD players will last how many years?

6. Show that the distribution (1) has mean μ. You will need to use the fact that the standard normal distribution has mean zero.

3.4 Normal Approximation to the Binomial Distribution

Performance Criteria:

3. (i) Use the normal distribution to approximate the binomial probabilities when appropriate.

The normal distribution is the "classic" continuous distribution, and the binomial distribution is the classic discrete distribution. In this section you will use Excel to explore a relationship between the two distributions. Of course we can't simply compare values of the binomial distribution with values of the normal distribution, since the normal distribution is a *density*, and its values do not actually represent probabilities.

1. (a) Consider a Bernoulli process with 10 trials. What are the possible numbers of successes? Put these values in the A column of an Excel spreadsheet, *beginning in cell A2*. Remember that these are the values x of a random variable X.
   (b) Suppose that the probability of a success is 0.1. In column B, next to each number of successes in column A, put the probability of that number of successes. It is suggested that you use the built-in binomial distribution function that Excel has.
   (c) Label the two columns at their tops, in the empty cells there.
   (d) Repeat part (b) for $p = 0.3$ and $p = 0.5$, putting the results in columns D and F respectively. (Yes, this leaves some empty columns.) Label them appropriately, maybe changing your labeling of column B if necessary.

2. The mean and standard deviation for the binomial distribution are

$$\mu = np\,, \qquad\qquad \sigma = \sqrt{npq}.$$

   Find the mean and standard deviation for each of the distributions from Exercise 1.

3. We will now consider a normal distribution whose mean and standard deviation are those you computed for the binomial distribution with $p = 0.1$. We want to compare this distribution with the corresponding binomial distribution, but we have to consider the following: Since the normal distribution is continuous, the probability at any point is zero, so the probabilities of the values of X that you listed in column A of your spreadsheet will all be zero! Instead, in column C we want $P(x - 0.5 < X < x + 0.5)$ for the value of x in the corresponding cell of column A. Do this; *you should not necessarily expect the values in column C to correspond with those in column B.*

4. Repeat Exercise 3 in columns E and G of your spreadsheet, but for the probabilities $p = 0.3$ and $p = 0.5$, respectively.

5. Find $B(4.5; 10, 0.3) - B(3.5; 10, 0.3)$ and $N(4.5; 3, \sqrt{2.1}) - N(3.5; 3, \sqrt{2.1})$.

6. Print your spreadsheet. Look at how the binomial and normal distributions compare for all three probability values. Think about it a little. For which of the three binomial distributions does the normal distribution most closely approximate the binomial distribution? Turn in your spreadsheet (with your name on it).

7. Go to Sheet 2 of your worksheet and find the binomial probabilities for a Bernoulli process with $n = 50$ and $p = 0.1$. Then find the corresponding normal distribution probabilities. How do the values compare in this case?

8. Summarize what you have seen. For what value or values of p does the normal distribution best approximate the binomial distribution? For what value or values of n does the normal distribution best approximate the binomial distribution?

What you should have realized is that the normal distribution is a better approximation of the binomial distribution when p is close to 0.5 and when n is large. A general rule of thumb that some people use is that the normal distribution is a good approximation of the binomial distribution when both

$$np \geq 10 \qquad \text{and} \qquad n(1-p) \geq 10\,.$$

In the past it has been useful to use the normal distribution as an approximation of the binomial distribution when computing probabilities for Bernoulli processes with large numbers of trials. With better computers and software, that is no longer necessary for most applications. There are some applications, however, where approximating a discrete distribution with a continuous distribution is useful.

3.5 Chapter 3 Exercises

1. A magazine article states that

   > Research from the National Highway Traffic Safety Administration shows that up to 80% of crashes can be attributed to driver inattentiveness.

   Assume that whether or not a crash is caused by inattentiveness is a Bernoulli process. With probability as given above. For each of the probabilities asked for, give

   - an expression involving the pdf b that gives the desired probability
   - an expression involving the cdf B that gives the desired probability
   - the probability, as a decimal rounded correctly to four places past the decimal

   You can/should use your calculator, Excel or some other assistance to find the last of these. **For the first two, use the notation given in the book.**

   (a) The probability that 7 of ten recent crashes were due to inattentiveness.
   (b) The probability that 3 or 4 of five crashes are due to inattentiveness.
   (c) The probability that 10 or fewer of 15 crashes are due to inattentiveness.
   (d) the probability that 5 or more of eight crashes are due to inattentiveness.

2. In one year there are 427 crashes in a small town. How many of those would we expect to be due to inattentiveness? What concept that we've studied does this illustrate?

3. Suppose that 20 marbles are to be drawn, with replacement, from a top hat (think Cat in the Hat) containing three yellow marbles and two blue marbles. Create a table in Excel with three columns, one for the number x of yellow marbles drawn, one for the appropriate pdf b values and one for the appropriate cdf B values. Label the top of each column with what it is and format the cells with numbers in them so that each probability is given to five places past the decimal.

4. Suppose that 100 Ω (ohm) resistors from a certain manufacturer actually have a mean of 99.92 Ω and standard deviation of 0.17 Ω. For each question below following, give each of the following, **connected by equal signs:**

   - probability statement of the form $P(\textit{something about } X)$, followed by
   - the expression involving the cumulative distribution N that gives the desired probability, followed by
   - the desired probability, to four places past the decimal.

   Find the probability that a randomly selected resistor has resistance

   (a) over 100 Ω (b) less than 99.7 Ω (c) between 99.8 and 100.2 Ω

5. Give a probability statement of the form $P(\textit{something about } Z)$ that is equivalent to your probability statement from part (c) of the previous exercise, followed by the expression involving the cumulative distribution N that gives the desired probability, followed by the corresponding value(s) obtained from the Standard Normal Distribution Table, followed by the final answer.

6. Suppose that five resistors are to be selected from a batch of fifty. Find the probability that two of the five have resistances over 100 Ω. **Indicate clearly, using appropriate notation, how you obtain your answer.**

4 More Distributions

Performance Criteria:

4. (a) Apply the hypergeometric distribution to solve applied problem.
   (b) Apply the negative binomial distribution to solve applied problem.
   (c) Apply the Poisson distribution to solve applied problem.
   (d) Approximate a binomial probability with the Poisson distribution, when appropriate.
   (e) Apply the exponential distribution to solve applied problem.
   (f) Apply the gamma distribution to solve applied problem.
   (g) Choose and apply the appropriate distribution(s) to solve an applied problem.

In the last chapter we worked with the binomial and normal distributions, the standard examples of discrete (the binomial) and continuous (the normal) distributions. In this chapter we will introduce and work with several other commonly encountered distributions.

4.1 Hypergeometric Distribution

Performance Criteria:

4. (a) Apply the hypergeometric distribution to solve applied problem.

1. An urn contains 40 marbles, exactly 15 of which are blue. Three marbles are drawn at random, *without replacement*. Drawing a blue will be considered a success, and the number of successes in the three draws will be observed.

   (a) Why is this experiment not a Bernoulli process?

   (b) One could draw a tree for this experiment. Draw just those branches of the tree for which there are exactly two successes. Determine the probability of ending up at the end of each of those branches, indicating clearly on your tree the conditional probabilities used.

   (c) Each of your probabilities from (b) should be the same. Write that probability in terms of "partial factorials" (things like $7 \cdot 6 \cdot 5$).

   (d) The number of branches with two successes can be written as a combination. What combination is it?

   (e) Write the probability of two successes in terms of your answers to (b) and (d).

2. Suppose again the urn containing 40 marbles with 15 of them being blue, but now the experiment is to draw five marbles without replacement. Find the probability of drawing exactly three blue marbles, writing your answer in a form like that of part 1(e) above.

In general, suppose there are N objects from which n are to be drawn without replacement, and k of the N objects are considered successes when drawn. Let X be the random variable that assigns to each set of n objects drawn the number of successes. The probability distribution function for the random variable X is the

Hypergeometric Distribution

$$h(x\,; n, k, N) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}, \qquad x = 0,\ 1,\ 2, ...,\ \min\{n, k\}.$$

Here it is assumed that $n, k \leq N$, and the notation $\min\{n, k\}$ means the smaller of n and k. *Note that the family of hypergeometric distributions is a THREE parameter family of distributions.*

3. You keep a jar of change at home, and (unbeknownst to you) it contains 274 coins, 87 of which are quarters. You are looking for quarters, so you take 10 coins from the jar.

   (a) What is the probability that exactly four of those coins are quarters? *Give your answer by first giving the distribution with the values of the variable and the parameters, then give it in decimal form, rounded to the thousandths place.*

(b) What is the probability that none of the coins are quarters? *Give your answer in the same way as you did in part (a).*

4. Suppose that an urn contains 10 marbles, exactly 6 of which are blue. An experiment consists of drawing three marbles without replacement.

   (a) Writing $h(x) = h(x\,; 3, 6, 10)$, give the probability distribution h. *Give all probabilities in fraction form, reduced as far as possible while keeping the same denominator for all probabilities.* **NOTE:** Remember that this is a discrete distribution, so it only takes values at the values of x that are in the range of the random variable X that assigns to each outcome of the experiment the number of blue marbles drawn.

   (b) Give the cumulative distribution $H(x\,; 3, 6, 10)$.

5. For the experiment of the previous exercise, find the probability of selecting

   (a) exactly one blue marble.

   (b) fewer than three blue marbles.

   (c) zero blue marbles.

   (d) at least one blue marble.

6. Consider again the jar of change containing 274 coins, 87 of which are quarters, and the experiment of randomly selecting 10 coins from the jar. Use Excel or some other technology to determine the probability of selecting *at least* four quarters. Indicate how you obtained your answer, in terms of either the probability function h or the cumulative probability function H. *Give your answer in decimal form, rounded to the thousandths place.*

Since the experiments of drawing marbles from an urn with or without replacement are similar, we might expect there to be a relationship between the binomial distribution and the hypergeometric distribution. The next exercise illustrates this relationship.

7. (a) A certain kind of part is claimed to have a defective rate of 2%. Assuming this holds for every shipment, what is the probability that three out of 100 randomly selected (without replacement) parts from a shipment of 10,000 will be defective? *Give your answer in decimal form, rounded to six places past the decimal.*

   (b) Suppose that 100 parts from the shipment are selected at random, but *WITH* replacement. What is the probability that three out of the 100 are defective? *Give your answer in decimal form, rounded to six places past the decimal.*

   (c) Compare your answers to (a) and (b) and explain what you see.

It should be clear that what you saw in the last exercise is due to the fact that the number n of parts drawn is much smaller than the total number N of parts. (Sometimes we write this as $n << N$.) A general rule of thumb is as follows:

$$\text{If } N \geq 20n, \text{ then } h(x\,; n, k, N) \approx b(x\,; n, \tfrac{k}{N})\,.$$

The condition $N \geq 20n$ should not be taken to be firm; the approximation will be fairly good if, for example, $N = 18.3n$. What we *can* deduce is that for something like $N = 10n$ the approximation is probably not very good, and if $N = 50n$ it is probably quite good.

8. Consider the results of Exercise 7.

   (a) The error in using the binomial distribution to approximate the hypergeometric distribution is the absolute value of the difference between your answers to 7(a) and (b). Find the error, to six places past the decimal.

   (b) The percent error is the error divided by the correct value. (This of course gives the error in decimal form.) Find the percent error, to the nearest tenth of a percent.

4.2 Negative Binomial Distribution

Performance Criteria:

4. (b) Apply the negative binomial distribution to solve applied problem.

1. Suppose we have an urn with three red marbles and two yellow marbles, and consider the experiment of drawing marbles *WITH replacement* until two yellow marbles have been drawn.

   (a) Draw a tree diagram for this experiment, stopping whenever two yellows have been drawn or at the fourth draw, whichever comes first.

   (b) Compute the probability that the second yellow is drawn on the third draw.

2. Suppose now that for the same urn, you will draw until seven yellow marbles have been obtained. In this exercise you will determine the probability that the seventh yellow marble is drawn on the tenth draw. This really boils down to two things: (1) Finding the probability of any single outcome for which the seventh yellow is obtained on the tenth draw. (2) Determining how many ways the seventh yellow can be obtained on the tenth draw.

   (a) Determine the probability of obtaining any one outcome for which the seventh yellow is drawn on the tenth draw. Give your answer as the product of two numbers to a power.

   (b) Now to obtain the seventh yellow on the tenth draw, six yellows must have been drawn on the first nine draws. How many ways can this happen?

   (c) Of course the probability of getting the seventh yellow on the tenth draw is the product of your answers to (a) and (b). What is that probability?

3. Consider now a Bernoulli process with probability of success p, and let $q = 1 - p$, as before. Fix a number n of successes, and draw until the kth success. Let X be the random variable that assigns to each outcome the number of the draw on which the nth success occurs. Give an expression for $P(X = x)$.

Negative Binomial Distribution

Consider conducting a Bernoulli process, with the goal of obtaining k successes. Let the random variable X be the number of trials needed to obtain k successes. Then the probability distribution function is

$$b^*(x\,; k, p) = \binom{x-1}{k-1} p^k q^{x-k}, \qquad x = k,\ k+1,\ k+2, ...,$$

where $q = 1 - p$.

4. Mr. W figures that the probability that any e-mail received is "spam" is 0.13. It is assumed that whether or not an e-mail received is spam is independent of all previous e-mails received. What is the probability that Mr. W receives his fourth spam e-mail for the day as his 15th overall e-mail of the day? Give your answer to four places past the decimal.

4.3 Poisson Distribution

Performance Criteria:

4. (c) Apply the Poisson distribution to solve applied problem.
   (d) Approximate a binomial probability with the Poisson distribution, when appropriate.

A **Poisson process** is an experiment consisting of counting the number of successes in a given period of time or region of space. It must meet the following conditions:

- The number of successes in two disjoint time periods or regions are independent of each other.
- The probability of a single success in a very short time interval or region is proportional to the length of time or size (length, area, volume) of the region.
- The probability of more than one success in such a short time interval or small region is negligible.

Poisson Distribution

Consider a Poisson process with average number λ of successes per unit of time or space. Let the random variable X be the number of successes in a fixed period of time t. Then

$$P(X = x) = p\,(x; \lambda t) = \frac{e^{-\lambda t}(\lambda t)^x}{x!}, \qquad x = 0,\, 1,\, 2,\, 3, \ldots$$

1. Suppose that an average of 315 semi-trucks pass through a weigh station every day, and assume that their arrival at the weigh station is a Poisson process.

   (a) Determine the rate λ at which trucks arrive, using whatever time units you wish, *but giving units with your answer*. (You may wish to read the next part of this exercise before deciding what time units to use - *whatever you use, keep all decimal places.*) Note that since we are assuming this is a Poisson process, this rate is good 24 hours a day. (This may be a reasonable assumption, since many truckers travel at night to avoid traffic.)

   (b) Determine the probability that exactly two trucks will arrive in a five minute period. Give your answer to the hundredth's place.

2. Consider the situation from the previous exercise, and suppose that we wish to know the probability that more than two trucks will arrive in a five minute period.

   (a) Write an infinite sum, or an expression involving an infinite sum, whose value is the desired probability.

   (b) Write a *finite sum*, or an expression involving a finite sum, whose value is the desired probability.

(c) Rewrite your answer to (b) in terms of the *cumulative* probability function $P(x; \lambda t)$.

(d) Find the desired probability, to the ten-thousandth's place.

3. In the "real world" it is often the case that a person will not necessarily know whether (or how well) a situation satisfies the conditions necessary for it to be modeled by a particular distribution. In these cases one often gathers some data and finds its distribution, then sees how well it matches the results given by a known distribution. You will investigate this process in this exercise.

   A machine is designed to spread fertilizer evenly over an area. You spread some fertilizer over a 10 square foot area that is marked off in a grid of one inch by one inch squares. The number of fertilizer pellets in each square is counted, giving these results:

No. of pellets in a square:	0	1	2	3	4	5	6	$\cdots$
Number of squares:	731	494	169	42	4	0	0	$\cdots$

   The numbers of squares containing the different numbers of pellets will be referred to as the *actual counts*. What you will now do is see what values you would *expect* to get in the second row if the spread of fertilizer followed a Poisson distribution. Those vales are what we call the *expected counts*

   (a) Find the average number of pellets per square. (This is simply a weighted average of the numbers 0, 1, 2, 3, ..., with the weights being the number of squares having those numbers of pellets in them.) The value you obtain is λ for the Poisson distribution. *Round to the nearest ten-thousandth, and give units with your answer.*

   (b) Determine the probabilities of 0, 1, 2, ... particles per square inch from the Poisson distribution. Round these values to the nearest ten-thousandth as well, and record your results in a table with two columns.

   (c) Add a third column to your table, giving the number of squares (out of the total in the 10 square foot area) that should contain the given number of pellets if the spreading of fertilizer really follows a Poisson distribution.

   (d) Compare the expected values from your table with the actual results of the experiment. Do you think that the Poisson distribution models the spread of the fertilizer well?

If p is small, $b(x\,; n, p) \approx p(x\,; np)$.

NOTE: This is used to approximate binomial probabilities with the Poisson distribution, since it can easily be computed as an exponential function.

4. A certain kind of part is claimed to have a defective rate of 2%. Assuming this holds for every shipment, what is the probability that 3 out of 100 randomly selected (without replacement) parts from a shipment of 10,000 will be defective? *Round all answers to the parts of this exercise to the ten-thousandth's place.*

   (a) Compute the desired probability, using the appropriate distribution.

   (b) This is not a Bernoulli experiment. However, it "almost" is; compute an approximation of the desired probability using the binomial distribution.

   (c) Approximate the binomial probability from part (b) using the Poisson distribution.

4.4 The Exponential Distribution

Performance Criteria:

4. (e) Apply the exponential distribution to solve applied problem.

Consider a Poisson process with parameters λ and t. Here λ is a constant, and the time t is a parameter to be fixed. If X is the random variable that assigns to an outcome the number of successes in a time period of length t, then $P(X = x) = p(x\,;\lambda t)$, where p is the Poisson distribution.

Now $p(0\,;\lambda t)$ represents the probability of zero successes in a time period of length t. If we think of fixing the value of zero for x and letting t be the variable, we can look at $p(0\,;\lambda t)$ as representing the probability that it will take at least time t to obtain one success. If we then take the new random variable X to be the length of time to the first success, we have

$$P(X \geq t) = p(0\,;\lambda t) = \frac{e^{-\lambda t}(\lambda t)^0}{0!} = e^{-\lambda t}\,.$$

So we now have a Poisson process with rate parameter λ, and we are considering the experiment of waiting until the first success and recording the time to that success. We define the random variable X to be the function that assigns to each outcome the time x to that outcome. (Note that we are replacing t with x in the above.) By the above, the cumulative probability function for this random variable is

$$F(x) = P(X \leq x) = 1 - P(X \geq x) = 1 - e^{-\lambda x}\,.$$

By Theorem 3.7, the probability density function f for this distribution is given by $f(x) = F'(x) = \lambda e^{-\lambda x}$. It is customary to then let $\lambda = \frac{1}{\beta}$, giving us the following.

Exponential Distribution

The **exponential distribution** is the continuous probability density function f given below, along with it's cumulative distribution function:

$$f(x) = \begin{cases} \dfrac{1}{\beta}\,e^{-\frac{x}{\beta}} & \text{for } x > 0 \\ 0 & \text{for } x \leq 0 \end{cases} \qquad\qquad F(x) = \begin{cases} 1 - e^{-\frac{x}{\beta}} & \text{for } x > 0 \\ 0 & \text{for } x \leq 0 \end{cases}$$

This distribution is a good model for time between "successes" for a Poisson process. The parameter β is the average time between successes, and is the reciprocal of the parameter λ from the Poisson distribution, the average number of successes per unit time. Note that

$$\beta = \frac{1}{\lambda} \quad \Rightarrow \quad \lambda = \frac{1}{\beta}.$$

1. Suppose that customers arriving at a drive-up ATM machine during lunch hour is a Poisson process (not a bad assumption). The average time between customers during this time is 8 minutes.

(a) A customer has just arrived at the machine. What is the probability that the next customer will arrive in exactly five minutes? (Remember that this is a *continuous* distribution!)

(b) What is the probability that the next customer will arrive sometime between three and five minutes from now?

(c) What is the probability that the next customer will arrive in less than ten minutes?

(d) What is the probability that the next customer will arrive in more than ten minutes?

2. Consider the same scenario as Exercise 1.

(a) Suppose that you just arrived at the ATM machine during the lunch hour. What is the probability that 3 people will arrive in the next 15 minutes? *Note that the random variable has changed! You need to use a different distribution here.*

(b) You just arrived at the machine. What is the probability that no customers will arrive in the next ten minutes?

(c) Compare your answer to (b) with your answer to 1(d). Think about this!

3. The exponential distribution models things like times to failure of electronic components, where β is the average time to failure. Suppose that the average time to failure for a certain component is 3.7 years. What is the probability that a randomly selected component will fail in less than 2 years?

4. In this exercise you will find the mean and variance of the exponential distribution, using the facts (which are obtained using integration by parts) that

$$\int u\, e^u\, du = e^u(u-1) + C \qquad \text{and} \qquad \int u^2\, e^u\, du = e^u(u^2 - 2u + 2) + C\,.$$

(a) Make the substitution $u = -\dfrac{x}{\beta}$ to show that $E(X) = \beta \displaystyle\int_0^{-\infty} u\, e^u\, du$. Then evaluate this integral, using one of the above. (What is $e^{-\infty}(-\infty - 1)$, and why?)

(b) Find $E(X^2)$ in the same manner, using the same substitution.

(c) Find σ^2.

For the exponential distribution, $\mu = \beta$ and $\sigma^2 = \beta^2$.

5. The result of this exercise may surprise you a bit! Consider an exponential distribution with $\beta = 3$.

(a) What do you think $P(X \le \mu)$ is?

(b) Find $P(X \le \mu)$, as a decimal to the nearest hundredth. Does your answer agree with your conjecture from (a)?

(c) For a probability distribution, the value of x such that $P(X \le x) = \frac{1}{2}$ is called the **median**. Find the median of the exponential distribution with $\beta = 3$.

4.5 The Gamma Distribution

> **Performance Criteria:**
>
> 4. (f) Apply the gamma distribution to solve applied problem.

Consider again the distribution

$$f(x) = \begin{cases} 0 & \text{if} \quad x < 0\,, \\ k\,x^{\alpha-1}\,e^{-\frac{x}{\beta}} & \text{if} \quad x \geq 0 \end{cases} \tag{1}$$

for parameters $\alpha > 0$ and $\beta > 0$, with k some constant to be determined.

1. (a) Make the substitution $y = \dfrac{x}{\beta}$ to show that

$$\int_0^\infty k\,x^{\alpha-1}\,e^{-\frac{x}{\beta}}\,dx = k\,\beta^\alpha\,\Gamma(\alpha)\,.$$

   (b) Determine what the value of k must be in order for the above function to be a probability density function.

> **The Gamma Distribution**
>
> For $\alpha > 0$ and $\beta > 0$, the continuous probability distribution with probability distribution function
>
> $$f(x) = \begin{cases} 0 & \text{for} \quad x \leq 0 \\ \dfrac{1}{\beta^\alpha \Gamma(\alpha)}\,x^{\alpha-1}\,e^{-\frac{x}{\beta}} & \text{for} \quad x > 0 \end{cases}$$
>
> is called the **gamma distribution**.

Note that when $\alpha = 1$ the gamma distribution becomes the exponential distribution, which sometimes models time to failure of an electronic component, or time until arrival of a person or vehicle at some point. One interpretation of the gamma distribution is that it models time to failure of *several* electronic components, or time to arrival for multiple people/vehicles. In these cases, α represents the number of components/people/vehicles, and β is again the average time between failures/arrivals.

2. Consider the weigh station of Exercise 1, Section 4.3, where the average time between trucks passing through the weigh station was 0.0762 hour. The scales used to weigh the trucks have to be recalibrated every fifty trucks. What is the probability of having to recalibrate less than four hours after the previous recalibration? Use your calculator to evaluate the appropriate integral, and give your answers to the thousandth's place.

3. Now we return to our ATM machine, with an average time between customers of 8 minutes (during the lunch hour). During the lunch hour, what is the probability that

   (a) That the amount of time for three customers to arrive (after the customer preceding all of them) will be ten to fifteen minutes?

   (b) That the amount of time for five customers to arrive (again, after some preceding customer) will be twenty minutes or less?

4. Now consider electrical components that have an average time to failure of 3.7 years. What is the probability it will be five years or longer before two of them fail?

5. (a) Use the substitution $y = \dfrac{x}{\beta}$ to show that the expected value of the gamma distribution is

$$\mu = \frac{\beta}{\Gamma(\alpha)} \int_0^\infty y^\alpha \, e^{-y} \, dy \, .$$

   (b) Use the fact that $y^\alpha = y^{(\alpha+1)-1}$ along with the definition of the gamma function to simplify the integral from part (a), getting the mean for the gamma distribution. (Your answer should be just an algebraic expression involving α and β.)

6. (a) Use a procedure like that of the previous exercise to find $E(X^2)$.

   (b) Use your answer to (a) along with the result of 3(b) to find the variance of the gamma distribution.

> For the gamma distribution, $\mu = \alpha\beta$ and $\sigma^2 = \alpha\beta^2$.

It is not always the case that α and β can be determined as they were for Exercises 1 and 2. In some situations an experiment is performed to determine α and β and the validity of the resulting gamma distribution model is checked against the data before using the model. In a previous Exercise you did this for a situation that might have involved a Poisson process. (The spread of fertilizer pellets.) Using the information given, you created a Poisson distribution to model the situation, and checked your model against the real data. In the second exercise below you will do the same thing again, but with a gamma distribution.

7. Suppose that you suspect some data follows a gamma distribution, and you want to determine α and β for the distribution. You find the mean and variance of the data to be 35 and 245, respectively.

   (a) The fact that $\mu = \alpha\beta$ means that $\alpha\beta = 35$. Multiply both sides of this equation by β, then substitute the value of σ^2 for $\alpha\beta^2$. Solve for β.

   (b) Substitute the value you obtained for β into the equation for the mean to find α.

   (c) Repeat the same process to find α and β if $\mu = 18.3$ and $\sigma^2 = 23.8$.

8. <u>(Somewhat True) Story</u>: When H_2Oman was young, he and his father used to go out early in the morning and catch unfortunate grasshoppers to use as fishing bait. (If you try this yourself, the early in the morning part is important - the grasshoppers are cold, so they don't move as fast!) Once when doing this process, older H_2Oman watched the young lad catching grasshoppers, and he (dad) recorded the times, in seconds, between when successive grasshoppers were caught. Here are the times he recorded:

30	13	29	58	14	40	17	9	27	36	21
24	11	26	33	48	19	63	35	23	47	12
15	24	32	52	19	14	44	21	36	27	

(a) Make a stem-and-leaf plot of this data, with split stems. (The top entry I found when doing a search for *stem-and-leaf plot split stems* has a nice explanation of this.)

(b) Enter the data in Excel and use Excel commands to find the mean and variance of the data. There are several variances - you want the one with a P in it's function call. *Note that the mean and variance are two parameters which characterize the distribution.*

(c) We will now assume that the data can be modelled with a gamma distribution. The gamma distribution should then have the same mean and variance as the data itself. Use the equations for the mean and variance of the gamma distribution to determine the values of the parameters α and β.

9. Now you need to check to see how well a gamma distribution with the parameters you just found models the data. You will find probabilities that the data falls in certain intervals, then find the predicted probabilities for the same intervals, using the gamma distribution.

(a) If you randomly select one of the data values above, what is the probability that it will lie in the interval $[15, 20)$? (Note that this interval includes 15, but not 20.) Put that value in the column for true probabilities in the spreadsheet. Then figure out how to use the spreadsheet functions to get all the probabilities for that column.

(b) Find the Excel gamma distribution and use it to compute the values for the modeled probability column. *Remember that you are dealing with a continuous distribution, so find the probabilities appropriately.* (Can you say "cumulative distribution?")

(c) To get a visual of how well the model matches the real data, you might want to plot the values from the two columns side by side. I think you can easily figure out how to do this. How does the model look?

4.6 A Summary of Distributions

> **Performance Criteria:**
>
> 4. (g) Choose and apply the appropriate distribution(s) to solve an applied problem.

When solving a problem there are two decisions to be made:

1) Which probability distribution to use.

2) Whether to use the probability distribution/density function (pdf), or the cumulative distribution function (cdf).

Let's address the second decision first:

- If the distribution to be used it is discrete, one should likely use the pdf if a probability is needed for only one, or maybe just a few, value(s) of the random variable. Otherwise the cdf will likely be easier to use.

- If the distribution to be used is continuous, the following guidelines will likely be effective:

  ◇ In the case of the normal distribution, one would likely use the cdf in the form of the normal distribution tables, or as a tool in *Excel*. The pdf could also be used directly with a tool that can approximate integrals of it using numerical integration.

  ◇ In the case of the exponential distribution, the CDF has a simple form that can easily be applied.

  ◇ The Poisson and gamma distributions are easy to work with in *Excel*, using their cdfs.

This now leaves us with the question of which specific distribution (or distributions) should be chosen for a given problem. Here are some guidelines for making a selection:

- First identify the random variable. If it consists of a count, then the distribution of interest will be discrete. If it consists of a measurement of a continuous (in theory) quantity, then the distribution of interest will be continuous.

- If one is determining the probability of a certain number of "successes" in a fixed number of trials, either the binomial or hypergeometric distribution should be used. The binomial distribution is used when items are selected with replacement, and the hypergeometric distribution is used when items are selected without replacement.

- The binomial distribution is also appropriate when determining the probability of a number of successes in a sample where the probability of success is the same for every item in the sample.

- If the number of successes is fixed in a Bernoulli process (like choosing with replacement), and trials are to be continued until that number of successes is attained, the negative binomial distribution should be used.

- If one is determining the probability of a number of successes in a fixed length of time, or in a fixed length, area or volume, the Poisson distribution is used.

- If the random variable is continuous and normally distributed, we use (of course!) the normal distribution.
- When calculating the probability of a given range of lengths of time to the first success for a Poisson process, the exponential distribution is used. If instead we are calculating the probability of a given range of lengths of time to a certain fixed number of successes in Poisson process, the gamma distribution is used.

All of the distributions we have encountered are summarized in the table on the next page. In the Chapter Exercises you will find a selection of scenarios for which you will select and apply the appropriate distribution.

Some Commonly Used Distributions

Distribution	Assumptions/ Conditions	Random Variable	Parameters	Notation	Formula
Binomial	Bernoulli Process - sampling with replacement	Number of successes in a fixed number of trials	Number n of trials, probability p of success, $q = 1 - p$	$b(x;n,p)$ $B(x;n,p)$	$b(x;n,p) = \binom{n}{x} p^x q^{n-x}$
Negative Binomial	Repeated independent trials with constant probability of success	Number of trials to a fixed number of successes	Number k of successes, probability p of success, $q = 1 - p$	$b^*(x;k,p)$ $B^*(x;k,p)$	$b^*(x;k,p) = \binom{x-1}{k-1} p^k q^{x-k}$
Hypergeometric	Sampling without replacement	Number of successes in a fixed number of trials	Number N of objects to be drawn from, number k of the N objects that are considered successes, number n of trials	$h\,(x\,;n,k,N)$ $H\,(x\,;n,k,N)$	$h\,(x\,;n,k,N) = \dfrac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$
Poisson	Poisson Process - successes uniformly distributed in time or space	Number of successes in a fixed length t of time or a fixed length/area/volume (also denoted by t)	Average number λ of successes per unit of time/length/area/volume, t units of time/length/ area/volume	$p(x;\lambda t)$ $P(x;\lambda t)$	$p(x;\lambda t) = \dfrac{e^{-\lambda t}(\lambda t)^x}{x!}$
Normal	Normally distributed continuous data	Data value	Mean μ and standard deviation σ	$n(x;\mu,\sigma)$ $N(x;\mu,\sigma)$	$n(x;\mu,\sigma) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$
Exponential	Poisson Process	Time to first success	Average length of time β between successes	$f(x)$, $F(x)$	$f(x) = \dfrac{1}{\beta} e^{-\frac{x}{\beta}}$ for $x > 0$ $F(x) = 1 - e^{-\frac{x}{\beta}}$ for $x > 0$
Gamma	Poisson Process	Time to some number of successes	Average length of time β between successes, number α of successes	$f(x)$, $F(x)$	$f(x) = \dfrac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}}$ for $x > 0$

4.7 Chapter 4 Exercises

For each exercise, provide the combination of the following that its asked for.

(i) A probability statement giving the event of interest, in terms of the random variable X.

(ii) An expression in terms of the pdf, with all parameters and the variable substituted, whose value is the desired probability.

(iii) An expression in terms of the *formula for the pdf*, with all parameters and the variable substituted, whose value is the desired probability.

(iv) An expression in terms of the cdf, with all parameters and the variable substituted, whose value is the desired probability.

(v) The desired probability, as a decimal rounded to four places past the decimal.

DO NOT label these things with the letters above. They should be strung together in order, with equal signs between them.

1. A couple wants to have two boys, but have also decided that they wish to have no more than four children. The probability in the general population of having a boy is 0.52. Assume that this probability holds for the couple, and that the gender of any child born is independent of previous children born to the same parents.

   (a) What is the probability that the couple will need to have four children in order to get two boys? Provide *i*, *ii*, *iii*, *iv*, *v*.

   (b) What is the probability that the couple will not obtain two boys by the time they have their fourth child? Provide *i*, *ii*, *iii*, *iv*, *v*.

2. Consider again the couple from the previous exercise. Suppose that they had decided they were going to have (exactly) three children, What is the probability that will get (at least) two boys? Provide *i*, *ii*, *iii*, *iv*, *v*.

3. A particular highway has an average of 3 fatal accidents per year. What is the probability of two fatal accidents over a six month period? Provide *i*, *ii*, *iii*, *iv*, *v*.

4. The manager of the Rubber Hits The Road tire shop, Bud, knows that on the first day the snow flies, customers arrive to have their snow tires put on at a rate of 13 per hour. Given that Rubber Hits The Road is open from 8:00 AM to 5:00 PM, what is the probability that they will change 125 sets of tires on the first day that it snows? Provide *i*, *ii*, *iii*, *iv*, *v*.

5. GW likes to have strawberries on his cereal in the morning, and he occasionally finds that some of the strawberries he bought are moldy. If (unbeknownst to him) 4 of the fifty strawberries in his refrigerator are moldy and he selects ten of them for breakfast, what is the probability that

   (a) exactly one of them is moldy?

   (b) two or more of them are moldy?

   Provide *i*, *ii*, *iii*, *iv*, *v* for each.

6. An assembly line worker's job is to install a particular part in a device, a task which they can do with a probability of success of 0.78 on each attempt. (Assume that a success on one attempt is independent of success on all other previous or future attempts.)

(a) Suppose that on a particular day the worker needs to install five such parts. What is the probability that it will take them exactly 8 attempts to do so? Provide *i*, *ii*, *iii*, *v*.

(b) Suppose that on another day they need to install seven such parts. What is the probability that it will take them between 10 and 15 (inclusive, meaning including both of those values) attempts to do so? Provide *i*, *ii*, *iv*. Excel does not provide the cdf for this distribution, so an answer must be obtained using the pdf. We'll go over the most efficient way to get a value for this in class on Wednesday.

7. When backpacking in the Wind River Mountains of Wyoming, one evening mosquitoes were landing on me at a rate of 12 per minute. *Think carefully about what the variable is in each of parts (b) and (c)!*

   (a) What was the average length of time between mosquitoes landing?

   (b) What is the probability that, from the time one mosquito landed, the next mosquito would land in between 4 and 7 seconds? Tell which distribution you use and give some indication of how it is used to obtain your answer. Round to four places past the decimal.

   (c) What is the probability that fewer than 20 mosquitoes would land in a two minute period?

8. A manufacturer of a certain part claims that only one part in 50 will fail under a pressure of 500 psi. You are going to test this claim by randomly selecting five parts from a batch of 300 and testing them for failure. A part will not be replaced in the batch after testing.

   (a) What is the probability that exactly one of the five parts will fail? Provide *i*, *ii*, *iii*, *v*.

   (b) What is the probability that *at least* one of the five parts will fail? Provide *i*, *ii*, *v*. *Give ii without using a summation.*

9. A rigged deck of cards has only 10 each of hearts, diamonds and clubs, with 9 extra spades to make up for the missing cards.

   (a) If ten cards are selected at random, with replacement, what is the probability that exactly five of them are spades? Provide *i*, *ii*, *iii*, *v*.

   (b) 100 cards are selected at random, with replacement. What is the probability that between 30 and 60 of them, inclusive, are spades? Provide *i*, *ii*, *iv*, *v*.

   (c) Cards are to be drawn, with replacement, until the third spade is obtained. What is the probability that exactly seven cards must be drawn to obtain the third spade? Provide *i*, *ii*, *iii*, *v*.

10. A device being made contains spacers that must be of thicknesses between 0.05 inches and 0.07 inches. In the past you have found that thicknesses of spacers from a supplier are normally distributed, with mean 0.0608 inches and standard deviation 0.0047. What is the probability that a randomly selected spacer meets the requirement of being between 0.05 inches and 0.07 inches in thickness? Provide *i*, *ii*, *iv*, *v*.

11. Give the probability asked for in the previous exercise in terms of the standard normal distribution (mean zero and standard deviation one) cdf, with values rounded to the hundredth's place, Then determine the probability using the standard normal distribution table. Your answer should be close to what you got in the previous exercise.

12. (a) During summer mornings, the average time between vehicle arrivals at the entrance to Lava Beds National Monument is 18 minutes. If the attendant wants to read a magazine article that takes 15 minutes to read without being interrupted, what is the probability that they will have time to do this? (*Remember that you are dealing with a continuous distribution for this situation.*)

    (b) At the same time but about 100 miles away, the average time between vehicles arriving at Crater Lake National Park is 43 seconds. After the arrival of one carload of visitors, what is the probability that the next carload will arrive sometime between 30 and 50 seconds later?

13. Recall the following situation, from a previous exercise: An assembly line worker's job is to install a particular part in a device, a task which they can do with a probability of success of 0.78 on each attempt. (Assume that a success on one attempt is independent of success on all other previous or future attempts.) Suppose that they need to install ten such parts a day.

    (a) What is the probability that it will take them 12 or more tries to install ten such parts? Provide *i*, *iv*, *ii*, *v* **in that order**. What you provide for *ii* should be a slight modification of what you provide for *iv*.

    (b) What is the probability that, during a five day work week, it will take the worker 12 or more tries to install all the parts on exactly three of the days? **You will need to use your result from (a), along with a different distribution.** Provide *i*, whichever of *ii* and *iv* is more appropriate, and *v*.

    (c) What is the probability that, during a five day work week, it will take the worker *less than* 12 tries to install all the parts on one or two of the days? Provide *i*, *ii*, *v*.

14. Consider again the spacers that must be of thicknesses between 0.05 inches and 0.07 inches, which are normally distributed, with mean 0.0608 inches and standard deviation 0.0047.

    (a) What is the probability that a randomly selected spacer has a thickness of 0.055 inches or more? Provide *i*, *ii*, *iv*, *v*.

    (b) In the last assignment you should have found that the probability of selecting a spacer meeting the desired specifications to be 0.9641. If you checked each spacer before installing it, what is the probability that you would select 20 satisfactory spacers before obtaining your first one that doesn't meet the specifications? Provide *i*, *ii*, *iii*, *v*.

15. A brand of solar panels has small "flaws" or "blemishes" at a rate of 0.037 flaws per square foot.

    (a) What is the probability that a 3 foot by 10 foot panel will have 4 or more flaws? Provide *i*, *ii*, *iv*, *v*.

    (b) How many flaws would you expect a panel to have, on average?

16. On average, 23 trucks pass through a weigh station each hour. Assuming that the arrival of trucks at the weigh station is a Poisson process, what is the probability that a person working the weigh station would see 175 or fewer trucks pass though during an eight hour day of work? Provide *i*, *ii*, *iv*, *v*.

5 Joint Probability Distributions

Performance Criteria:

5. Understand and work with joint distributions.

   (a) Create a joint distribution table for a discrete probability distribution of two random variables.

   (b) Give probabilities of events for a discrete probability distribution of two random variables.

   (c) Find marginal probabilities for a discrete joint distribution of two random variables.

   (d) Give the two marginal distributions for a discrete joint distribution of two random variables.

   (e) Give a conditional probability associated with a discrete joint distribution of two random variables.

   (f) Give a conditional probability distribution for a discrete joint distribution of two random variables.

   (g) Determine whether two random variables are independent.

   (h) Find the value of a constant for which a given function of two variables is a joint probability density function.

   (i) Find probabilities for a continuous joint distribution.

   (j) Give the two marginal distributions for a continuous joint distribution of two random variables.

   (k) Give a conditional probability distribution for a continuous joint distribution of two random variables.

   (l) Determine whether two continuous random variables are independent.

   (m) Find the expected value and covariance of a discrete joint probability distribution.

   (n) Find the expected value and covariance of a continuous joint probability distribution.

   (o) Calculate probabilities using a multinomial probability distribution or multivariate hypergeometric distribution.

In this chapter we study joint probability distributions, which arise when we are considering two random variables having the same sample space.

5.1 Discrete Joint Distributions

Performance Criteria:

5. (a) Create a joint distribution table for a discrete probability distribution of two random variables.

   (b) Give probabilities of events for a discrete probability distribution of two random variables.

The table below shows political registrations of 200 residents of the state of Montana.

		PARTY REGISTRATION		
		Dem	**Rep**	**Ind**
GENDER	**Male**	35	45	24
	Female	47	33	16

1. Consider the experiment of randomly selecting one person out of this group. What is the probability that they are

   (a) male? (b) Republican? (c) male and Republican?

   (d) male or Republican? (e) male, given that they are Republican?

   (f) Republican, given that they are male?

The above type of table is sometimes called a **two-way table**. Using terminology analogous to what we used when we have worked with *Excel*, we call each of the spaces occupied by the numbers in the table **cells**. The numbers in the cells are called **cell counts**. Note that each cell represents a particular gender/party registration combination.

Considering the experiment of randomly selecting an individual from this group, let us define two random variables X and Y. X assigns to any outcome (individual selected) a number representing their party registration (zero for Democrat, one for Republican, two for Independent). Y assigns to each outcome a zero or one depending on gender (zero for male, one for female). We can make a new table based on the values of the random variables, with the number in each cell being the probability that an individual who is randomly selected from the group has the gender and party registration that the cell represents.

		X		
		0	**1**	**2**
Y	**0**	0.175	0.225	0.120
	1	0.235	0.165	0.080

2. Give each of the following probabilities. Take $(X = x, Y = y)$ to mean $X = x$ *and* $Y = y$.

   (a) $P(X = 1, Y = 0)$ (b) $P(X = 1)$ (c) $P(Y = 1)$

   (d) $P(X = 1 \mid Y = 0)$ (e) $P(X = 2 \text{ or } Y = 1)$

We will now define a function f of two variables, called a **joint probability distribution** (in this case, a *discrete* joint probability distribution), by

$$f(x, y) = P(X = x, Y = y) \qquad \text{for} \qquad x = 0, 1, 2, \quad y = 0, 1$$

Thus, $f(2, 0) = 0.120$, $f(1, 1) = 0.165$ and so on. At the risk of being a bit redundant, we now show a table of values for the function f:

		x		
	$f(x, y)$	0	1	2
y	0	0.175	0.225	0.120
	1	0.235	0.165	0.080

◇ **Example 5.1(a):** Give each of the probabilities from Exercise 2 in terms of the probability function f.

The only one that might need some explanation is $P(X = 1 \mid Y = 0)$, where we must divide the probability $P(X = 1, Y = 0)$ by $P(Y = 0)$:

$$P(X = 1 \mid Y = 0) = \frac{P(X = 1, Y = 0)}{P(Y = 0)} = \frac{f(1, 0)}{f(0, 0) + f(1, 0) + f(2, 0)}.$$

In the next section we will encounter a concept that will allow us to give a simpler expression for the denominator. Here are the other probabilities in terms of f:

$$P(X = 1, Y = 0) = f(1, 0) \qquad\qquad P(X = 1) = f(1, 0) + f(1, 1)$$

$$P(Y = 1) = f(0, 1) + f(1, 1) + f(2, 1)$$

$$P(X = 2 \text{ or } Y = 1) = f(2, 0) + f(2, 1) + f(0, 1) + f(1, 1).$$

The concept from the next section will allow us to give a cleaner expression for this last probability as well.

3. Consider the experiment of rolling a single die (one half of a pair of dice).

   (a) Give the sample space.

   (b) Define a random variable X on the sample space by $X = 0$ if the number rolled is odd, $X = 1$ if the number rolled is even. Give $\text{Ran}(X)$.

(c) Define the random variable Y on the same sample space to be the number of letters in the spelling of the number rolled. For example, $Y(5) = 4$ because the word *five* has four letters. Give $\text{Ran}(Y)$.

(d) Create a table for the joint probability distribution $f(x, y)$. It should have the form of the table on the previous page, but give all probabilities as fractions.

We now make a formal definition, based on things that should be intuitively clear. First, though, we recall that for two sets A and B, we define the **Cartesian product** of A and B, denoted by $A \times B$, to be the set of all ordered pairs (x, y) for which x is in A and y is in B. Recall also that $\text{Ran}(X)$ is the range of X, and similarly for Y.

Discrete Joint Probability Distribution

Let X and Y be random variables *defined on the same discrete sample space*. A function f, defined on $\text{Ran}(X) \times \text{Ran}(Y)$, is a **joint probability distribution** if

1) $f(x, y) \geq 0$ for all $(x, y) \in \text{Ran}(X) \times \text{Ran}(Y)$,

2) $\displaystyle\sum_{\text{Ran}(X)} \sum_{\text{Ran}(Y)} f(x, y) = 1$,

3) $P(X = x, Y = y) = f(x, y)$ for all $(x, y) \in \text{Ran}(X) \times \text{Ran}(Y)$.

Although the following theorem may be obvious, we state it because of its importance when we introduce continuous joint distributions.

Theorem 5.1: Let f be a discrete joint probability distribution for the random variables X and Y, and let A be any subset of $\text{Ran}(X) \times \text{Ran}(Y)$. Then

$$P((X, Y) \in A) = \sum_{(x,y) \in A} \sum f(x, y)\,.$$

4. The following table gives the joint probability distribution function f for two discrete random variables X and Y.

		x			
	$f(x, y)$	0	1	2	3
y	0	$\frac{5}{24}$	$\frac{4}{24}$	$\frac{3}{24}$	$\frac{2}{24}$
	1	$\frac{3}{24}$	$\frac{2}{24}$	$\frac{1}{24}$	$\frac{1}{24}$
	2	$\frac{2}{24}$	$\frac{1}{24}$	0	0

Give each of the following probabilities of events, first as an expression in terms of f evaluated at specific values of x and y, then give a numerical answer, **as a fraction**. Take $(X = a, Y = b)$ to mean $X = a$ *and* $Y = b$, and take $P(X = a \mid Y = b)$ to mean the conditional probability of $X = a$ given that $Y = b$. **You might consider giving one of the probabilities in terms of its complementary event**.

(a) $P(X = 1, Y = 0)$ (b) $P(X = 1)$ (c) $P(X + Y \leq 1)$

(d) $P(X = 2 \text{ or } Y = 1)$ (e) $P(X + Y \leq 3)$ (f) $P(Y = 2|X = 0)$

5. (a) Suppose that three marbles are to be drawn *without replacement* from an urn containing three blue marbles, two red marbles and five yellow marbles. Let X be the number of blue marbles drawn and Y be the number of red marbles drawn. Give a table like the one above for the joint probability distribution, *giving all probabilities in fraction form.* It is suggested that you make all fractions have denominators of 720; check your answers by making sure that the sum of all the probabilities is one.

   (b) Repeat part (a) under the conditions that the marbles are drawn *with replacement.* Here I would suggest that you make the denominators of all the fractions 1000.

5.2 Discrete Marginal Distributions

Performance Criteria:

5. (c) Find marginal probabilities for a discrete joint distribution of two random variables.

   (d) Give the two marginal distributions for a discrete joint distribution of two random variables.

Consider again the table from the last section, but with probabilities in place of counts:

		PARTY REGISTRATION		
		Dem	**Rep**	**Ind**
GENDER	**Male**	0.175	0.225	0.120
	Female	0.235	0.165	0.080

This table gives us probabilities of specific party/gender combinations, but does not directly give probabilities such as whether a randomly selected individual will be Republican, or be male, and so on. In terms of the random variables X and Y, we are asking for $P(X = 1)$ and $P(Y = 0)$. It should be obvious that to compute probabilities involving only one of the random variables X or Y one needs to add up the values in each each column and row of the distribution table:

		X			
		0	**1**	**2**	
Y	**0**	0.175	0.225	0.120	0.520
	1	0.235	0.165	0.080	0.480
		0.410	0.390	0.200	1.000

These additional probabilities are called **marginal probabilities**. They give the values of two additional distributions g and h defined by

$$g(x) = P(X = x) \text{ for } x = 0, 1, 2 \qquad \text{and} \qquad h(y) = P(Y = y) \text{ for } y = 0, 1\,.$$

For this particular example, g and h have the distributions

x	0	1	2
$g(x)$	0.41	0.39	0.20

y	0	1
$h(y)$	0.52	0.48

The functions g and h are called the **marginal distributions** associated with the joint distribution f defined in the last section. The value of 1.000 in the lower right is the sum of the probabilities for both g and h - as it should be, the total probability for each of those marginal distributions is one. For any joint distribution f, we define g and h more formally as follows.

Marginal Distributions

Let f be a discrete joint probability distribution for the random variables X and Y. We define the **marginal distribution functions** g and h on $\mathrm{Ran}(X)$ and $\mathrm{Ran}(Y)$, respectively, by

$$g(x) = \sum_{\mathrm{Ran}(Y)} f(x,y) \quad \text{and} \quad h(y) = \sum_{\mathrm{Ran}(X)} f(x,y)\,.$$

Note carefully what these two formulas say. To compute the value of g for a particular x, we fix the value of x and add up all the values of $f(x,y)$ as y ranges over all possible values of the random variable Y. A similar process is carried out for finding values of h. For the example we have been looking at, this means that

$$g(x) = f(x,0) + f(x,1) \qquad \text{and} \qquad h(y) = f(0,y) + f(1,y) + f(2,y)\,.$$

We can perhaps see this more readily by adding the marginal probabilities to our table for the distribution function $f(x,y)$:

		x			
	$f(x,y)$	0	1	2	$h(y)$
y	0	0.175	0.225	0.120	0.520
	1	0.235	0.165	0.080	0.480
	$g(x)$	0.410	0.390	0.200	

◇ **Example 5.2(a):** Three marbles are drawn, without replacement, from an urn containing 3 blue, 2 yellow and 5 red marbles. Let X be the random variable that assigns to any outcome the number of red marbles drawn, and let Y assign to any outcome the number of yellow marbles drawn. The table for the joint probability distribution f is shown below. Determine the marginal distributions $g(x)$ and $h(y)$.

		x			
	$f(x,y)$	0	1	2	3
y	0	$\frac{1}{120}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{12}$
	1	$\frac{1}{20}$	$\frac{1}{4}$	$\frac{1}{6}$	0
	2	$\frac{1}{40}$	$\frac{1}{24}$	0	0

To get the marginal distributions, we simply sum up the rows and columns to obtain the following table:

		x				
	$f(x,y)$	0	1	2	3	$h(y)$
	0	$\frac{1}{120}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{12}$	$\frac{7}{15}$
y	1	$\frac{1}{20}$	$\frac{1}{4}$	$\frac{1}{6}$	0	$\frac{7}{15}$
	2	$\frac{1}{40}$	$\frac{1}{24}$	0	0	$\frac{1}{15}$
	$g(x)$	$\frac{1}{12}$	$\frac{5}{12}$	$\frac{5}{12}$	$\frac{1}{12}$	1

From this we can see that the marginal distributions are

x	0	1	2	3
$g(x)$	$\frac{1}{12}$	$\frac{5}{12}$	$\frac{5}{12}$	$\frac{1}{12}$

y	0	1	2
$h(y)$	$\frac{7}{15}$	$\frac{7}{15}$	$\frac{1}{15}$

◇ **Example 5.2(b):** For the distribution from the previous example, give each of the following probabilities in terms as efficiently as possible in terms of f, g and h.

$$P(X=2) \qquad P(X=1 \mid Y=2) \qquad P(X=1 \text{ or } Y=2)$$

Using the definitions of f, g, h and conditional probability, along with the addition rule, we have

$$P(X=2)=g(2) \qquad P(X=1 \mid Y=2)=\frac{f(1,2)}{h(2)}$$

$$P(X=1 \text{ or } Y=2)=g(1)+h(2)-f(1,2)$$

---

1. Add the marginal probabilities to your table from Exercise 3(d) of the previous section, then give the marginal distribution g of X and the marginal distribution h of Y.

2. Consider again the joint distribution for gender and political affiliation. Use to to give each of the following probabilities.

   (a) $P(X=2)$ in terms of f

   (b) $P(X=2)$ in terms of g or h

   (c) $P(X=1, Y=0)$ in terms of f

   (d) $P(Y=0 \mid X=1)$ in terms of f and g or h

   (e) $P(X=1 \text{ or } Y=0)$ in terms of f alone

   (f) $P(X=1 \text{ or } Y=0)$ in terms of f, g and h

3. The following table gives the joint probability distribution function f for two discrete random variables X and Y.

		x			
	$f(x,y)$	0	1	2	3
y	0	$\frac{5}{24}$	$\frac{4}{24}$	$\frac{3}{24}$	$\frac{2}{24}$
	1	$\frac{3}{24}$	$\frac{2}{24}$	$\frac{1}{24}$	$\frac{1}{24}$
	2	$\frac{2}{24}$	$\frac{1}{24}$	0	0

Give the two marginal distributions $g(x)$ and $h(y)$.

x : y :

$g(x)$: $h(y)$:

4. Continue using the joint distribution given by the table in the previous exercise. Give each of the following probabilities in terms of the marginal distributions g and h, and use the joint distribution f as well, when necessary. Then give a numerical value for each, as a fraction.

(a) $P(X = 2)$ (b) (b) $P(Y = 2)$ (c) $P(Y \leq 1)$

(d) $P(X = 2 \text{ or } Y = 1)$ (e) $P(X = 3|Y = 1)$ (f) $P(Y = 2|X = 0)$

3. Add the marginal probabilities to your table from Exercise 5(a) of the previous section, then give the marginal distribution g of X and the marginal distribution h of Y.

5.3 Discrete Conditional Distributions, Independent Random Variables

Performance Criteria:

5. (e) Give a conditional probability associated with a discrete joint distribution of two random variables.
   (f) Give a conditional probability distribution for a discrete joint distribution of two random variables.
   (g) Determine whether two random variables are independent.

Consider one last time the table for the Montanans:

		PARTY REGISTRATION		
		Dem	**Rep**	**Ind**
GENDER	**Male**	35	45	24
	Female	47	33	16

Suppose that we wish to determine the probability that a randomly selected male voter from this group would be a Democrat. Because there are $35 + 45 + 24 = 104$ males, the probability of selecting a Democrat from among them is $\frac{35}{104} = 0.337$. You'll recall that what we have computed here is a **conditional probability**, in this case the probability that one of the voters is a Democrat, *given that they are male*. Suppose that we wished to determine the same probability from the table of values for the joint distribution function $f(x, y)$:

		x			
	$f(x,y)$	0	1	2	$h(y)$
y	0	0.175	0.225	0.120	0.520
	1	0.235	0.165	0.080	0.480
	$g(x)$	0.410	0.390	0.200	

In this case we would compute the desired probability by taking $\frac{0.175}{0.520} = 0.337$. You can see here that the quotient that gives us the desired probability is

$$\frac{f(0,0)}{h(0)}$$

We would find the probability of a voter being Republican, given that they are male, to be $\frac{f(1,0)}{h(0)}$ and the probability that a voter is Independent, given that they are male, to be $\frac{f(2,0)}{h(0)}$. These three values are the probabilities of a probability distribution function that we will denote by $f(x \mid 0)$. This is defined more formally on the next page.

Discrete Conditional Distributions

Let f be a discrete joint probability distribution for the random variables X and Y with marginal distributions g and h. For a fixed $y \in \text{Ran}(Y)$ we define the **conditional distribution function** $v(* \mid y)$ on $\text{Ran}(X)$ by

$$v(x \mid y) = \frac{f(x,y)}{h(y)} \quad \text{if} \quad h(y) > 0\,.$$

Similarly, for a fixed $x \in \text{Ran}(X)$, if $g(x) > 0$ the conditional distribution function $w(* \mid x)$ is defined by

$$w(y \mid x) = \frac{f(x,y)}{g(x)}$$

for all $y \in \text{Ran}(Y)$

The purpose of writing $v(* \mid y)$ instead of $v(x \mid y)$ is to indicate that x is a variable, whereas y is fixed (for the purposes of the conditional distribution).

◇ **Example 5.3(a):** Find the conditional distributions $v(x \mid 0)$ and $w(y \mid 2)$ for the probability distribution from Example 5.2(a), shown below.

		x				
	$f(x,y)$	0	1	2	3	$h(y)$
y	0	$\frac{1}{120}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{12}$	$\frac{7}{15}$
	1	$\frac{1}{20}$	$\frac{1}{4}$	$\frac{1}{6}$	0	$\frac{7}{15}$
	2	$\frac{1}{40}$	$\frac{1}{24}$	0	0	$\frac{1}{15}$
	$g(x)$	$\frac{1}{12}$	$\frac{5}{12}$	$\frac{5}{12}$	$\frac{1}{12}$	1

Here we have $v(0|\,0) = \frac{1}{120}/\frac{7}{15} = \frac{1}{56}$, $v(1|\,0) = \frac{1}{8}/\frac{7}{15} = \frac{15}{56}$, $v(2|\,0) = \frac{1}{4}/\frac{7}{15} = \frac{15}{28}$ and $v(3|\,0) = \frac{1}{12}/\frac{7}{15} = \frac{5}{28}$, so the function $v(x|\,0)$ is given by

x	0	1	2	3
$v(x\|\,0)$	$\frac{1}{56}$	$\frac{15}{56}$	$\frac{15}{28}$	$\frac{5}{28}$

Similar computations give us

y	0	1	2
$w(y\|\,2)$	$\frac{3}{5}$	$\frac{2}{5}$	0

You will recall that if two events A and B are independent, the occurrence of B makes it no more or less likely that A will occur. That is, $P(A|B) = P(A)$. This leads us to the definition that A and B are independent if, and only if, $P(A \cap B) = P(A)P(B)$. We define independent random variables in a similar manner.

Independent Random Variables

Let X and Y be discrete random variables with joint distribution function f and marginal distribution functions g and h, respectively. If

$$f(x, y) = g(x)h(y)$$

for all $x \in \text{Ran}(X)$ and $y \in \text{Ran}(Y)$, then we say that the random variables X and Y are independent.

◇ **Example 5.3(b):** Determine whether the random variables X and Y from the distribution given in Example 5.3(a) (the table for the distribution is reproduced below) are independent.

		x				
	$f(x,y)$	0	1	2	3	$h(y)$
y	0	$\frac{1}{120}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{12}$	$\frac{7}{15}$
	1	$\frac{1}{20}$	$\frac{1}{4}$	$\frac{1}{6}$	0	$\frac{7}{15}$
	2	$\frac{1}{40}$	$\frac{1}{24}$	0	0	$\frac{1}{15}$
	$g(x)$	$\frac{1}{12}$	$\frac{5}{12}$	$\frac{5}{12}$	$\frac{1}{12}$	1

We see that $g(0)h(0) = \frac{7}{15}\frac{1}{12} = \frac{7}{180} \neq \frac{1}{120} = f(0, 0)$, so the random variables X and Y are dependent.

---

1. The following table gives the joint probability distribution function f for two discrete random variables X and Y. Give the conditional distributions $v(x|0)$ and $w(y|3)$ in the manner shown in Example 5.3(a). Are the random variables X and Y independent?

		x			
	$f(x,y)$	0	1	2	3
y	0	$\frac{5}{24}$	$\frac{4}{24}$	$\frac{3}{24}$	$\frac{2}{24}$
	1	$\frac{3}{24}$	$\frac{2}{24}$	$\frac{1}{24}$	$\frac{1}{24}$
	2	$\frac{2}{24}$	$\frac{1}{24}$	0	0

5.4 Continuous Joint Probability Density Functions

Performance Criteria:

5. (h) Find the value of a constant for which a given function of two variables is a joint probability density function.

   (i) Find probabilities for a continuous joint density function.

As with the single random variable situation, when we can have two random variables that are continuous rather than discrete. In that case we again have a probability density function, but of two variables. And as in the single random variable case, evaluating such a function at a single point has no real meaning - it is only when we integrate values over some region in the $\mathbb{R}^2$ plane that we get a probability. That is, the probability has density, but no mass at any single point. Here is the definition of such a function:

Continuous Joint Probability Density Function

Let X and Y be continuous random variables defined on the same sample space. A function $f : \mathbb{R}^2 \to \mathbb{R}$ is a **joint probability density function** if

1) $f(x, y) \geq 0$ for all $(x, y) \in \mathbb{R}^2$,

2) $\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1$,

3) For any subset A of $\mathbb{R}^2$, $P[(X, Y) \in A] = \displaystyle\iint_A f(x, y)\, dx\, dy$

◇ **Example 5.4(a):** Find the value of c for which

$$f(x) = \begin{cases} cx^2y & \text{for } (x, y) \in [0, 1] \times [0, 2] \\ 0 & \text{otherwise} \end{cases}$$

is a continuous joint probability density function.

It should be clear that f meets condition (1) above. We must choose c so that condition (2) is met. We see that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\, dx\, dy = \int_0^2\int_0^1 cx^2y\, dx\, dy =$$

$$c\int_0^2 \left[\tfrac{1}{3}x^3y\right]_0^1 dy = \tfrac{1}{3}c\int_0^2 y\, dy = \tfrac{1}{3}c\left[\tfrac{1}{2}y^2\right]_0^2 = \tfrac{2}{3}c\,,$$

so c must be $\frac{3}{2}$.

◇ **Example 5.4(b):** For the continuous joint probability density function

$$f(x) = \begin{cases} \frac{3}{2}x^2y & \text{for } (x,y) \in [0,1] \times [0,2] \\ 0 & \text{otherwise} \end{cases}$$

of the previous example, find $P(\frac{1}{2} \leq X \leq 1, 0 \leq Y \leq 1)$.

Here we apply (3) from the definition, where A is the set $[\frac{1}{2}, 1] \times [0,1]$:

$$P(\tfrac{1}{2} \leq X \leq 1, 0 \leq y \leq 1) = \int_0^1 \int_{\frac{1}{2}}^1 \tfrac{3}{2}x^2y \, dx\, dy =$$

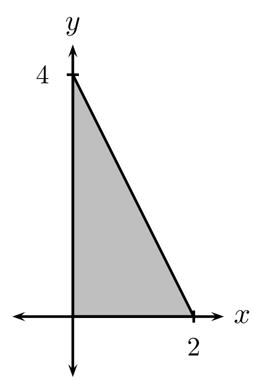
$$\int_0^1 \left[\tfrac{1}{2}x^3y\right]_{\frac{1}{2}}^1 dy = \int_0^1 \left[\tfrac{1}{2}y - \tfrac{1}{16}y\right] dy = \tfrac{7}{16}\int_0^1 y\, dy = \tfrac{7}{16}\left[\tfrac{1}{2}y^2\right]_0^1 = \tfrac{7}{32}$$

Although the integrals for both of the previous examples were computed "by hand," I would encourage you to use some technology like the **Wolfram Alpha®** *Double Integral Calculator* or your handheld calculator to make such computations.

◇ **Example 5.4(c):** Find $P(2X + Y \leq 4)$ for the continuous joint probability density function

$$f(x) = \begin{cases} 2e^{-2x-y} & \text{for } x \geq 0,\ y \geq 0 \\ 0 & \text{otherwise} \end{cases}.$$

Setting up the iterated integral for the desired probability will be easier if we first determine the region in $\mathbb{R}^2$ where we are integrating. We first consider the equation $2x + y = 4$, which is a line with x-intercept (found by setting $y = 0$) two and y-intercept four. Because the point $(0,0)$ satisfies the inequality $2x + y \leq 4$ the region of interest is the triangle shown to the right (because we only have nonzero probability in the first quadrant). We then see that the probability is given by either of the iterated integrals



$$\int_0^2 \int_0^{4-2x} 2e^{-2x-y}\, dy\, dx = \int_0^4 \int_0^{2-\frac{1}{2}y} 2e^{-2x-y}\, dx\, dy,$$

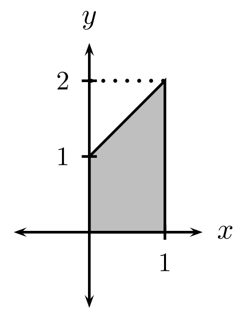
resulting in $P(2X + Y \leq 4) = 1 - 5e^{-4} \approx 0.9084$.

◇ **Example 5.4(d):** For the continuous joint probability density function

$$f(x) = \begin{cases} \frac{3}{2}x^2y & \text{for } (x,y) \in [0,1] \times [0,2] \\ 0 & \text{otherwise} \end{cases},$$

find $P(Y \leq X + 1)$.

Because the density function is nonzero only on the rectangular region $[0,1]\times[0,2]$, we need only integrate over the portion of that region that satisfies the inequality $y \le x+1$. The graph of the line $y = x+1$ has slope one and y-intercept one and we are considering the points in the rectangle that are below that line, as shown in the graph to the right. Thus the desired probability is obtained by integrating over that region:



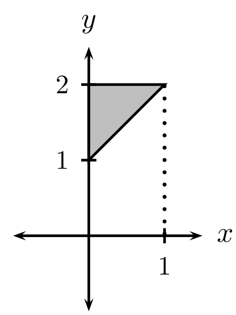
$$P(Y \le X+1) = \int_0^1 \int_0^{x+1} \tfrac{3}{2}x^2y \; dy\, dx = \frac{31}{40}$$

We should make two observations regarding the previous exercise:

- If we were to change the order of integration, *two* iterated integrals would be required. This is because when y is a fixed value between zero and one we enter the region at $x = 0$ and leave at $x = 1$, but if y is a fixed value between one and two we enter at $x = y-1$ and leave at $x = 1$. Therefore we would have

$$P(Y \le X+1) = \int_0^1 \int_0^1 \tfrac{3}{2}x^2y \; dx\, dy + \int_1^2 \int_{y-1}^1 \tfrac{3}{2}x^2y \; dx\, dy$$

- Because the density function is nonzero only on the rectangular region $[0,1]\times[0,2]$, We could have instead obtained the desired probability by integrating over the triangle shown to the right and subtracting from one:



$$P(Y \le X+1) = 1 - \int_0^1 \int_{x+1}^2 \tfrac{3}{2}x^2y \; dy\, dx = 1 - \frac{9}{40} = \frac{31}{40}$$

5.5 More on Continuous Joint Probability

Performance Criteria:

5. (j) Give the two marginal distributions for a continuous joint density function of two random variables.
   (k) Give a conditional probability distribution for a continuous joint density function of two random variables.
   (l) Determine whether two continuous random variables are independent.

Marginal Distributions

Let f be a continuous joint probability density function for the random variables X and Y. We define the **marginal distribution functions** $g:\mathbb{R}\rightarrow\mathbb{R}$ and $h:\mathbb{R}\rightarrow\mathbb{R}$, respectively, by

$$g(x)=\int_{-\infty}^{\infty} f(x,y)\ dy \qquad \text{and} \qquad h(y)=\int_{-\infty}^{\infty} f(x,y)\ dx$$

◇ **Example 5.5(a):** Find the marginal distributions g and h for the continuous joint probability density function

$$f(x)=\begin{cases} 6x^2y & \text{for } (x,y)\in[0,1]\times[0,1] \\ 0 & \text{otherwise} \end{cases}$$

From the above definitions we have

$$g(x)=\int_0^1 6x^2y\ dy=\Big[3x^2y^2\Big]_0^1=3x^2 \quad \text{for } x\in[0,1],\ 0 \text{ otherwise}$$

and

$$h(y)=\int_0^1 6x^2y\ dx=\Big[2x^3y\Big]_0^1=2y \quad \text{for } y\in[0,1],\ 0 \text{ otherwise}$$

Conditional Distributions

Let f be a continuous joint probability density function for the random variables X and Y. For a fixed $y \in \mathbb{R}$ with $h(y) > 0$ we define the **conditional distribution function** $v(\ * \mid y) : \mathbb{R} \to \mathbb{R}$ by

$$v(x \mid y) = \frac{f(x,y)}{h(y)}$$

The notation $* \mid y$ is meant to indicate that y is fixed and the x is a variable.

Similarly, for a fixed $x \in \mathbb{R}$ with $g(x) > 0$ the conditional distribution function $w(\ * \mid x) : \mathbb{R} \to \mathbb{R}$ is defined by

$$w(y \mid x) = \frac{f(x,y)}{g(x)}$$

◇ **Example 5.5(b):** Find the conditional distributions $v(x \mid y)$ and $w(y \mid x)$ for the continuous joint probability density function

$$f(x) = \begin{cases} 6x^2y & \text{for } (x,y) \in [0,1] \times [0,1] \\ 0 & \text{otherwise} \end{cases}$$

Applying the above definitions and using the results of Example 5.5(a) we have

$$v(x \mid y) = \frac{f(x,y)}{h(y)} = \frac{6x^2y}{2y} = 3x^2 \quad \text{for } x \in [0,1],\ 0 \text{ otherwise}$$

and

$$w(y \mid x) = \frac{f(x,y)}{g(x)} = \frac{6x^2y}{3x^2} = 2y \quad \text{for } y \in [0,1],\ 0 \text{ otherwise.}$$

We should make particular note of the results of this last example. Looking at the first part, we see that

$$v(x \mid y) = \frac{f(x,y)}{h(y)} = g(x).$$

This tells us that computing probabilities for the random variable Y for any fixed value of the random variable Y is the same as computing probabilities of X regardless of the value of Y; that is, X seems to be independent of Y. Multiplying both sides of the second equality by $h(y)$ results in

$$f(x,y) = g(x)h(y)$$

which, as in the case of two discrete random variables, we take to be the definition of independence:

Independent Random Variables

Let X and Y be continuous random variables with joint distribution function f and marginal distribution functions g and h, respectively. If

$$f(x,y) = g(x)h(y)$$

for all $(x,y) \in \mathbb{R}^2$, then we say that the random variables X and Y are statistically independent.

◇ **Example 5.5(c):** The continuous joint probability density function for two continuous random variables X and Y is

$$f(x) = \begin{cases} \frac{6}{5}(x^2+y) & \text{for } (x,y) \in [0,1] \times [0,1] \\ 0 & \text{otherwise} \end{cases}$$

Are X and Y independent?

First we see that

$$g(x) = \int_0^1 \tfrac{6}{5}(x^2+y)\,dy = \tfrac{6}{5}\Big[x^2y + \tfrac{1}{2}y^2\Big]_0^1 = \tfrac{6}{5}(x^2+\tfrac{1}{2}) \quad \text{for } x \in [0,1],\ 0 \text{ otherwise}$$

and

$$h(y) = \int_0^1 \tfrac{6}{5}(x^2+y)\,dx = \tfrac{6}{5}\Big[\tfrac{1}{3}x^3 + xy\Big]_0^1 = \tfrac{6}{5}(y+\frac{1}{3}) \quad \text{for } y \in [0,1],\ 0 \text{ otherwise}$$

It looks doubtful that $f(x,y) = g(x)h(y)$, but I suppose we should check to be sure:

$$g(x)h(y) = \tfrac{6}{5}(x^2+\tfrac{1}{2}) \cdot \tfrac{6}{5}(y+\frac{1}{3}) = \tfrac{36}{25}(x^2y + \tfrac{1}{3}x^2 + \tfrac{1}{2}y + \tfrac{1}{6}) \neq \tfrac{6}{5}(x^2+y)$$

Therefore the random variables X and Y are not independent.

Consider the following joint probability density function for two continuous random variables X and Y:

$$f(x,y) = \begin{cases} \frac{1}{4}(2x+y) & \text{for } (x,y) \in [0,1] \times [0,2] \\ 0 & \text{otherwise} \end{cases}$$

1. Find $P(X \geq \frac{1}{2}, 0 \leq Y \leq \frac{3}{2})$. Give the integral used, and **evaluate it by hand, giving its value in fraction form**. Check your answer using a TI-89 or the Wolfram online double integral calculator (do a search for *online double integral calculator*.

2. Find $P(X+Y \geq 1)$ as follows: (a) Sketch the region over which the probability density function must be integrated to find the desired probability. (b) Set up the integral. (c) Evaluate the integral using a TI-89 or the Wolfram double integral calculator. You should get $\frac{7}{8}$; if you don't, try to correct your integral so that you do.

3. Find and simplify the marginal distribution $h(y)$, then use it to find $P(1 \leq y \leq 2)$.

4. Give the conditional distribution $f_{X|Y}(x|y)$ in simplified form.

5. Use your answer to (d) to find $P(X \leq \frac{1}{2} \mid Y = 1)$. Do this by setting $y = 1$ and integrating over the desired range of x values.

5.6 Expected Value and Covariance of Joint Distributions

Performance Criteria:

5. (m) Find the expected value and covariance of a discrete joint probability distribution.

   (n) Find the expected value and covariance of a continuous joint probability distribution.

Expected Value of Joint Random Variables

Let X and Y be discrete random variables with joint distribution function f. The expected value of the distribution, denoted $\mu_{XY} = E(XY)$ is

$$E(XY) = \sum_{x\in\text{Ran}(X)} \sum_{y\in\text{Ran}(Y)} xyf(x,y)\,.$$

If X and Y are continuous random variables with joint distribution function f, the expected value $\mu_{XY} = E(XY)$ of the distribution is

$$E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f(x,y)\; dx\, dy.$$

◇ **Example 5.6(a):** Find the expected value of the joint probability distribution, for discrete random variables X and Y, given in the table below.

		x		
	$f(x,y)$	0	1	2
y	0	0.175	0.225	0.120
	1	0.235	0.165	0.080

$$E(XY) = (0)(0)(0.175) + (1)(0)(0.225) + (2)(0)(0.120) + (0)(1)(0.235) + (1)(1)(0.165) + (2)(1)(0.080) = 0.325$$

◇ **Example 5.6(b):** Find the expected value of the joint probability distribution, for continuous random variables X and Y, given by

$$f(x) = \begin{cases} \frac{6}{5}(x^2 + y) & \text{for } (x, y) \in [0,1] \times [0,1] \\ 0 & \text{otherwise} \end{cases}$$

$$E(XY) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f(x,y)\, dx\, dy = \int_0^1 \int_0^1 xy \cdot \tfrac{6}{5}(x^2 + y)\, dx\, dy =$$

$$\tfrac{6}{5} \int_0^1 \int_0^1 (x^3 y + xy^2)\, dx\, dy = \tfrac{6}{5} \int_0^1 \left[\tfrac{1}{4}x^4 y + \tfrac{1}{2}x^2 y^2\right]_0^1 dy =$$

$$\tfrac{6}{5} \int_0^1 (\tfrac{1}{4}y + \tfrac{1}{2}y^2)\, dy = \left[\tfrac{1}{8}y^2 + \tfrac{1}{6}y^3\right]_0^1 = \tfrac{3}{16}$$

Covariance of Joint Random Variables

Let X and Y be discrete random variables with joint distribution function f. Let μ_X and μ_Y be the expected values of the marginal distributions $g(x)$ and $h(y)$. The **covariance** σ_{XY} of X and Y is

$$\sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)] = \sum_{x \in \mathrm{Ran}(X)} \sum_{y \in \mathrm{Ran}(Y)} (x - \mu_X)(y - \mu_Y) f(x,y)\,.$$

If X and Y are continuous random variables, then the covariance is given by

$$\sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y) f(x,y)\, dx\, dy.$$

As with the variance of a single random variable, the definition of the covariance of a joint distribution does not give the most efficient way to compute the covariance. Forthat we turn to the following result.

Theorem 5.2

Let X and Y be random variables (discrete or continuous) with joint distribution function f. The covariance of X and Y is given by

$$\sigma_{XY} = E(XY) - \mu_X \mu_Y\,.$$

◇ **Example 5.6(c):** The joint probability distribution from Example 5.5(a) is given below, with the marginal distributions added. Find the covariance of the random variables X and Y.

		x			
	$f(x,y)$	0	1	2	$h(y)$
y	0	0.175	0.225	0.120	0.520
	1	0.235	0.165	0.080	0.480
	$g(x)$	0.410	0.390	0.200	

We can see that the expected values for X and Y are

$$\mu_X = (0)(0.41) + (1)(0.39) + (2)(0.20) = 0.79 \quad \text{and} \quad \mu_Y = (0)(0.52) + (1)(0.48) = 0.48$$

We computed the expected value in Example 5.5(a), so we can now compute

$$\sigma_{XY} = E(XY) - \mu_X\mu_Y = 0.325 - (0.79)(0.48) = -0.0542$$

---

1. The table below and to the right gives the joint probability distribution for two random variables X and Y. Use it for the following.

		x	
	$f(x,y)$	0	1
y	0	$\frac{5}{32}$	$\frac{15}{32}$
	1	$\frac{3}{32}$	$\frac{9}{32}$

   (a) Give $g(x)$ and $h(y)$ in table form, *with all values reduced.*

   (b) Compute $\sum_{x=0}^{1} x\, g(x)$. *Reduce your answer.*

   (c) Compute $\sum_{x=0}^{1}\sum_{y=0}^{1} x\, f(x,y)$; *reduce again.*

   (d) Your answers to (b) and (c) should be the same, and both are μ_X. μ_Y can be computed analogously, using either method. Find μ_Y.

   (e) Find the expected value $E(XY)$.

   (f) Find the covariance σ_{XY} using the formula $\sigma_{XY} = E(XY) - \mu_X\mu_y$.

   (g) Are X and Y independent? Show/explain. **This requires four mathematical statements!**

2. Give the expected value and covariance for the following joint probability density function for two continuous random variables X and Y:

$$f(x,y) = \begin{cases} \frac{1}{4}(2x+y) & \text{for } (x,y) \in [0,1]\times[0,2] \\ 0 & \text{otherwise} \end{cases}$$

5.7 Multinomial and Multivariate Hypergeometric Distributions

Performance Criteria:

5. (o) Calculate probabilities using a multinomial probability distribution or multivariate hypergeometric distribution.

Consider an experiment with the following characteristics:

- It consists of n repeated trials.
- Each trial can result in any of k outcomes $E_1, E_2, E_3, ..., E_k$
- The probabilities of the outcomes of the outcomes, which we will denote $p_1, p_2, p_3, ..., p_k$, respectively, remain the same for each trial.
- The outcome of any single trial is independent of the outcomes of the previous trials.

We will refer to such an experiment as a **multinomial experiment**. Note that a Bernoulli experiment is simply a multinomial experiment for which $k = 2$.

Multinomial Probability

For a multinomial experiment with outcomes and probabilities as defined above, let $X_1, X_2, X_3, ..., X_k$ be random variables that assign the number of times the outcomes $E_1, E_2, E_3, ..., E_k$ occur, respectively. Then the joint probability distribution for these random variables is

$$f(x_1, x_2, x_3, ..., x_k; p_1, p_2, p_3, ..., p_k; n) = \left(\begin{array}{c} n \\ x_1, x_2, ..., x_k \end{array} \right) p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}.$$

where

$$\left(\begin{array}{c} n \\ x_1, x_2, ..., x_k \end{array} \right) = \frac{n!}{x_1! x_2! \cdots x_k!}$$

It is instructive to compare this with the formula for the binomial distribution.

The multinomial distribution would apply to an experiment like drawing ten marbles, with replacement, from an urn containing 30 red marbles, 20 blue marbles, 40 yellow marbles and green marbles. In that case, the probability of getting two red, two blue, five yellow and one green marble is

$$f(2, 2, 5, 1; \frac{3}{10}, \frac{2}{10}, \frac{4}{10}, \frac{1}{10}; 10) = \frac{10!}{2!2!4!1!} \left(\frac{3}{10}\right)^2 \left(\frac{2}{10}\right)^2 \left(\frac{4}{10}\right)^5 \left(\frac{1}{10}\right)^1.$$

Suppose that we had the same urn, same number of draws, but we were drawing *without replacement*. You might guess that some variation of the hypergeometric distribution would be used, and you are correct. Note that the hypergeometric situation can be interpreted as follows: We are drawing from N objects with only two possible outcomes, with respective numbers k and

$N - k$ in the original urn. We draw n objects and want the probability of getting x of the first outcome and $n - x$ of the second outcome. The probability is given by

$$\frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$$

Lets change notation a bit by letting a_1 be the number of the N total objects that are of the first kind, and let a_2 be the number of objects of the second kind. Also, let x_1 and x_2 be the numbers of each of those kinds of objects that we ant the probability of obtaining when we draw n objects. Then the above expression becomes

$$\frac{\binom{a_1}{x_1}\binom{a_2}{x_2}}{\binom{N}{n}}$$

Now suppose we have N objects of k types, a_1 of type A_1, a_2 of type A_2, and so on through a_k of type A_k, and we are selecting n objects, without replacement.

Multivariate Hypergeometric Distribution

For an experiment as just described, let $X_1, X_2, X_3, ..., X_k$ be random variables that assign the number of objects of types $A_1, A_2, A_3, ..., A_k$ selected, respectively. Then the joint probability distribution for these random variables is

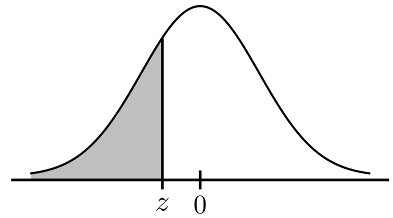
$$f(x_1, x_2, x_3, ..., x_k; a_1, a_2, a_3, ..., a_k, n, N) = \frac{\binom{a_1}{x_1}\binom{a_2}{x_2}\cdots\binom{a_k}{x_k}}{\binom{N}{n}}.$$

A Index of Symbols

$A \subseteq B$	A is a subset of B, 111
$A \nsubseteq B$	A is not a subset of B, 111
$A \cup B$	union of sets A and B, 112
$A \cap B$	intersection of sets A and B, 112
$A \times B$	Cartesian product of sets A and B, 114
$\mathcal{P}(A)$	power set of the set A, 8

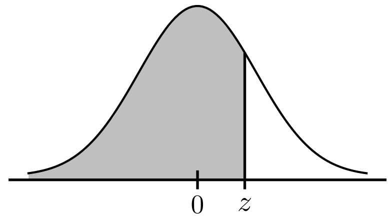
B Standard Normal Distribution Tables

$$Z = \frac{X - M}{SD}$$



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

$$Z = \frac{X - M}{SD}$$



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

C Sets

C.1 Introduction

We might "define" a **set** to be a collection of objects, but then one could ask what a collection is. Because we cannot really define a set without using other terms whose meanings are unclear, a set is what is called an *undefined term.* So we will assume that we all have some basic understanding of what a set is. We call the objects in a set **elements**. We also say that a set contains its elements. Here are two important conditions that sets *MUST* meet:

(1) No element can occur more than once in a set.

(2) There is no order to elements in a set. Thus the sets $\{1,2,3\}$ and $\{3,2,1\}$ are the same set. (However, it is common practice when listing elements of a set to list them from smallest to largest if they are numbers.)

Suppose that we have a set in mind. We have two immediate concerns, determining a symbol used to represent the set (its name), and describing the set in a way that it is clear to all what elements it contains. *We always use capital letters for names of sets.* (There may be occasional exceptions to this, but they will be rare.) A few sets that we will refer to regularly have permanently designated symbols. Perhaps the three most important sets in mathematics are the natural numbers, the integers and the real numbers. The **natural numbers** are the numbers 1,2,3,4,..., and we will denote them by $\mathbb{N}$. The **integers** are the natural numbers and their negatives, and zero. That is, they are the numbers $..., -3, -2, -1, 0, 1, 2, 3, ...$ We denote the integers by $\mathbb{Z}$.

The **real numbers** are difficult to describe without getting painfully technical, but basically they are all of the numbers you ever came across before you tried to take the square root of a negative number. They include all of the natural numbers, fractions, decimals, even-numbered roots of non-negative numbers, all odd-numbered roots, π, e, and of course negatives of any of these. The real numbers will be denoted by $\mathbb{R}$. In this course we will be primarily interested in the real numbers and the set called the **whole numbers**, which consist of the natural numbers and zero.

As stated before, the objects in a set are called elements of the set. We use the symbol $\in$ to indicate that something is an element of a particular set. For example, $5 \in \mathbb{N}$ means "five is an element of the natural numbers". (Often we will say instead "five is *in* $\mathbb{N}$".) Similarly, we could write $-\sqrt{3} \in \mathbb{R}$. At times we will also wish to indicate symbolically that something is *NOT* an element of a set. For this we use the symbol $\notin$; for example, $-7 \notin \mathbb{N}$.

NOTE: A major emphasis of this course will be the clear written and verbal communication of ideas, and it is extremely important that correct notation and terminology are used by all of us. Start now in making an effort to use correct notation!

C.2 Describing Sets

There are a variety of ways of describing a set in writing, but we will only need a few (at least for now) in this class. Let us consider two examples. Suppose that set A is all the natural numbers between 3 and 7, including both of those numbers, and set B is all the real numbers between 3 and 7, including 3 but NOT including 7. The first way we can describe a set is with a written verbal description, which is what I have just done for A and B. If a set is finite, we can describe it by simply listing the elements of the set. For example, we write

$$A = \{3, 4, 5, 6, 7\}.$$

The symbols $\{\ \}$ indicate that we are dealing with a set. We could also describe this set using something called **set builder notation**, which gives directions as to how to build a set:

$$A = \{n \in \mathbb{N} \mid 3 \leq n \leq 7\}.$$

This is read as "A is the set of natural numbers n such that n is greater than or equal to three and less than or equal to seven."

To represent set B, we can use set builder notation as well:

$$B = \{x \in \mathbb{R} \mid 3 \leq x < 7\}.$$

(We usually use the letters j, k, m and n for integers, and letters like x, y and z for real numbers.) We can also use something called **interval notation** to describe certain sets of real numbers. (Interval notation is *NEVER* appropriate for describing sets of natural numbers or integers.) The set of all real numbers between two given numbers, possibly including either or both of those numbers, is described in interval notation by

- listing the two numbers, smallest first then largest, separated by a comma,
- enclosing the pair of numbers with square brackets [ ] and/or parentheses ( ), using [or] if the number is to be included, (or) if it is not. Of course, as you probably already know, we can also have the combinations [ ) and ( ].

The interval notation for set B is then $[3, 7)$.

When using interval notation to describe a set of real numbers that is greater than some number, or less than some number, we will need to incorporate the symbols ∞ and $-\infty$. (Infinity and negative infinity.) The set $\{x \in \mathbb{R} \mid x \geq 5\}$ has interval notation $[5, \infty)$, and the set $\{x \in \mathbb{R} \mid x \leq 5\}$ has interval notation $(-\infty, 5]$. Note that ∞ and $-\infty$ are not numbers, and are therefore not included in the sets.

1. (a) Describe the natural numbers greater than 7 using set builder notation.
   (b) Describe the set of real numbers greater than 7 using interval notation.
   (c) Describe the set of natural numbers between 2 and 6, including both, by listing.
   (d) Describe the set of real numbers between 2 and 6, including both, by set builder notation *AND* interval notation.

C.3 Finite, Countable and Uncountable Sets

A set that contains finitely many elements will be called (surprise!) a **finite set**. (And of course, a set containing infinitely many elements is an **infinite set**.) A **one-to-one correspondence** between two sets A and B is a method or scheme for pairing each element A with *exactly one* element of B in such a way that every element of B has been paired with some element of A. For example, we can construct a one-to-one correspondence between

$$A = \{0, 1, \pi, e\} \qquad \text{and} \qquad B = \{5, 10, 15, 20\}$$

by pairing 0 with 15, 1 with 5, π with 10 and e with 20.

2. Give a one-to-one correspondence between the natural numbers and the set $\{2, 4, 6, 8, ...\}$ by giving a function f that takes an element of one set and gives out a corresponding element of the other.

The above exercise shows that it is quite possible to have a one-to-one correspondence between infinite sets. Any set for which there exists a one-to-one correspondence with a subset of the natural numbers (which of course could be the set of all natural numbers) is called a **countable set**. A set that is not finite or countable is called (surprise!) **uncountable**.

3. Show that the set of integers is countable by setting up a one-to-one correspondence between them and the natural numbers.

It is often acceptable to describe infinite sets by listing, *as long as they are countable sets.* For example, the odd natural numbers could be denoted by O and described by

$$O = \{1, 3, 5, 7, ...\}.$$

One problem with listing to describe an infinite set is that there is no reason to expect that a pattern shown will continue to hold. However, *we will use listing to describe an infinite set when we think it clearly describes the set in question and, in those cases, we will assume that any pattern we see in the list continues.* We can now tidy up a bit.

NOTES:

(1) Interval notation is only appropriate for uncountable sets of real numbers, and listing is only appropriate for finite or countable sets. Both types of sets can be described using set builder notation.

(2) An interval of the real numbers with endpoints a and b that are both finite numbers is called a **finite interval**; the interval $B = [3, 7)$ is such an interval. Any other interval, like $(-\infty, 2]$, is an **infinite interval**. (Some people refer to infinite intervals as **unbounded intervals**.) *Note that a finite interval is NOT a finite set.*

(3) When denoting an interval for which the two endpoints are numbers, *the smaller number is always listed first.* If the interval involves either (or both) of the symbols $-\infty$ or ∞, $-\infty$ is always written first and ∞ is always written second.

(4) The symbols $-\infty$ and ∞ do not represent numbers, so they cannot be included in sets of numbers. Therefore, $-\infty$ is *always* preceded by (and ∞ is always followed by).

C.4 The Universal Set and the Empty Set

Whenever we are working with sets, there is usually some large set "in the background" that the sets under consideration are subsets of. For example, the background set for the set $A = \{3, 4, 5, 6, 7\}$ that we have been working with is $\mathbb{N}$ and the background set for $B = [3, 7)$ is $\mathbb{R}$. The relevant background set is called the **universal set**. If there is no designated universal set, we just denote the universal set by U and assume it is some set containing all of the sets under consideration.

We also have a need for a set that has no elements in it, which we call **empty set**, or **null set**. The need for the empty set is analogous to the fact that when counting things, we need the number zero to indicate when there is nothing to count. The symbol for the empty set is $\emptyset$. (From here on, you need to take care to use the symbol 0 for zero and $\emptyset$ for the empty set!) Note that the symbol $\emptyset$ is taken to include the set brackets $\{$ and $\}$, so $\{\emptyset\}$ *is incorrect notation for the empty set.* (This notation does have meaning in certain situations, but it is rarely needed.)

C.5 Cardinality of a Set

The **cardinality** of a finite set A is the number of elements in the set. For example, the cardinality of $\{2, 4, 6, 8, 10\}$ is five, and the cardinality of the empty set is zero. We will say that infinite sets have infinite cardinality. (There is a bit more to that story, but we have other things to attend to ...) The cardinality of the null set is zero. We denote the cardinality of a set A by $|A|$.

C.6 Subsets

We are very interested in situations where one set is "contained" in another. If every element of a set A is also an element of another set B we say that A is a **subset** of B. Symbolically we denote A is a subset of B by $A \subseteq B$. Another way of stating this in words is that "A is contained in B". *A set A is NOT a subset of a set B if there is an element in A (one is enough!) that is not in B.* When we want to indicate symbolically that a set A is *NOT* a subset of a set B we write $A \nsubseteq B$. Thus if $A = \{2, 4, 6, 8\}$ and $B = \{4, 5, 6, 7, 8, \}$, we have $A \nsubseteq B$ because $2 \in A$ and $2 \notin B$.

4. In each case, determine whether C is a subset of D. If C is not a subset of D, write a brief statement (like the one just given above for A and B) explaining why it is not.

   (a) $C = \{1, 3\}, \quad D = \{1, 2, 3, 4\}$ (b) $C = \{1, 3, 5\}, \quad D = \{1, 2, 3, 4\}$

   (c) $C = \{1, 3\}, \quad D = [1, 3]$ (d) $C = \{1, 3\}, \quad D = (1, 3)$

   In (d) we determine from context that $(1, 3)$ is an interval, not an ordered pair!

   (e) $C = \{1, 2, 3\}, \quad D = \{1, 2, 3\}$ (f) $C = [1, 3], \quad D = \{1, 2, 3\}$

5. Can a set be a subset of itself? If so, must it be?

6. There is no reason that a set cannot contain just one element; such sets are sometimes referred to as "singletons", but you don't need to know this terminology. Two of the statements

   $$5 \in A\,, \qquad 5 \subseteq A\,, \qquad \{5\} \subseteq A$$

   use proper notation and one does not. Which does not, and why? This exercise is meant to point out the difference between an element of a set and a singleton set as a subset of a set.

7. Suppose that A is *any* set. What is the largest possible subset of A?

Exercises 4(e), 5 and 7 illustrate that *every set is a subset of itself*. Because of a rule of logic that I don't want to go into here, *the empty set is a subset of every set*. If A is a subset of B but is not equal to B itself, we say that A is a **proper subset** of B. We are not usually interested in making a big deal out of whether or not a subset is proper.

8. Give all subsets of the set $\{a, b\}$.

9. Finish the statement "If $A \subseteq B$, then ..." with something about the cardinalities of A and B. Use correct notation and remember that $|A|$ and $|B|$ are *numbers*, so whatever you say about them must make sense for numbers.

C.7 Set Operations

A binary operation is a process that takes two things of a certain kind and gives back one thing of the same kind. The simplest example of a binary operation is addition, which takes two numbers and returns a third number. (Note that even when we are adding more than two numbers, we still add two at a time.) In this section you will see some binary operations on sets. Note that each of these *is a method for combining two sets to get a SINGLE new set*. Given any two sets A and B, the **union** of them is the set of elements that are in either A or B, or both. The notation for the union of A and B is $A \cup B$. So symbolically,

$$A \cup B = \{x \mid x \in A \text{ or } x \in B\}.$$

As emphasized above, $A \cup B$ is a *single* set!

The set of all elements that are in *both* A and B is called the **intersection** of the two sets, denoted by $A \cap B$. That is,

$$A \cap B = \{x \mid x \in A \text{ and } x \in B\}.$$

If the sets A and B have no elements in common, then their intersection is the empty set; in that case we say that the two sets are **disjoint**.

10. Let $A = \{2, 4, 6, 8, 10\}$ and $B = \{3, 4, 5, 6\}$. Find $A \cap B$ and $A \cup B$.

Given the *single set* A with universal set U, the set of all elements of U not in A is called the **complement** of A. It is denoted variously by A', $\overline{A}$, or A^C. Complementation is a **unary operation** - it is an operation that needs only one "input" (assuming that the universal set is known). Given two sets A and B, the set of all elements of A that are not in B is called the **difference of A and B**; sometimes this is called the relative complement of B with respect to A. We will denote this set by $A - B$, and we will often say "A minus B" instead of "the difference of A and B." Symbolically we now have

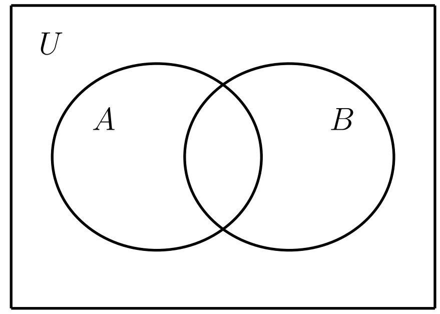
$$A' = \{x \mid x \notin A\} \qquad \text{and} \qquad A - B = \{x \mid x \in A \text{ and } x \notin B\}.$$

(Here it is understood in the definition of A' that the elements x must come from the universal set U.) We sometimes use $A \backslash B$ instead of $A - B$ for the difference of A and B. Of course, like when subtracting numbers, $A - B \neq B - A$ in general. (Can you think of a situation in which they would be the same?)

11. Let $U = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$, $A = \{2, 4, 6, 8, 10\}$, and $B = \{3, 4, 5, 6\}$.

    (a) Find A' and B', $A - B$.
    (b) Find $A' \cap B'$ and $(A \cap B)'$. Are they the same?
    (c) Give a set $C \subseteq U$ such that A and C are disjoint.

A good tool for visualizing set operations is a drawing called a **Venn diagram**. A Venn diagram usually consists of a rectangle representing the universal set and, inside the rectangle, circles or ovals representing other sets. If two sets intersect, then their ovals overlap, and if they are disjoint their ovals do not overlap. If one set is a subset of another, then its oval is inside the oval of the second. So the Venn diagram to the right represents the situation from the previous exercise. You might find Venn diagrams to be useful when working on the rest of the exercises in this section.



12. (a) It is possible to have $A \cap B = A$. A statement of the form "If ..., then ..." is called a **conditional statement**. Write a conditional statement giving the most general conditions under which $A \cap B = A$. "Most general" means the least specific conditions necessary.
    (b) Is it possible to have $A \cup B = A$? If so, give an example, and give the most general conditions under which it can happen *in the form of a conditional statement.*
    (c) Is it possible to have $A \cap B = A \cup B$? Again, give an example if it is possible and give the most general conditions under which it can happen.

13. (a) Draw graphical representations of $[-3, 2)$ and $[-1, \infty)$ on separate real (number) lines, one directly over the other. Then draw a graphical representation of the set $[-3, 2) \cap [-1, \infty)$ on another real line directly below those two. Give the interval notation for that set. Describe how to use the first two graphs to obtain the third.
    (b) Repeat the above for $[-3, 2) \cup [-1, \infty)$.

14. Repeat Exercises 10 and 11 for $U = \mathbb{R}, A = (-2, 4.1],\ B = [1, 8]$.

15. (a) Sketch the graph of the set $\{x \in \mathbb{R} \mid x \geq -2\} \cap \{x \in \mathbb{R} \mid x \neq 3\}$ on a number line.
    (b) *Carefully* write your answer to part (a) using interval notation. Even though you have found the intersection of two sets, your answer should be the union of two intervals!

16. Let $A = (3, 5]$ and $B = \{3, 4, 5, 6, 7\}$. Find $A \cup B$, $A \cap B$, $A - B$ and $B - A$.

17. Describe the set $\mathbb{R} - \{2\}$.

    (a) with interval notation (b) with set builder notation (c) in words

18. (a) Fill the first blank of the following statement with $<$, $>$, $\leq$, $\geq$ or $=$ and the second blank with $+$, $-$, $\cdot$ or $\div$:

    $$|A \cup B| \ \_\_\_\_ \ |A| \ \_\_\_\_ \ |B|$$

    Your answer should be valid whether or not A and B are disjoint, or regardless of whether one is a subset of the other.
    (b) You should not have used $=$ in the first blank of part (a). Under what circumstances would $=$ be correct there? Write a conditional statement that summarizes this.
    (c) Try to finish the statement $|A \cup B| =$ in such a way that it holds unconditionally. Your answer should contain the quantities $|A|$, $|B|$ and $|A \cap B|$.

C.8 Partitions of Sets

Mathematical concepts (as well as concepts in other technical disciplines) are often presented in a very terse form that is at first difficult to understand. Once such a concept is understood, we find the idea to be very simple and we wonder why its definition seems to be so confusing or unilluminating. That is because the definition is constructed to be as concise as possible, with no "loopholes". Historically, the idea came first, followed by a definition that may or may not have been correct. As time went on the definition was modified to be correct and as precise as possible. The process is usually reversed when the rest of us come to the idea later - the definition is given and we must sometimes struggle to see the idea that it describes. In the next exercise you will grapple with the definition of a concept that is really quite simple, a partition of a set. (Don't neglect to think about what the word partition means to you in your own field, or just in general!) Here is the definition:

A **partition** of a set A is a set $\{A_1,\ A_2, ...A_n\}$ of subsets of A for which

- none of $A_1, A_2, ..., A_n$ are empty,
- $A_1 \cup A_2 \cup \cdots \cup A_n = A$ and
- $A_i \cap A_j = \emptyset$ whenever $i \neq j$.

19. Determine whether each of the following is a partition of the set $A = \{1, 2, 3, 4\}$. For any that are not, explain why.

    (a) $\{\{1\},\ \{1, 2\},\ \{1, 2, 3\},\ \{1, 2, 3, 4\}\}$.
    (b) $\{\{1\},\ \{2\},\ \{3, 4\}\}$.
    (c) $\{\{1, 2, 3, 4\},\ \{1, 2\},\ \{3, 4\}, \emptyset\}$.
    (d) $\{\{1\},\ \{2\},\ \{3\}\}$.

20. What set or sets could be included in

$$\{[0, 1), [1, 3), (3, 4)\}$$

    to obtain a partition of the interval $[0, 6]$? If there is more than one answer possible, give two.

If you haven't recognized it by now, a partition of a set is just a set of its subsets that don't "overlap" (they're disjoint), and which "combine" to give the whole set (their union is the whole set). You probably think of a number of partitions already:

- The integers partition into evens and odds.
- The real numbers partition into positives, negatives and zero.
- The real numbers partition into rationals and irrationals.

C.9 Cartesian Products of Sets

Given two sets A and B we can create a new set consisting of all possible ordered pairs made up of an element of A and an element of B, *in that order.* This new set is called the **Cartesian product** of A and B, denoted by $A \times B$. Each ordered pair is enclosed in parentheses, with its two elements separated by a comma, like $(2, 5)$. Of course this notation also represents the interval of real numbers between 2 and 5, not including either. *One must determine from the context what the meaning of* $(2, 5)$ *is.*

21. If $A = \{a, b, c\}$ and $B = \{1, 2, 3, 4\}$, write out $A \times B$. Indicate symbolically that your answer is a *set.*

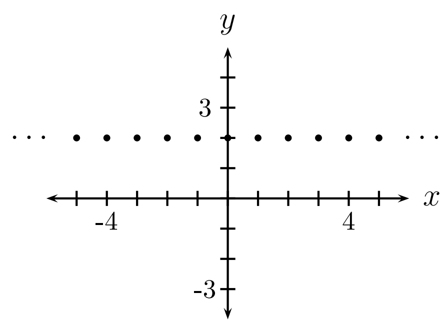
22. Discuss the validity of the statement $A \times B = B \times A$.

23. If A has m elements and B has n elements, how many elements does $A \times B$ have?

A Cartesian product that you are all familiar with is $\mathbb{R} \times \mathbb{R}$, the set of all ordered pairs of real numbers. This is xy-plane from algebra and calculus. When you graph an equation like $y = x^2$, you are really graphing a subset of $\mathbb{R} \times \mathbb{R}$; in particular, you are graphing the set

$$\{(x, y) \in \mathbb{R} \times \mathbb{R} \mid y = x^2\}.$$

Another subset of $\mathbb{R} \times \mathbb{R}$ is $\mathbb{Z} \times \{2\}$, whose graph is shown to the right.



24. Take $\mathbb{R} \times \mathbb{R}$ to be the universal set. For each of the following, sketch and label xy-axes, then indicate the given set in your diagram.

    (a) $\mathbb{Z} \times \mathbb{Z}$. This set is sometimes referred to as the lattice of integer points.

    (b) $\mathbb{R} \times \{0\}$. What is this set?

    (c) $\{3\} \times \mathbb{R}$

    (d) $[-1, 3] \times [2, 4]$

    (e) $(-1, 3] \times [2, 4)$ Note that you must somehow distinguish some sides of the resulting rectangle from others! Use solid lines when pairs on the boundary are included in the set, dashed when they are not. Put open or closed circles on the corners, indicating whether or not those points are included.

    (f) $[2, \infty) \times \mathbb{R}$

    (g) $\{(x, y) \in \mathbb{R} \times \mathbb{R} \mid x = -y\}$

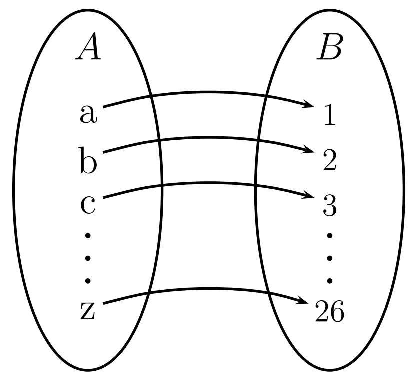
    (h) $\{(x, y) \in \mathbb{R} \times \mathbb{R} \mid x + y \in \mathbb{Z}\}$

25. Using the set definition for part (g) of the previous exercise as an example, describe the sets from parts (a), (c), (e) and (f) of that exercise using set builder notation.

D Functions

Those of you who have taken a physics class, or perhaps even a high school physical science class, know that light is a rather elusive concept. It has been found that for some purposes light can be thought as rays emanating from a source. This works well when thinking about things like reflection or refraction of light. However, when examining the fact that light is actually made up of parts (spectra, the wavelengths making up light), it is more productive to think of light as being made up of waves.

Similarly, we will find that there is more than one way to think of the concept of a function. In this chapter we will look at functions in the way that is probably familiar to all of you. That is, we think of a **function** as a rule for assigning to each element of a set A an element of a second set B. Let's look at a concrete example. Suppose that $A = \{a, b, c, d, ..., z\}$ and $B = \mathbb{R}$. A simple rule would be to assign to a the number 1, to b the number 2, and so on. The diagram to the right illustrates this.



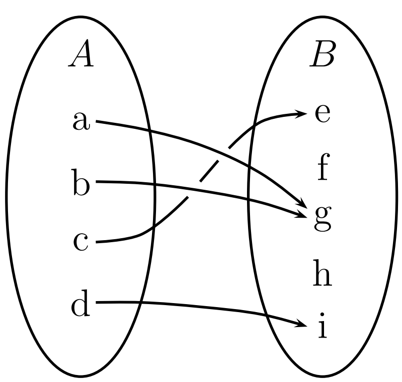
The function is *the rule AND the set* A. (We must specify those two - when the second set B is not specified, we can take it to be any set that contains *at least* all elements that the function assigns to elements of A, maybe more.) There are two conditions that must be met here for our rule to be a function:

(i) The rule must assign something in B to *every* element of A.

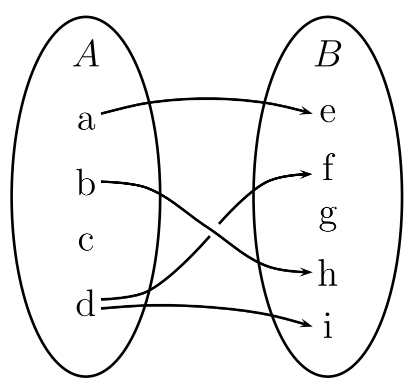
(ii) The rule cannot assign to an element of A more than one element in B.

1. Below are diagrams showing ways to assign to each element of $A = \{a, b, c, d\}$ an element of a set B whose elements are shown in each case. (Assume that the elements shown are *all* the elements of that set B.) For each diagram, determine whether the assignment rule shown is a function. That is, does it satisfy (i) and (ii) above? If not, tell which is (are) violated, and how.

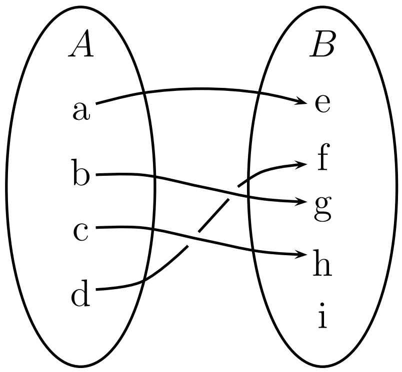
(a)



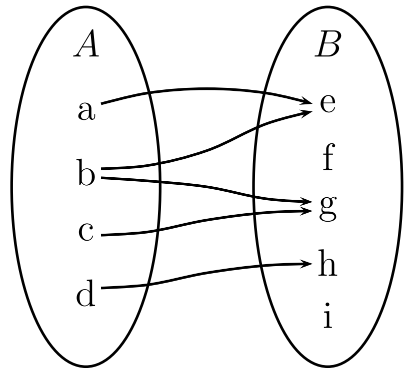
(b)



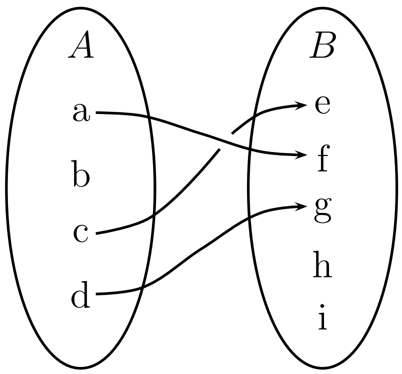
(c)



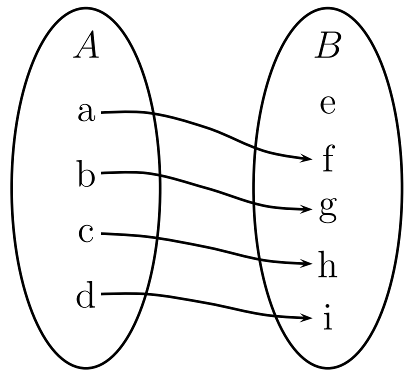
(d)



(e)



(f)



Notice here that we have to be careful not to read into (i) and (ii) anything that is not there. As you should now see, it is possible that

- different elements of A can be assigned the same element of B (Exercise 1(a)), and
- not every element of B has to be assigned to an element of A (Exercises 1(a),(f)).

We will always name our functions, usually (but not necessarily always) with the lower case letters f, g and h. Suppose that the function described before Exercise 1 is f. Then we indicate that f assigns 3 to c by writing $f(c) = 3$. If we use a different rule, we use different letter. For example we could again let $A = \{a, b, c, d, ..., z\}$ and $B = \mathbb{R}$, and assign to each vowel the number 0 and to each consonant the number 1. (Remember vowels and consonants?) This is a function because every letter will be assigned a value (every letter is either a vowel or a consonant) and no letter can be assigned more than one value (no letters are both vowels *and* consonants). If we call this function g we then have $g(m) = 1$ and $g(u) = 0$. (Of course there could be a little confusion concerning $g(g) = 1$, but we will rarely, if ever, see such a problem again.) Notice that f and g both assign real numbers to letters of the alphabet, but the assignment rules used are different.

Now for a little terminology. The set A is called the **domain** of the function. We will call the B the **target set** of the function. The term "domain" is used universally, but different people have different names for the set B. (I looked in three different books, and each author had a different name for this set. I like the name "target set" because it seems natural when we think of the function as sort of "shooting things from A into B", as illustrated in the diagrams we have been looking at.) Using this language, (i) and (ii) say that every element in the domain must be assigned an element in the target set, and a domain element can only be assigned one element of the target set.

On the other hand, not every element in the target set needs to be "hit" by the function when assigning values to the elements of the domain. The subset of the target set consisting of all elements that *are* assigned to elements of the domain is called the **range** of the function.

For the sake of brevity, we have some notation for the domain and range of a function; the domain of a function f is denoted by $\text{Dom}(f)$ and the range is denoted by $\text{Ran}(f)$. When referring to arbitrary elements of the domain and range of a function we generally use x for domain elements and y for range elements. Thus the domain of a function might be $\{x \in \mathbb{R} \mid \text{blah, blah}\}$ and the range might be $\{y \in \mathbb{R} \mid \text{yada, yada}\}$. As an example, consider the function $h(x) = x^2 + 3$, where x is only allowed to have *integer* values. Then

$$\text{Dom}(h) = \mathbb{Z} \qquad \text{and} \qquad \text{Ran}(h) = \{3, 4, 7, 12, 19, ...\}$$

2. Consider the two functions f and g described previously. The domain of both is the letters of the alphabet, and the target set for both is the real numbers. Give the range of each, using listing or set builder notation, whichever seems most appropriate.

Prior to this course, you have likely thought of a function as being just a rule telling how the function assigns values, like $f(x) = x^2$. *Whenever we are given a function as just a rule, we will assume that the domain of the function is the largest subset of the real numbers for which the function is defined, and the target set is all of the real numbers.* The range may or may not be all real numbers.

3. Give the domain and range for each of the following functions *using interval notation* and the Dom and Ran notations.

   (a) $f(x) = -x^2$

   (b) $g(x) = 2x - 1$

   (c) $h(x) = \sqrt{5 - x}$

   (d) $f(x) = \dfrac{1}{x + 3}$

4. Restate the domain and range for each of the functions in Exercise 3, using set builder notation. (Just write $\mathbb{R}$ in the case that the set being described is all real numbers. Note that the set consisting of all real numbers except 2 can be clearly and concisely described by $\{x \in \mathbb{R} \mid x \neq 2\}$.

Suppose that we are considering the function f, whose domain is all real numbers, that assigns to each real number its square. We will sometimes show the domain and target set of the function by writing $f : \mathbb{R} \to \mathbb{R}$. So we might write something like

$$\text{Define the function } f : \mathbb{R} \to \mathbb{R} \text{ by } f(x) = x^2.$$

When the same rule is used with different domain sets, *the resulting functions are not the same*. This is because *a function consists of both a rule and a domain.* For example, the function defined by $g : \mathbb{N} \to \mathbb{R}$, $g(x) = x^2$ is not the same as the function f that was just defined. When we use a rule on a set smaller than the largest set that it is mathematically feasible on, we say that we have **restricted** the domain of the function.

5. Restricting the domain of a function will sometimes change the range of the function.

   (a) What are the ranges of the functions f and g just given?

(b) Is it possible to restrict the domain of f without changing the range? (Of course you should give an example if you answer yes!)

6. Suppose that the domain of each of the functions in Exercise 3 is restricted to the largest subset of $\mathbb{Z}$ that is allowable. Give the domain and range of each function, using any correct notation.

The letter x used in the definition $g(x) = x^2$ is what is called a "dummy variable"; it is simply a letter used for the purpose of showing how the function g assigns values to domain elements. We could just as well use any other letter for the dummy variable. We will usually use x for the dummy variable when the domain is some continuous subset of the real numbers, and n when the domain is $\mathbb{N}$ or $\mathbb{Z}$. Here we would use the dummy variable x with f, and the variable n with g.

We should note at this point that neither the domain nor the target set of a function need be sets of numbers. For example, we could consider the function that assigns to every triangle the perimeter of the triangle. For the purpose of these notes, *after the following exercise we will assume that the target set of every function is some set of numbers.* This will keep us from having to belabor certain points that are not essential to your understanding of functions. The domains of our functions will generally be sets of numbers as well.

7. Describe some more functions whose domains are all triangles. There are at least six or seven that come to mind for me right now!

8. Try to determine another function whose domain is not a set of numbers - the target set may or may not be a set of numbers. Be sure to specify the domain and the rule for assigning a value to each element of the domain. Be creative - try not to just give some minor variation on the example that I just gave!

9. Give a function whose domain *and* target set are not sets of numbers.

E Review of Calculus

E.1 Differentiation

Differentiation refers to the act of taking a derivative. First we note that there are really two kinds of derivatives for a function f of a variable x:

- The derivative of f at a point $x = a$, denoted by $f'(a)$, is supposed to represent the *instantaneous rate of change* of the function *at that point* alone. Graphically it is the slope of the line tangent to the graph of f at the point on the graph with x-coordinate a.
- The derivative *function* of f, denoted by f' is the function whose value at any point $x = a$ is the derivative of f at that point.

Based on these things, two things are immediately apparent:

- The derivative of a constant function $f(x) = C$ at *any* point a is zero. The derivative function is then $f'(x) = 0$.
- The derivative of a linear function $f(x) = mx + b$ at any point is just the slope of the line, m. The derivative function is $f'(x) = m$, a constant function.

For our purposes, we can pretty much get by with the following facts about derivatives.

Let f and g be differentiable functions and let c be any constant.

(1) $f'(cx) = cf'(x)$ (2) $[f(x) \pm g(x)]' = f'(x) \pm g'(x)$

(1) If $f(x) = x^n$, then $f'(x) = nx^{n-1}$.

(2) If $f(x) = e^{g(x)}$, then $f'(x) = g'(x)e^{g(x)}$

In the last of these, if $g(x) = x$ we then have that the e^x is its own derivative, which makes the exponential function a very important function in mathematics. The "rule" given above incorporates the **chain rule**, which you should recall.

We can combine the first three things above, along with the fact that the derivative of a constant is zero, to see that

$$\text{if} \qquad f(x) = 5x^3 - 7x^2 + 3x - 1, \qquad \text{then} \qquad f'(x) = 15x^2 - 14x + 3$$

Using the rule for the derivative of an exponential function we have

$$\text{if} \qquad g(x) = e^{-5x}, \qquad \text{then} \qquad g'(x) = -5e^{-5x}$$

and

$$\text{if} \qquad h(x) = e^{x^2}, \qquad \text{then} \qquad h'(x) = 2xe^{x^2}$$

E.2 Basics of Integration

Recall that there are two types of integrals:

- **Indefinite Integrals (Antiderivatives):** An example would be $\int (x^2 - 3x^2 + 5)\,dx$. The symbols $\int$ and dx around an expression mean to find an expression whose derivative is the original expression. In this case we have

$$\int (x^2 - 3x^2 + 5)\,dx = \frac{1}{3}x^3 - \frac{3}{2}x^2 + 5x + C,$$

where C represents any constant. You should verify for yourself that the derivative of $\frac{1}{3}x^3 - \frac{3}{2}x^2 + 5x + C$ is in fact $x^2 - 3x + 5$.

- **Definite Integrals:** An example of this would be $\int_0^3 (x^2 - 3x^2 + 5)\,dx$, which represents a *number*, rather than another expression, like the indefinite integral does. To determine the number, we apply the Fundamental Theorem of Calculus:

> If f is continuous on the interval $[a, b]$ and F is such that $F'(x) = f(x)$ for all $x \in [a, b]$, then
>
> $$\int_a^b f(x)\,dx = F(b) - F(a).$$

We use the notation $[F(x)]_a^b$ to mean $F(b) - F(a)$, so for our example we have

$$\begin{aligned}
\int (x^2 - 3x^2 + 5)\,dx &= \left[\frac{1}{3}x^3 - \frac{3}{2}x^2 + 5x\right]_0^3 \\
&= \left[\frac{1}{3}(3)^3 - \frac{3}{2}(3)^2 + 5(3)\right] - \left[\frac{1}{3}(0)^3 - \frac{3}{2}(0)^2 + 5(0)\right] \\
&= 9 - \frac{27}{2} + 15 \;=\; \frac{21}{2}
\end{aligned}$$

We are only interested in definite integrals in this course, but note that to evaluate the definite integral of a function f we need to find an anti-derivative F. The three main rules we need for integrating the functions we'll be interested in are

> $$\int u^n\,du = \frac{1}{n+1}u^{n+1} \text{ if } n \neq -1 \qquad \int \frac{1}{u}\,du = \ln|u| \qquad \int e^u\,du = e^u$$

F Series

F.1 Sequences and Series

For our purposes, a **sequence** is simply an infinite list of numbers or functions. Each number or function is called a **term** of the sequence. We stipulate that there is a definite order in which the terms are listed, and they usually (but don't have to) follow some sort of pattern. Here are some examples of sequences:

Example 1: 1, 2, 3, 4, ...	Example 2: $1, -1,\ 1, -1,\ 1, -1, ...$
Example 3: $\frac{5}{3},\ \frac{5}{9},\ \frac{5}{27}.\ \frac{5}{81}, ...$	Example 4: 1, 1, 2, 6, 24, 120, ...
Example 5: $\frac{1}{2},\ \frac{1}{4},\ \frac{1}{8},\ \frac{1}{16},\ \frac{1}{32}, ...$	Example 6: $1,\ x,\ \frac{1}{2}x^2,\ \frac{1}{6}x^3,\ \frac{1}{24}x^4, ...$

Two broad categories of sequences that we will deal with are numerical sequences, like Examples 1-5, and sequences of functions, like Example 6. A sequence is called a **geometric sequence** if each term is obtained by *multiplying* the previous term by a fixed amount, called the **ratio**. The ratio is denoted by r.

1. Determine which of the above sequences are geometric, and give the ratio for each.

Sequences arise naturally in combinatorial situations, as well as probability scenarios.

2. Consider the following problem: 23 people are at a party, and each must talk one-on-one to every other person at the part *exactly once*. How many conversations will there have been by the time the party is (finally) over? One way to tackle this problem is to find out how many conversations there would be if there were only one at a party, then how many conversations if there were two people, then three, and so on. One would then hope to see a pattern in the sequence of numbers of conversations for 1, 2, 3, ... people.

   (a) Obtain the sequence described. The numbers are called the **triangular numbers**. Can you see why?

   (b) If you know how, construct Pascal's Triangle. What do you notice?

   (c) Can you figure out how many conversations there will be with 23 people *without writing out every term of the sequence*? This is a somewhat challenging problem if you have not done this sort of thing before. Give it a try!

Series

If we add all of the terms of a sequence together, we get something called a **series**. Here is an example of a series:

$$1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \cdots$$

This particular series is important enough that it gets its own name; it is the **harmonic series**. Some of the terminology from sequences carries over to series; there are numerical series and series of functions. For a scientist or engineer, the most important series are those whose terms are functions. Geometric series are series whose terms are a geometric sequence.

A series can have terms containing a variable, and a series whose terms are monomials in successive powers of x is called a **power series**. Here is an example:

$$\frac{2}{3}+\frac{4}{9}x+\frac{8}{27}x^2+\frac{16}{81}x^3+\cdots$$

F.2 Convergence of Series

There are several ways that we can get a sequence from a given series. Obviously, since a series is just a sum of a sequence, the terms of a series form a sequence. Also, when we have a power series we can form a sequence by taking just the coefficients of the powers of x, remembering that the constant term can be thought of as a constant times x^0. (The sign needs to go with the constant in this case.)

Perhaps the most important method of obtaining a sequence from a series is to find what are called **partial sums** of the series. The first partial sum is simply the first term of the series. The second partial sum is the sum of the first and second terms, the third partial sum is the sum of the sum of the first three terms of the series, and so on. If we then list the partial sums in order, we have a sequence! For example, the sequence of partial sums for the series

$$1+2+3+4+\cdots \qquad \text{is} \qquad 1,3,6,10,15,...$$

Hmmmmm... that sequence looks familiar!

3. Find the sequence of partial sums of the following series. *Remember that the first partial sum is just the first term.*

   (a) $1+1+2+3+5+8+13+\cdots$

   (b) $1-1+1-1+1-1+\cdots$

   (c) $1+\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\frac{1}{16}+\cdots$

   (d) $1+3+5+7+9+\cdots$

4. A well known series is

$$\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\frac{1}{16}+\cdots$$

   The denominators of the terms of this series are 2, 4, 8, 16, 32, ...

   (a) The first term of the series is $\frac{1}{2}$, the second term is $\frac{1}{4}$, and so on. What is the nth term of the series, which is denoted by a_n? What is the limit of a_n as $n\to\infty$? State your answer by writing $\lim_{n\to\infty} a_n =$ ______.

   (b) Find the first three partial sums of the series, which we denote by s_1, s_2 and s_3. Speculate as to what s_4 is, then check your answer by adding the first four terms.

   (c) Find the nth partial sum of the series, s_n. What is $\lim_{n\to\infty} s_n$?

   (d) Call your answer to (c) L, and note that $s_n < L$ for all n. But since the limit of the partial sums is L, their values are getting closer and closer to L as n gets larger and larger. How close is s_n to L?

   (e) Find out how large n has to be in order for s_n to be within 0.0001 of L.

The value L that you obtained in (c) is called the **limit of the series**, and we say that the series **converges to L.** If no limit L exists, then we say that the series **diverges**.

F.3 Convergence of Geometric Series

In the previous section you determined whether or not a geometric series converged by computing its partial sums and seeing if they had a finite limit. Of course, some geometric series converge and others do not. Consider, for example, the two series

$$1 + 2 + 4 + 8 + 16 + \cdots\ , \qquad\qquad 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots$$

Clearly the limit of the partial sums of the first series will be ∞, so that series diverges. In general, determining whether a series converges can be a tricky business. In the case of geometric series, however, there is a simple way to determine whether the series converges and, if it does, we have a nice formula for the limit:

> **Theorem:** The geometric series $a + ar + ar^2 + ar^3 + \cdots$ converges if $|r| < 1$ and diverges if $|r| \geq 1$. If it converges, it converges to
>
> $$S = \frac{a}{1-r}$$

5. Use the above result to tell why the series $1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots$ converges, and find its sum.

6. Repeat for the series $\frac{1}{3} + \frac{1}{9} + \frac{1}{27} + \frac{1}{81} + \cdots$

G The Gamma Function

Given parameters $\alpha > 0$ and $\beta > 0$, a very useful two-parameter family of continuous probability distributions are those of the form

$$f(x) = \begin{cases} 0 & \text{if} \quad x < 0\,, \\ k\,x^{\alpha-1}\,e^{-\frac{x}{\beta}} & \text{if} \quad x \geq 0 \end{cases} \tag{1}$$

where k is some constant to be determined.

1. Suppose that the value of the parameter α is fixed at one. Do a u-substitution to determine the value of the constant k in this case. What distribution does (1) then represent?

> **The Gamma Function**
>
> For all $\alpha > 0$ we define the **gamma function** Γ by
>
> $$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}\,e^{-x}\,dx\,.$$

2. (a) Show that $\Gamma(1) = 1$.
   (b) Use integration by parts to show that $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$.
   (c) Use the results of (c) and (a) to find $\Gamma(2)$. *Indicate clearly what happens here.*
   (d) Use the results of (c) and (d) to find $\Gamma(3)$. *Again, indicate clearly what happens.*
   (e) Repeat a similar process to find $\Gamma(4)$ and $\Gamma(5)$.

This exercise indicates the following.

> For $\alpha = 1, 2, 3, ...,\ \ \Gamma(\alpha) = (\alpha - 1)!$

3. (a) Make the the substitution $x = 2y^2$ into the expression for $\Gamma(\frac{1}{2})$. Then use the result of Exercise 1(a) from Section 4.2 to show that $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.
   (b) Use $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$ and the result of part (a) to get the exact value of $\Gamma(\frac{7}{2})$. (**Hint:** Don't multiply fractions out.) Speculate on the exact value of $\Gamma(\frac{27}{2})$.

H Solutions to Exercises

H.1 Chapter 1 Solutions

Section 1.1

1. (a) {H, T}, {H}
   (b) {TTT, HTT, THT, TTH, HHT, HTH, THH, HHH}, {HHT, HTH, THH, HHH}
   (c) {H, TH, TTH, TTTH,...}, {H, TH, TTH, TTTH}
   (d) {ABC, ACB, BAC, BCA, CAB, CBA, BAC, BCA}, {BAC,BCA}

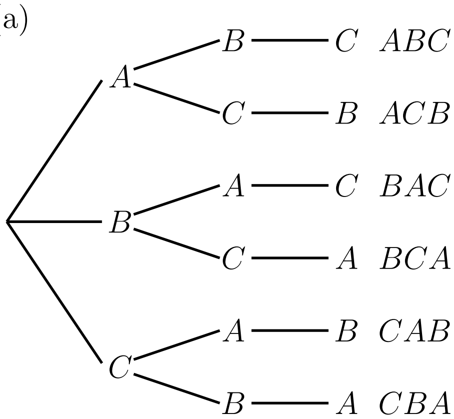
2. (a) {T1, T2}
   (b) {T1, T2, T3, T4, T5, T6, H1, H2}
   (c) {H4, H5, H6, T4, T5, T6}

3. (a) There are 36 outcomes possible, because there are six possible outcomes on one die and, for each of those outcomes, another six on the second die.
   (b) $\{(2,6),(3,5),(4,4),(5,3),6,2)\}$

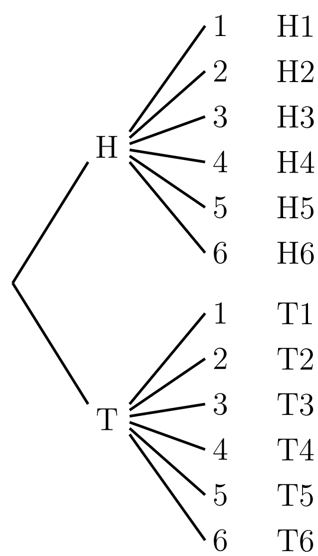
4. $(10.12, \infty)$ or $\{x \in \mathbb{R} \mid x > 10.12\}$

5. $[10, 15]$ or $\{t \in \mathbb{R} \mid 10 \leq t \leq 15\}$

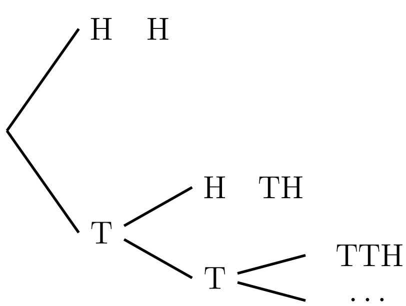
6. (a)



   (b)



   (c)



7. There are 24 ways: TChJCo, TChCoJ, TJChCo, TJCoCh, TCoJCh, TCoChJ, JTChCo, JTCoCh, JChTCo, JChCoT, JCoTCh, JCoChT, ChTJCo, ChTCoJ, ChJTCo, ChJ-CoT, ChCoTJ,ChCoJT, CoTChJ, CoTJCh, CoChTJ, CoChJT, CoJTCh, CoJChT

8. (a) The events are not mutually exclusive; both events contain HHT, HTH and THH.
   (b) The events are mutually exclusive.
   (c) The events are mutually exclusive.

9. (a) {AB, AC, AD, BC, BD, CD}
   (b) {AA, AB, AC, AD, BB, BC, BD, CC, CD, DD}

10. (a) {AB, AC, AD} (b) {AA, BB, CC, DD}
    (c) {AB, AC, AD, BC, BD, CD}

11. {(AB,CDE), (AC,BDE), (AD,BCE), (AE,BCD), (BC, ADE), (BD,ACE), (BE,ACD), (CD,ABE), (CE,ABD), (DE,ABC)}

12. {(AB, CD, E), (AB, CE, D), (AB, DE, C), (AC, BD, E), (AC, BE, D), (AC, DE, B), (AD, BC, E), (AD, BE, C), (AD, CE, B), (AE, BC, D), (AE, BD, C), (AE, CD, B), (BC, DE, A), (BD, CE, A), (BE, CD, A)}

Section 1.2

1. $\frac{4}{8}$

2. (a) 36 (c) $\frac{3}{36}$ (d) $\frac{3}{36} + \frac{2}{36} + \frac{1}{36} = \frac{6}{36}$
   (e) The sum of the numbers is five or more. The probability of this is $1 - \frac{6}{36} = \frac{30}{36}$.

Section 1.3

1. (a) Mr. W can dress six ways, BB, BG, BR, KB, KG, KR (b) $\frac{1}{6}$

2. (a) $26 \cdot 26 \cdot 10 \cdot 10 \cdot 9 = 608,400$ (b) $1 \cdot 1 \cdot 10 \cdot 10 \cdot 1 = 100$ (c) $\dfrac{52000}{608400}$

3. (a) 2 (b) $3 \cdot 2 \cdot 1 = 6$ (c) $4 \cdot 3 \cdot 2 \cdot 1 = 24$ (d) $n(n-1)(n-2) \cdots 2 \cdot 1$

4. (a) $4 \cdot 3 = 12$ (b) $8 \cdot 7 \cdot 6 = 336$ (c) $18 \cdot 17 \cdot 16 \cdot 15 \cdot 14 = 1,028,160$

5. (a) 6 (b) 56

6. $\dfrac{16!}{5!\,2!\,9!} = 240,240$ 7. $\dfrac{8!}{3!\,5!} = 56$ 8. $\dfrac{5!}{2!\,3!} = 10$ 9. $\dfrac{5!}{2!\,2!\,1!} = 30$

Section 1.4

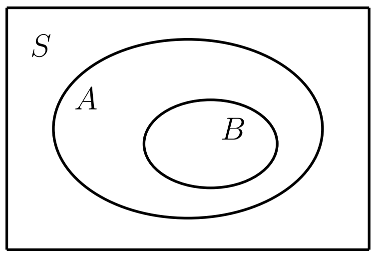
1. (a) $P(F) = \frac{12}{52} = \frac{3}{13}$ (b) $P(E) = \frac{20}{52} = \frac{5}{13}$ (c) $P(F \cup E) = \frac{32}{52} = \frac{8}{13}$
   (d) $P(A \cup B) = P(A) + P(B)$

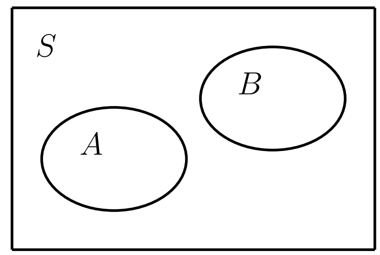
2. (a) C and F are not mutually exclusive because the Jack, Queen and King of clubs are in both events.
   (b) $P(C) = \frac{13}{52}$, $P(F) = \frac{12}{52} = \frac{1}{4}$ (c) $P(A \cup C) = \frac{22}{52} = \frac{11}{26}$

3. (a) $P(\text{male or non-smoker}) = P(\text{male}) + P(\text{smoker}) - P(\text{male and smoker})$
$$= \frac{10}{27} + \frac{20}{27} - \frac{8}{27} = \frac{22}{27}$$

4. (a) $P(M' \cap N) = 0.444$ (b) $P(M') = 0.186 + 0.444 = 0.630$
   (c) $0.074 + 0.444 = 0.518$

5. See Venn diagram below and to the left. Since $B \subset A$, $A \cap B = B$, which then gives $P(B) - P(A \cap B) = 0$.





6. See Venn diagram above and to the right. Since $A \cap B = \emptyset$, $P(A \cap B) = 0$.

7. (a) $P(A \cup B) = 0.17 + 0.35 = 0.52$ (b) $P(A \cap B) = 0.42 - 0.35 = 0.07$
   (c) $P[(A \cup B)'] = 1 - 0.52 = 0.48$

8. (a) $P(D') = 1 - 0.14 = 0.86$ (b) $P(D \cup B) = 0.14 + 0.227 - 0.098 = 0.269$
   (c) $P(D' \cap B') = P[(D \cup B)'] = 1 - 0.269 = 0.731$ Note the use of DeMorgan's Law!

Section 1.5

1. (a) $\frac{7}{27}$ (b) $\frac{2}{10}$

2. (a) $\frac{1}{4}$ (b) $\frac{1}{3}$ (c) 1

3. (a) $\frac{5}{17}$ (b) $\frac{8}{20}$ (c) $\frac{8}{10}$

4. (a) $\frac{0.098}{0.14} = 0.70$ (b) $\frac{0.129}{0.227} = 0.568$

5. $P(A|B) = \frac{0.07}{0.17} = 0.41$ $P(B|A) = \frac{0.07}{0.42} = 0.17$

6. (a) $\frac{4}{52} = \frac{1}{13}$ (b) $\frac{4}{12}$ (c) $\frac{1}{13}$

7. (a) $P(H \cap A) = \frac{1}{12}$, $P(H) = \frac{6}{12} = \frac{1}{2}$, $P(A) = 212 = \frac{1}{6}$ The events are independent because $P(H \cap A) = P(H)P(A)$.
   (b) $P(A) = \frac{1}{6}$, $P(B) = \frac{6}{12} = \frac{1}{2}$, $P(A \cap B) = \frac{1}{6}$ The events are not independent.

8. The events are not independent.

Section 1.6

1. (a) The events should not be independent. If we get a yellow on the first selection, then we are less likely to get a yellow on the second. $P(A) = \frac{3}{5}$, $P(B) = \frac{12}{20}$, $P(A \cap B) = \frac{6}{20}$. Since $P(A \cap B) \neq P(A)P(B)$, the events are not independent.
   (b) $P(A) = \frac{3}{5}$, $P(B|A) = \frac{2}{4}$

(c) $P(B|A) = \dfrac{P(B \cap A)}{P(A)} \Rightarrow \dfrac{2}{4} = \dfrac{P(B \cap A)}{\frac{3}{5}} \Rightarrow P(B \cap A) = \dfrac{2}{4} \cdot \dfrac{3}{5} = \dfrac{6}{20}$

2. $\frac{17}{29} \cdot \frac{16}{28} = \frac{68}{203}$

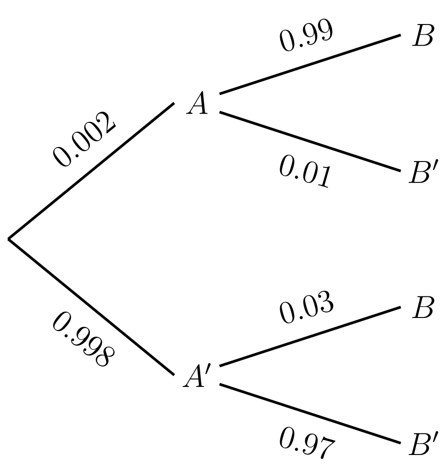
Section 1.7

1. (a) $P(A) = \frac{2}{1000} = 0.002, \; P(B|A) = 0.99, \; P(B|A') = 0.03$

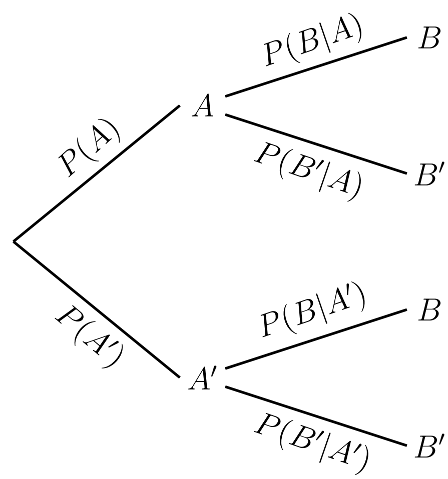
   (b) $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$ (c) See below.

   (d) $P(B) = (0.002)(0.99) + (0.998)(0.03) = 0.03192$

   (e) $P(A|B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{(0.002)(0.99)}{0.03192} = 0.062 = 6.2\%$



Exercise 1(c)



Exercise 2(a)

2. (a) See above. (b) $P(B) = P(A)P(B|A) + P(A')P(B|A')$

   (c) $P(B|A) = \dfrac{P(A \cap B)}{P(A)} \Rightarrow P(A \cap B) = P(A)P(B|A)$

   (d) $P(A|B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{P(A)P(B|A)}{P(A)P(B|A) + P(A')P(B|A')}$

   (e) $P(A|B) = \dfrac{(0.002)(0.99)}{(0.002)(0.99) + (0.998)(0.03)} = 0.062$

3. **Experiment:** Select a random e-mail message sent to a person.

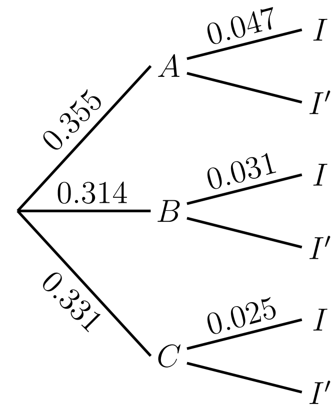
   Event A**:** The message is spam. **Event** B**:** The message is removed.

   $P(A) = 0.73, \; P(B|A) = 0.97, \; P(B|A') = 0.05$

   $P(A|B) = \dfrac{(0.73)(0.97)}{(0.73)(0.97) + (0.27)(0.05)} = 0.981$

4. (a) $P(I) = (0.355)(0.047) + (0.314)(0.031) + (0.331)(0.025) = 0.034694$

See tree diagram below and to the right.

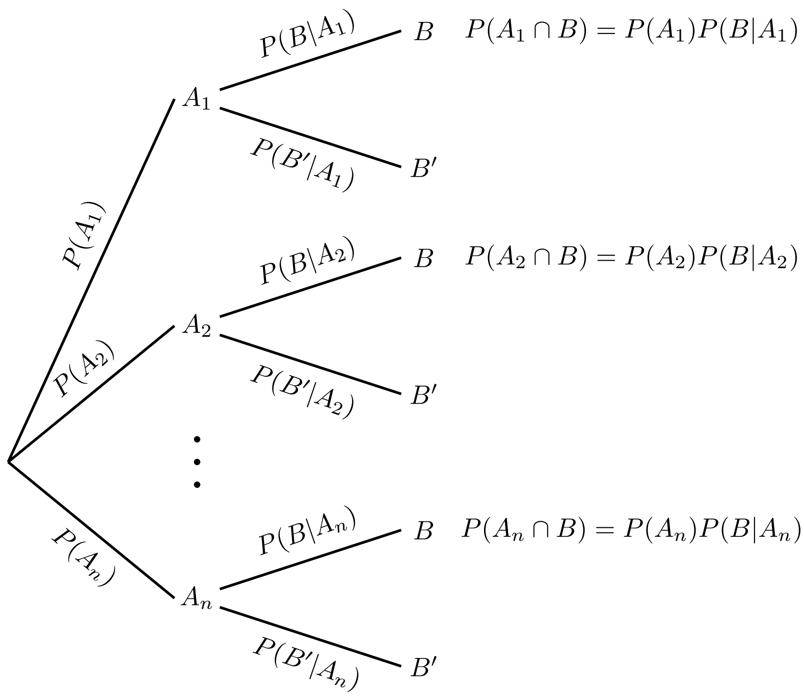


(b) $P(A) = 0.355,\ P(B) = 0.314,\ P(C) = 0.331$

$P(I|A) = 0.047,\ P(I|B) = 0.031,\ P(I|C) = 0.025$

(c) $P(I) = P(A)P(I|A) + P(B)P(I|B) + P(C)P(I|C)$

5.



6. $P(C|I) = \dfrac{P(C \cap I)}{P(I)} = \dfrac{(0.331)(0.025)}{0.034694}$

H.2 Chapter 2 Solutions

Section 2.1

1. (a) $\{0, 1, 2, 3, 4, 5, 6\}$ (b) $\{1, 2, 3, ...\}$ (c) $(0, \infty)$
2. $X((2, 5)) = 2$
3. (a) $Y(2, 5) = 7$ (b) $(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)$
   (c) $(6, 6)$ (d) $(1, 2), (2, 1)$
4. (a) $X(\text{TTTTTH}) = 6$
   (b) No, each value of the random variable results from only one outcome.
5. (a) $1 \le X \le 2$ (b) $X = 0$
6. (a) $X \le 5$ or $X < 6$ (b) $X = 3$
7. See table to the right. Every value of the random variable does not occur the same number of times.

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

8. (a) $P(Y = 4) = \frac{3}{36} = \frac{1}{12}$ (b) $P(Y = 7) = \frac{6}{36} = \frac{1}{6}$ (c) $P(Y \le 4) = \frac{6}{36} = \frac{1}{6}$
   (d) $P(Y > 4) = 1 - P(Y \le 4) = 1 - \frac{1}{6} = \frac{5}{6}$ (e) $P(Y = 15) = 0$
   (f) $P(Y \ne 10) = 1 - P(Y = 10) = 1 - \frac{3}{36} = \frac{33}{36} = \frac{11}{12}$
9. (a) $\text{Ran}(X) = \{0, 1, 2, 3\}$ (b) $P(X = 3) = (\frac{3}{10})^3(\frac{7}{10}) \cdot 4 = \frac{756}{10000} = \frac{189}{2500}$
   (c) $P(X \ge 1) = 1 - P(X = 0) = 1 - (\frac{7}{10})^4 = 1 - \frac{2401}{10000} = \frac{7599}{10000}$
10. (a) $\text{Ran}(X) = \{0, 1, 2, 3\}$ (b) $P(X = 3) = \frac{3 \cdot 2 \cdot 1 \cdot 7}{10 \cdot 9 \cdot 8 \cdot 7} \cdot 4 = \frac{168}{5040} = \frac{7}{210}$
    $P(X \ge 1) = 1 - P(X = 0) = 1 - \frac{7 \cdot 6 \cdot 5 \cdot 4}{10 \cdot 9 \cdot 8 \cdot 7} = 1 - \frac{1}{6} = \frac{5}{6}$

Section 2.2

1. (a) (a) $F(1) = \frac{1}{4} + \frac{1}{2} = \frac{3}{4}$ (b) $F(.83) = \frac{1}{4}$ (c) $F(1.99999) = \frac{1}{4} + \frac{1}{2} = \frac{3}{4}$
   (d) $F(2) = \frac{1}{4} + \frac{1}{2} + \frac{1}{4} = 1$ (e) $F(73) = \frac{1}{4} + \frac{1}{2} + \frac{1}{4} = 1$
2. (a) $X(\text{THTH}) = 2$ (b) $X(\text{HHHH}) = 4$
   (c) $0, 1, 2, 3, 4$ (d) $f(3) = P(X = 3) = \frac{4}{16} = \frac{1}{4}$
   (e) $F(3) = P(X \le 3) = 1 - P(X = 4) = 1 - \frac{1}{16} = \frac{15}{16}$
   (f) $P(X \le 2) = f(0) + f(1) + f(2) = F(2) = \frac{1}{16} + \frac{4}{16} + \frac{6}{16} = \frac{11}{16}$
   (g) $P(X \ge 2) = f(2) + f(3) + f(4) = 1 - F(1) = 1 - (\frac{1}{16} + \frac{4}{16}) = \frac{11}{16}$
   (h) $f(1) + f(3) = \frac{4}{16} + \frac{4}{16} = \frac{1}{2}$

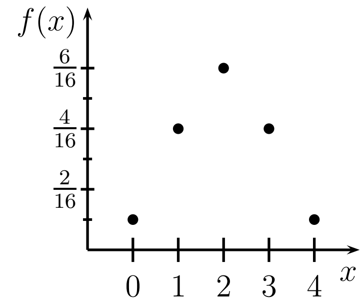
(i) $P(1 \le X \le 3) = f(1) + f(2) + f(3) = \frac{4}{16} + \frac{6}{16} + \frac{4}{16} = \frac{14}{16} = \frac{7}{8}$

(j) $P(1 \le X \le 3) = F(3) - F(0) = \frac{15}{16} - \frac{1}{16} = \frac{14}{16} = \frac{7}{8}$

3. (a)

x	0	1	2	3	4
$f(x)$	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

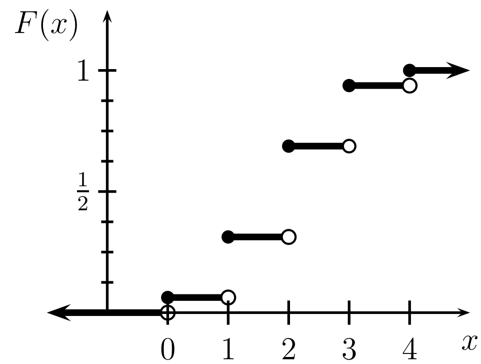
(b)



(c)

$$F(x) = \begin{cases} 0 & \text{for } x < 0 \\ \frac{1}{16} & \text{for } 0 \le x < 1 \\ \frac{5}{16} & \text{for } 1 \le x < 2 \\ \frac{11}{16} & \text{for } 2 \le x < 3 \\ \frac{15}{16} & \text{for } 3 \le x < 4 \\ 1 & \text{for } x \ge 4 \end{cases}$$

(d)



4.

x	0	1	2	3
$f(x)$	$\frac{3}{15}$	$\frac{7}{10}$	$\frac{4}{15}$	$\frac{1}{15}$

5. (a) $P(X \le 2) = F(2) = \frac{14}{15}$ (b) $P(X < 2) = F(1) = \frac{10}{15}$

(c) $P(X = 2) = F(2) - F(1) = f(2) = \frac{4}{15}$ (d) $P(X \ge 1) = 1 - F(0) = 1 - \frac{3}{15} = \frac{12}{15}$

(e) $P(X > 1) = 1 - F(1) = 1 - \frac{10}{15} = \frac{5}{15}$ (f) $P(X \le -2) = F(-2) = 0$

(g) $P(X \ge 5) = 0$ (h) $P(X \le 5) = F(5) = 1$

(i) $P(1 \le X \le 3) = F(3) - F(0) = 1 - \frac{3}{15} = \frac{12}{15}$

6. (a)

x	1	2	3	4	$\cdots$
$f(x)$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	$\cdots$

(b)

$$F(x) = \begin{cases} 0 & \text{for } x < 1 \\ \frac{1}{2} & \text{for } 1 \le x < 2 \\ \frac{3}{4} & \text{for } 2 \le x < 3 \\ \frac{7}{8} & \text{for } 3 \le x < 4 \\ \frac{15}{16} & \text{for } 4 \le x < 5 \\ \vdots & \quad\vdots \end{cases}$$

(c) $f(x) = \dfrac{1}{2^x}$ (d) $F(x) = \dfrac{2^n - 1}{2^n}$ for $n \le x < n + 1$

8. $1 = f(1) + f(2) + f(3) + f(4) = 1c + 2c + 3c + 4c = 10c$, so $c = \frac{1}{10}$

9. $1 = f(1) + f(2) + f(3) + f(4) = \frac{c}{1} + \frac{c}{2} + \frac{c}{3} + \frac{c}{4} = \frac{25c}{12}$, so $c = \frac{12}{25}$

10. $$1 = f(0) + f(1) + f(2) = c\binom{3}{0}\binom{5}{2} + c\binom{3}{1}\binom{5}{1} + c\binom{3}{2}\binom{5}{0}$$
$$= 10c + 15c + 3c = 28c, \text{ so } c = \tfrac{1}{28}$$

Section 2.3

1. (a) $\frac{76}{983}$ (b) $\frac{35+33+15+11+6}{983}$ (c) $1 - \frac{181+32}{983} = \frac{770}{983}$

2. 1, because the total probability is one

3. (a) $0.045 + 0.03 + 0.015 + 0.01 = 0.10$ (b) $1 - 0.10 = 0.90$
   (c) can't tell (d) $0.085 + 0.135 = 0.22$ (e) 1

4. (a) 0.4 (b) 0.5 (c) 0.9 (d) 0.1

Section 2.4

1. (a) $P(0 \le X \le 1) = \int_0^1 \frac{1}{2}x\, dx = \frac{1}{4}x^2\Big]_0^1 = \frac{1}{4}(1)^2 - \frac{1}{4}(0)^2 = \frac{1}{4}$

   (b) $P(\frac{1}{2} \le X \le 2) = \int_{\frac{1}{2}}^2 \frac{1}{2}x\, dx = \frac{1}{4}x^2\Big]_{\frac{1}{2}}^2 = \frac{1}{4}(2)^2 - \frac{1}{4}(\frac{1}{2})^2 = 1 - \frac{1}{16} = \frac{15}{16}$

   (c) $P(X = 1) = 0$ (d) $P(-1 \le X \le 1) = \int_{-1}^0 0\, dx + \int_0^1 \frac{1}{2}x\, dx = 0 + \frac{1}{4} = \frac{1}{4}$

   (e) $P(X \le 3) = 1$ since the only non-zero probability is from zero to two.

   (f) $P(X \le -2) = 0$ for the same reason.

2. (a) $P(2 \le X \le 4) = \int_2^4 \frac{1}{x^2}\, dx = -\frac{1}{x}\Big]_2^4 = -\frac{1}{4} - \left(-\frac{1}{2}\right) = -\frac{1}{4} + \frac{1}{2} = \frac{1}{4}$

   (b) $P(0 \le X \le 2) = P(1 \le X \le 2) = \int_1^2 \frac{1}{x^2}\, dx = -\frac{1}{x}\Big]_1^2 = -\frac{1}{2} - \left(-\frac{1}{1}\right) = \frac{1}{2}$

   (c) $P(X \ge 1) = 1$ since all of the probability is from one on.

   (d) $P(X \le 5) = P(1 \le X \le 5) = \int_1^5 \frac{1}{x^2}\, dx = -\frac{1}{x}\Big]_1^5 = -\frac{1}{5} - \left(-\frac{1}{1}\right) = \frac{4}{5}$

   (e) $P(X > 3) = \int_3^\infty \frac{1}{x^2}\, dx = -\frac{1}{x}\Big]_3^\infty = -\frac{1}{\infty} - \left(-\frac{1}{3}\right) = 0 + \frac{1}{3} = \frac{1}{3}$

   (f) $P(X = 7) = 0$

3. $1 = \int_0^5 C\, dx = Cx]_0^5 = 5C - 0C = 5C$, so $C = \frac{1}{5}$.

4. $1 = \int_1^4 Cx\, dx = \frac{1}{2}Cx^2\Big]_1^4 = 8C - \frac{1}{2}C = \frac{15}{2}$, so $C = \frac{2}{15}$.

5. (a) $F(x) = 0$ for $x \le 0$. For $0 < x \le 5$, $F(x) = \displaystyle\int_{-\infty}^{x} f(t)\ dt = \int_0^x \tfrac{1}{5}\ dt = \tfrac{1}{5}t\Big]_0^x =$ $\tfrac{1}{5}x$. For $x > 5$, $F(x) = 1$. Thus $F(x) = \begin{cases} 0 & \text{for } x < 0 \\ \frac{1}{5}x & \text{for } 0 \le x \le 5 \\ 1 & \text{for } x > 5 \end{cases}$

(b) $F(x) = 0$ for $x \le 0$. For $0 < x \le 2$, $F(x) = \displaystyle\int_{-\infty}^{x} f(t)\ dt = \int_0^x \tfrac{1}{2}t\ dt = \tfrac{1}{4}t^2\Big]_0^x =$ $\tfrac{1}{4}x^2$. For $x > 2$, $F(x) = 1$. Therefore $F(x) = \begin{cases} 0 & \text{for } x < 0 \\ \frac{1}{4}x^2 & \text{for } 0 \le x \le 2 \\ 1 & \text{for } x > 2 \end{cases}$

6. (a) $P(X < 3) = P(X \le 3) = F(3) = \frac{\sqrt{3}}{2}$

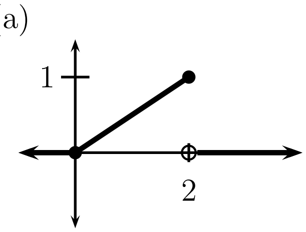
(b) $P(X \ge 1) = 1 - P(X < 1) = 1 - P(X \le 1) = 1 - F(1) = 1 - \frac{\sqrt{1}}{2} = \frac{1}{2}$

(c) $P(1 < X \le 3) = F(3) - F(1) = \frac{\sqrt{3}}{2} - \frac{1}{2} = \frac{\sqrt{3}-1}{2}$ (d) $P(X = 2) = 0$

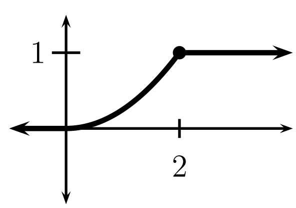
(e) $P(X \ge -1) = 1 - P(X \le -1) = 1 - F(-1) = 1 - 0 = 1$

(f) $P(X = -1) = F(-1) = 0$

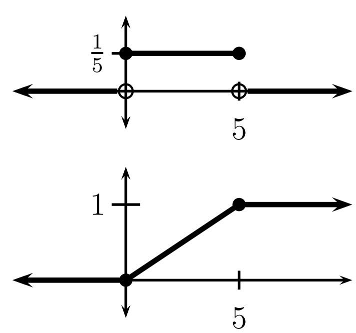
7. (a)



(b)



8.



9. Since F "picks up" probability as x increases, and it has reached value one by the time it reaches $x = 5$, we must have $c(5-2)^2 = 1$, so $c = \frac{1}{9}$.

10. (a) $P(X \le 4) = F(4) = \frac{4}{9}$ (b) $P(X < 4) = F(4) = \frac{4}{9}$

(c) $P(X = 4) = 0$ (d) $P(X \ge 3) = 1 - P(X \le 3) = 1 - F(3) = 1 - \frac{1}{9} = \frac{8}{9}$

(e) $P(3 \le X < 4) = F(4) - F(3) = \frac{4}{9} - \frac{1}{9} = \frac{3}{9}$

(f) $P(1 < X < 4) = F(4) - F(1) = \frac{4}{9} - 0 = \frac{4}{9}$

(g) $P(X \ge 7) = 1 - P(X \le 7) = 1 - F(7) = 1 - 1 = 0$ (h) $P(X \le 7) = F(7) = 1$

11. $f(x) = F'(x) = \frac{2}{9}(x-2)$ for $2 < x < 5$ and $f(x) = 0$ otherwise.

$$P(3 \le X < 4) = \int_3^4 \frac{2}{9}(x-2)\, dx = \int_3^4 \left(\frac{2}{9}x - \frac{4}{9}\right)\, dx = \frac{1}{9}x^2 - \frac{4}{9}x\Big]_3^4 = \left(\frac{16}{9} - \frac{16}{9}\right) - \left(\frac{9}{9} - \frac{12}{9}\right) = \frac{3}{9}$$

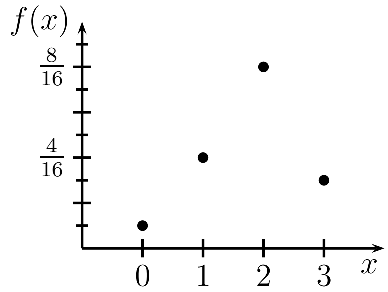
12. Note that $F(x) = \frac{1}{2}x^{\frac{1}{2}}$, so $f(x) = F'(x) = \frac{1}{4}x^{-\frac{1}{2}} = \frac{1}{4\sqrt{x}}$ for $0 \le x \le 4$ and $f(x) = 0$ otherwise.

Section 2.5

1. (a) $\frac{0+1+2+3}{4} = \frac{3}{2} = 1.5$

   (b) $\mu = E(X) = 0 \cdot \frac{1}{16} + 1 \cdot \frac{4}{16} + 2 \cdot \frac{8}{16} + 3 \cdot \frac{3}{16} = \frac{4}{16} + \frac{16}{16} + \frac{9}{16} = \frac{29}{16}$

   (c)



   (d) $\sigma^2 = (0 - \frac{29}{16})^2 \cdot \frac{1}{16} + (1 - \frac{29}{16})^2 \cdot \frac{4}{16} + (2 - \frac{29}{16})^2 \cdot \frac{8}{16} + (3 - \frac{29}{16})^2 \cdot \frac{3}{16} = \frac{83}{128} \approx 0.65$

   (e) $\sigma = \sqrt{\frac{83}{128}} \approx 0.81$

3. $E(X - \mu) = 0$

4. $\frac{11}{8},\ \frac{47}{64}$

5. $\mu = 2,\ \sigma^2 = 1,\ 2$

7. $\mu = p,\ p - p^2 = p(1-p)$

8. $\mu = \frac{4}{3},\ \sigma^2 = \frac{2}{9}$

9. $\mu = \frac{5}{2},\ \sigma^2 = \frac{25}{12}$

10. $\mu = \frac{14}{5},\ \sigma^2 = \frac{33}{50}$

11. The distribution has no expected value, since the integral diverges.

H.3 Chapter 3 Solutions

Section 3.1

1. (a) Not a Bernoulli process, because the probability for success on each trial is different and because the trials are not independent.
   (b) Not a Bernoulli process because there are more than two outcomes.
   (c) Bernoulli process, success is drawing a blue and $p = \frac{7}{17}$.
   (d) Bernoulli process, success is encountering a green light and $p = \frac{60}{140} = \frac{3}{7}$.
   (e) Not a Bernoulli process, then number of trials is not fixed.
   (f) Bernoulli process, success is testing a defective part, p is unknown.

2. (b) $b(x; 3, \frac{7}{17})$ (c) $b(x; 5, \frac{3}{7})$ (d) $b(x; 10, p)$

3. (a) $b\left(x; 3, \dfrac{7}{17}\right) = \begin{pmatrix} 3 \\ x \end{pmatrix} \left(\dfrac{7}{17}\right)^x \left(\dfrac{10}{17}\right)^{3-x}, \qquad x = 0,\ 1,\ 2,\ 3$

x	0	1	2	3
$b(x; 3, \frac{7}{17})$	$\frac{1000}{4913}$	$\frac{2100}{4913}$	$\frac{1470}{4913}$	$\frac{343}{4913}$

   (b) $b\left(x; 5, \dfrac{3}{7}\right) = \begin{pmatrix} 5 \\ x \end{pmatrix} \left(\dfrac{3}{7}\right)^x \left(\dfrac{4}{7}\right)^{5-x}, \qquad x = 0,\ 1,\ 2,\ 3,\ 4,\ 5$

x	0	1	2	3	4	5
$b(x; 5, \frac{4}{7})$	$\frac{1024}{16807}$	$\frac{3840}{16807}$	$\frac{5760}{16807}$	$\frac{4320}{16807}$	$\frac{1620}{16807}$	$\frac{243}{16807}$

4. (a) 0.13 (b) 9.7×10^{-6} (c) 0.99 (d) 0.14

5. (a)

x	0	1	2	3
$b(x; 3, \frac{1}{5})$	$\frac{64}{125}$	$\frac{48}{125}$	$\frac{12}{125}$	$\frac{1}{125}$

   (b)
$$B(x; 3, \tfrac{1}{5}) = \begin{cases} 0 & \text{for } x < 0 \\ \frac{64}{125} & \text{for } 0 \leq x < 1 \\ \frac{112}{125} & \text{for } 1 \leq x < 2 \\ \frac{124}{125} & \text{for } 2 \leq x < 3 \\ 1 & \text{for } x \geq 3 \end{cases}$$

   (c) 1, 3, $\frac{3}{5}$ (d) $E(X) = 0(\frac{64}{125}) + 1(\frac{48}{125}) + 2(\frac{12}{125}) + 3(\frac{1}{125}) = \frac{3}{5}$
   (e) $\sigma^2 = E(X^2) - [E(X)]^2 = \frac{105}{125} - \frac{9}{25} = \frac{12}{25}$

6. $\sigma^2 = npq = 3(\frac{1}{5})(\frac{4}{5}) = \frac{12}{25}$

7. (a) 0, 1, 2, 3,
   (b) $b(0; n, p) = q^3,\ b(1; n, p) = 3pq^2,\ b(2; n, p) = 3p^2q,\ b(3; n, p) = p^3$
   (c) $q^3 + 3pq^2 + 3p^2q + p^3 = 1$
   (d) $(p + q)^3 = (p + q)(p^2 + 2pq + q^2) = p^3 + 3p^2q + 3pq^2 + q^3$
   (e) $p + q = 1$, so $(p + q)^3 = 1$

Section 3.2:

1. (a) 0.0314
   (b) $P(Z \geq 2.26) = 1 - p(Z \leq 2.26) = 1 - 0.9881 = 0.0119$
   (c) $P(-1.54 \leq Z \leq -0.13) = 0.4483 - 0.0618 = 0.3865$
   (d) $P(-2.09 \leq Z \leq 1.27) = 0.8980 - 0.0183 = 0.8797$

2. (a) 0.20 (b) -0.55 (c) 0.48 (d) 1.41

Section 3.3:

1. (a) $Z = \frac{6-5.3}{1.7} = 0.41, \qquad P(x \leq 6) = P(Z \leq 0.41) = 0.6591$
   (b) $P(3 \leq X \leq 7) = P(-1.35 \leq Z \leq 1.00) = 0.8413 - 0.0885 = 0.7528$
   (c) $P(X \geq 8) = 1 - P(X \leq 8) = 1 - P(Z \leq 1.59) = 1 - 0.9441 = 0.0554$

2. $P(X \leq 247) = P(Z \leq -1.40) = 0.0808$

3. 6.7 years 4. 3.1 years 5. 3.9-6.7 years

H.4 Chapter 4 Solutions

Section 4.1:

1. (a) Since the marbles are drawn without replacement, the probability of a success on each trial does not remain the same.

   (c) $\dfrac{15 \cdot 14 \cdot 35}{40 \cdot 39 \cdot 38}$ (d) $\dbinom{3}{2}$ (e) $\dbinom{3}{2}\dfrac{15 \cdot 14 \cdot 35}{40 \cdot 39 \cdot 38}$

2. $\dbinom{5}{3}\dfrac{15 \cdot 14 \cdot 13 \cdot 35 \cdot 34}{40 \cdot 39 \cdot 38 \cdot 37 \cdot 36}$

3. (a) $h(4; 274, 10, 87) = 0.219$ (b) $h(0; 274, 10, 87) = 0.020$

4. (a)

x	0	1	2	3
$h(x;10,3,6)$	$\frac{1}{30}$	$\frac{3}{10}$	$\frac{1}{2}$	$\frac{1}{6}$

   (b)

$$H(x;10,3,6) = \begin{cases} 0 & \text{for } x < 0 \\ \frac{1}{30} & \text{for } 0 \le x < 1 \\ \frac{10}{30} & \text{for } 1 \le x < 2 \\ \frac{25}{30} & \text{for } 2 \le x < 3 \\ 1 & \text{for } x \ge 3 \end{cases}$$

5. (a) $h(1; 10, 3, 6) = \frac{3}{10}$ (b) $H(2; 10, 3, 6) = \frac{25}{30}$ (c) $h(0; 10, 3, 6) = \frac{1}{30}$

   (d) $1 - H(0; 10, 3, 6) = 1 - \frac{1}{30} = \frac{29}{30}$

7. (a) 0.1837175 (b) 0.182276

   (c) The answers are very close. This is because the number of defective parts is so small relative to the total number of parts that the probability of selecting a defective part doesn't change much with or without replacement.

8. (a) The error is 0.000899. (b) The percent error is 0.5%.

Section 4.2

1. (a) $2\left(\frac{2}{5}\right)^2\left(\frac{3}{5}\right) = \frac{24}{125}$

2. (a) $\left(\dfrac{2}{5}\right)^7\left(\dfrac{3}{5}\right)^3$ (b) $\dbinom{9}{6}$ (c) $\dbinom{9}{6}\left(\dfrac{2}{5}\right)^7\left(\dfrac{3}{5}\right)^3 = \dfrac{290304}{9765625} = 0.02973$

1. $P(X = x) = \dbinom{x-1}{k-1}(p)^k(q)^{x-k}$

2. $b^*(15; 4, 0.13) = 0.0225$

Section 4.3

1. (a) $\lambda = 315 \text{ trucks/day} \cdot \dfrac{1 \text{ day}}{1440 \text{ minutes}} = 0.21875 \text{ trucks/minute}$

   (b) $p(2; (0.21875)(5)) = p(2; 1.09375) = \dfrac{e^{-1.09375}(1.09375)^2}{2!} = 0.20$

2. $\displaystyle\sum_{x=3}^{\infty} p(x; 1.09375) = 1 - \sum_{x=0}^{2} p(x; 1.09375) = 1 - P(2; 1.09375) = 0.0983$

4. (a) 0.1832 (b) 0.1823 (c) 0.1804

Section 4.4

1. (a) $P(X = 5) = 0$

   (b) $P(3 \leq X \leq 5) = F(5) - F(3) = (1 - e^{-\frac{5}{8}}) - (1 - e^{-\frac{3}{8}}) = e^{-\frac{3}{8}} - e^{-\frac{5}{8}} = 0.152$

   (c) $P(X < 10) = 1 - e^{-\frac{10}{8}} = 0.713$

   (d) $P(x > 10) = 1 - P(X < 10) = 1 - 0.713 = 0.287$

2. (a) $\lambda = \frac{1}{8}, \quad p(3; , \dfrac{1}{8} \cdot 15) = \dfrac{e^{-\frac{15}{8}}(\frac{15}{8})^3}{3!} = 0.168$

   (b) $p(0; \frac{10}{8}) = e^{-\frac{10}{8}} = 0.287$ (c) The answers are the same.

3. $P(X \leq 2) = \frac{1}{3.7} e^{-\frac{2}{3.7}} = 0.1574$

4. (b) $P(X \leq 3) = \frac{1}{3} e^{-1} = 0.1226$

   (c) $\dfrac{1}{2} = \displaystyle\int_{-\infty}^{x} \frac{1}{3} e^{-\frac{t}{3}} = -e^{-\frac{t}{3}}\Big]_{-\infty}^{x} = e^{-\frac{x}{3}} \Rightarrow \ln\left(\frac{1}{2}\right) = \ln(e^{-\frac{x}{3}}) = -\frac{x}{3} \Rightarrow x = -\frac{\ln\frac{1}{2}}{3}$

H.5 Chapter 5 Solutions

Section 5.1

1. (a) $\frac{35+45+24}{200} = \frac{104}{200} = \frac{13}{25}$ (b) $\frac{45+33}{200} = \frac{78}{200} = \frac{39}{100}$ (c) $\frac{45}{200} = \frac{9}{40}$

   (d) $\frac{35+45+24+33}{200} = \frac{137}{200}$ (e) $\frac{45}{45+33} = \frac{45}{78} = \frac{15}{23}$ (f) $\frac{45}{35+45+24} = \frac{45}{104}$

2. (a) $P(X = 1, Y = 0) = 0.225$ (b) $P(X = 1) = 0.225 + 0.165 = 0.390$

   (c) $P(Y = 1) = 0.235 + 0.165 + 0.080 = 0.480$

   (d) $P(X = 1 \mid Y = 0) = \dfrac{0.225}{0.175 + 0.225 + 0.120} = \dfrac{0.225}{0.52} = 0.433$

   (e) $P(X = 2 \text{ or } Y = 1) = 0.120 + 0.080 + 0.235 + 0.165 = 0.600$

3. (a) $S = \{1, 2, 3, 4, 5, 6\}$

   (b) $\text{Ran}(X) = \{0, 1\}$

   (c) $\text{Ran}(Y) = \{3, 4, 5\}$

   (d) See to the right.

		x		
	$f(x, y)$	0	1	2
y	0	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
	1	$\frac{2}{6}$	$\frac{1}{6}$	0

4. (a) $P(X = 1, Y = 0) = f(1, 0) = \frac{4}{24} = \frac{1}{6}$

   (b) $P(X = 1) = f(1, 0) + f(1, 1) + f(1, 2) = \frac{4}{24} + \frac{2}{24} + \frac{1}{24} = \frac{7}{24}$

   (c) $P(X + Y \leq 1) = f(0, 0 + f(1, 0) + f(0, 1) = \frac{5}{24} + \frac{4}{24} + \frac{3}{24} = \frac{12}{24} = \frac{1}{2}$

   (d) $P(X = 2 \text{ or } Y = 1) = f(2, 0)+f(2, 1)+f(2, 2)+f(0, 1)+f(1, 1)+f(3, 1) = \frac{10}{24} = \frac{5}{12}$

   (e) $P(X + Y \leq 3) = 1 - [f(2, 2) + f(3, 1) + f(3, 2)] = 1 - [0 + \frac{1}{24} + 0] = \frac{23}{24}$

   (f) $P(Y = 2|X = 0) = \dfrac{f(0, 2)}{f(0, 0) + f(0, 1) + f(0, 2)} = \dfrac{\frac{2}{24}}{\frac{5}{24} + \frac{3}{24} + \frac{2}{24}} = \dfrac{2}{10} = \dfrac{1}{5}$

5. (a)

		x			
	$f(x, y)$	0	1	2	3
y	0	$\frac{60}{720}$	$\frac{180}{720}$	$\frac{90}{720}$	$\frac{6}{720}$
	1	$\frac{120}{720}$	$\frac{180}{720}$	$\frac{36}{720}$	0
	2	$\frac{30}{720}$	$\frac{18}{720}$	0	0
	3	0	0	0	0

(b)

		x			
	$f(x,y)$	0	1	2	3
y	0	$\frac{125}{1000}$	$\frac{225}{1000}$	$\frac{135}{1000}$	$\frac{27}{1000}$
	1	$\frac{150}{1000}$	$\frac{180}{1000}$	$\frac{54}{1000}$	0
	2	$\frac{60}{1000}$	$\frac{36}{1000}$	0	0
	3	$\frac{8}{1000}$	0	0	0

Section 5.2

1. (a)

x	0	1	2
$g(x)$	$\frac{3}{6}$	$\frac{2}{6}$	$\frac{1}{6}$

(b)

y	0	1
$h(y)$	$\frac{3}{6}$	$\frac{3}{6}$

2. (a) $P(X=2) = f(2,0) + f(2,1)$ (b) $P(X=2) = g(2)$
   (c) $P(X=1, Y=0) = f(1,0)$
   (d) $P(Y=0 \mid X=1) = \frac{f(1,0)}{g(1)}$
   (e) $P(X=1 \text{ or } Y=0) = f(1,0) + f(1,1) + f(0,0) + f(2,0)$
   (f) $P(X=1 \text{ or } Y=0) = g(1) + h(0) - f(1,0)$

3.

x	0	1	2	3
$g(x)$	$\frac{10}{24}$	$\frac{7}{24}$	$\frac{4}{24}$	$\frac{3}{24}$

y	0	1	2
$h(y)$	$\frac{14}{24}$	$\frac{7}{24}$	$\frac{3}{24}$

4. (a) $P(X=2) = g(2) = \frac{4}{24} = \frac{1}{6}$ (b) (b) $P(Y=2) = h(2) = \frac{3}{24} = \frac{1}{8}$
   (c) $P(Y \le 1) = h(0) + h(1) = \frac{14}{24} + \frac{7}{24} = \frac{21}{24} = \frac{7}{8}$
   (d) $P(X=2 \text{ or } Y=1) = g(2) + h(1) - f(2,1) = \frac{4}{24} + \frac{7}{24} - \frac{1}{24} = \frac{10}{24} = \frac{5}{12}$
   (e) $P(X=3|Y=1) = \dfrac{f(3,1)}{h(1)} = \dfrac{\frac{1}{24}}{\frac{7}{24}} = \dfrac{1}{7}$ (f) $P(Y=2|X=0) = \dfrac{f(0,2)}{g(0)} = \dfrac{\frac{2}{24}}{\frac{10}{24}} = \dfrac{1}{5}$

5. (a)

x	0	1	2	3
$g(x)$	$\frac{210}{720}$	$\frac{378}{720}$	$\frac{126}{720}$	$\frac{6}{720}$

(b)

y	0	1	2
$h(y)$	$\frac{336}{720}$	$\frac{336}{720}$	$\frac{48}{720}$

Section 5.3

1.

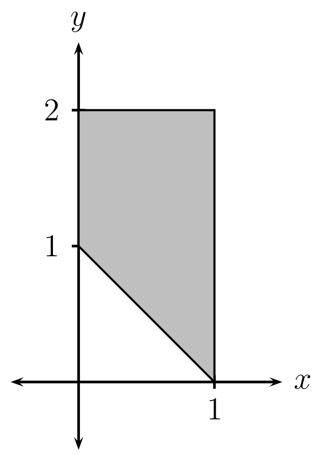
x	0	1	2	3
$v(x|0)$	$\frac{5}{14}$	$\frac{4}{14}$	$\frac{3}{14}$	$\frac{2}{14}$

y	0	1	2
$w(y|3)$	$\frac{2}{3}$	$\frac{1}{3}$	0

Because $f(0,0) = \frac{5}{24} \neq \frac{35}{144} = \frac{10}{24} \cdot \frac{14}{24} = g(0)h(0)$ X and Y are not independent.

Section 5.5

1. $P(X \geq \frac{1}{2}, 0 \leq Y \leq \frac{3}{2}) = \frac{1}{4}\int_0^{\frac{3}{2}} \int_{\frac{1}{2}}^{1} (2x+y)\,dx\,dy = \frac{27}{64}$



2. (a) See graph to right.

   (b) $P(X+Y \geq 1) = \int_0^1 \int_{1-x}^{2} \frac{1}{4}(2x+y)\,dy\,dx$

3. $h(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \int_0^1 \frac{1}{4}(2x+y)\,dx = \frac{1}{4}(1+y),$

   $P(1 \leq Y \leq 2) = \int_1^2 h(y)\,dy = \frac{5}{8}$

4. (a) $v(x\,|\,y) = \dfrac{f(x,y)}{h(y)} = \dfrac{2x+y}{1+y}$

   (b) $P(X \leq \frac{1}{2} \mid Y = 1) = \int_0^{\frac{1}{2}} v(x\,|\,1)\,dx = \frac{3}{8}$

Section 5.6

1. (a)

x	0	1
$g(x)$	$\frac{1}{4}$	$\frac{3}{4}$

y	0	1
$h(y)$	$\frac{5}{8}$	$\frac{3}{8}$

   (b) $\sum_{x=0}^{1} x\,g(x) = \frac{3}{4}$ (c) $\sum_{x=0}^{1}\sum_{y=0}^{1} x\,f(x,y) = \frac{3}{4}$ (d) $\mu_Y = \sum_{y=0}^{1} y\,h(y) = \frac{3}{8}$

   (e) $E(XY) = \frac{9}{32}$ (f) $\sigma_{XY} = 0$

   (g) $f(0,0) = \frac{5}{32} = \frac{5}{8} \cdot \frac{1}{4} = g(0)h(0), \quad f(1,0) = \frac{15}{32} = \frac{5}{8} \cdot \frac{3}{4} = g(1)h(0),$

   $f(0,1) = \frac{3}{32} = \frac{1}{4} \cdot \frac{3}{8} = g(0)h(1), \quad f(1,1) = \frac{9}{32} = \frac{3}{4} \cdot \frac{3}{8} = g(1)h(1)$

2. $g(x) = x + \frac{1}{2}$ for $x \in [0,1]$, 0 elsewhere, $h(y) = \frac{1}{4}y + \frac{1}{4}$ for $y \in [0,2]$, 0 elsewhere

   $\mu_X = E(X) = \frac{7}{12}, \quad \mu_Y = E(Y) = \frac{7}{6}, \quad \sigma_{XY} = E(XY) - \mu_x\mu_y = -\frac{1}{72}$

I Solutions to Appendix Exercises

I.1 Appendix C Solutions

1. (a) $B = \{x \in \mathbb{N} \mid x > 7\}$ (b) $[7, \infty)$ (c) $\{2, 3, 4, 5, 6\}$
   (d) $\{x \in \mathbb{R} \mid 2 \leq x \leq 6\}$, $[2, 6]$

4. (a) $C \subseteq D$ (b) $C \nsubseteq D$ because $5 \in C$ and $5 \notin D$. (c) $C \subseteq D$
   (d) $C \nsubseteq D$ because $1 \in C$ and $1 \notin D$. Note that it is sufficient to provide just *one* element in C that is not in D.
   (e) $C \subseteq D$ (f) $C \nsubseteq D$ because $1.5 \in C$ and $1.5 \notin D$.

5. Every set is a subset of itself.

6. $5 \in A$ and $\{5\} \subseteq A$ are valid statements, $5 \subseteq A$ is not.

7. The largest possible subset of A is A itself.

8. $\emptyset, \{a\}, \{b\}, \{a, b\}$

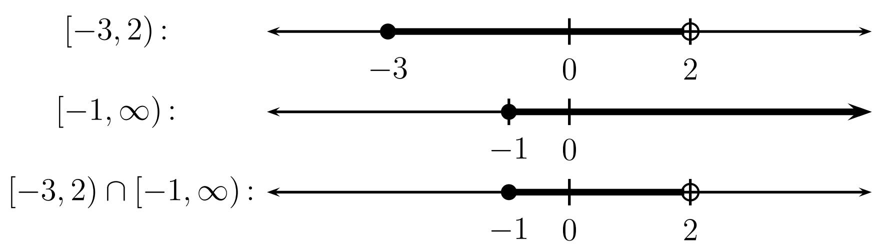
9. $|A| \leq |B|$

10. $A \cap B = \{4, 6\}$, $A \cup B = \{2, 3, 4, 5, 6, 8, 10\}$

11. (a) $A' = \{1, 3, 5, 7, 9\}$, $B' = \{1, 2, 7, 8, 9, 10\}$, $A - B = \{2, 8, 10\}$
    (b) $A' \cap B' = \{1, 7, 9\}$, $(A \cap B)' = \{1, 2, 3, 5, 7, 8, 9, 10\}$. The two sets are not the same.
    (c) C can be any set of elements of U other than 2, 4, 6, 8, 10.

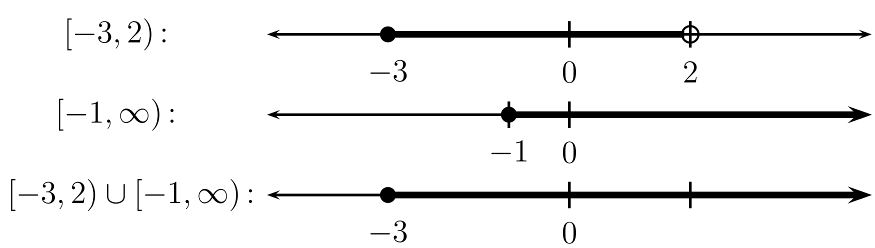
12. (a) If $A \subseteq B$, then $A \cap B = A$. (b) If $B \subseteq A$, then $A \cup B = A$.
    (c) If $A = B$, then $A \cap B = A \cup B$.

13. (a)



$[-1, 2)$ - This is obtained by shading only the part of the number line that is shaded on the graphical representations of *BOTH* of the two original sets.

(b)



$[-3, \infty)$ - This is obtained by shading the part of the number line that is shaded on the graphical representations of *EITHER* of the two original sets.

14. $A \cap B = [1, 4.1]$, $A \cup B = (-2, 8]$, $A' = (-\infty, -2] \cup (4.1, \infty)$, $B' = (-\infty, 1) \cup (8, \infty)$, $A' \cap B' = (-\infty, -2] \cup (8, \infty)$, $(A \cap B)' = (-\infty, 1) \cup (4.1, \infty)$, $A - B = (-2, 1)$.

15. (a) 

(b) $[-2, 3) \cup (3, \infty)$

16. $A \cup B = [3, 5] \cup \{6, 7\}$, $A \cap B = \{4, 5\}$, $A - B = (3, 4) \cup (4, 5)$, $B - A = \{3, 6, 7, \}$

17. (a) $(-\infty, 2) \cup (2, \infty)$ (b) $\{x \in \mathbf{R} \mid x \neq 2\}$ (c) All real numbers except 2.

18. (a) $\leq, +$ (b) If A and B are disjoint, then $|A \cup B| = |A| + |B|$
(c) $|A \cup B| = |A| + |B| - |A \cap B|$

19. (a) Not a partition because $A_1 \cap A_2 = \{1\} \neq \emptyset$.
(b) Partition.
(c) Not a partition because intersections of some of the sets are non-empty AND because one of the sets is the empty set.
(d) Not a partition because the union of all the sets is not the entire set A.

20. Include the sets $\{3\}$ and $[4, 6]$, for one example.

21. $A \times B = \{(a, 1), (a, 2), (a, 3), (a, 4), (b, 1), (b, 2), (b, 3), (b, 4), (c, 1), (c, 2), (c, 3), (c, 4)\}$

22. In general, $A \times B \neq B \times A$. It IS true in the case that $A = B$.

23. $|A \times B| = mn$

25. (a) $\{(x, y) \in \mathbf{Z} \times \mathbf{Z}\}$ or $\{(x, y) \mid x, y \in \mathbf{Z}\}$ (c) $\{(x, y) \in \mathbf{R} \times \mathbf{R} \mid x = 3\}$
(e) $\{(x, y) \in \mathbf{R} \times \mathbf{R} \mid -1 < x \leq 3 \text{ and } 2 \leq y < 4\}$ (f) $\{(x, y) \in \mathbf{R} \times \mathbf{R} \mid x \geq 2\}$

I.2 Appendix D Solutions

1. (a) The rule is a function.
(b) The rule is not a function because the element $c \in A$ is not assigned an element in B and because d is assigned two values in B. (Of course either one of these alone is enough to show that the rule is not a function.)
(c) The rule is a function.
(d) The rule is not a function because the element $b \in A$ is assigned more than one element in B.
(e) The rule is not a function because the element $b \in A$ is not assigned an element in B.
(f) The rule is a function.

2. $\text{Ran}(f) = \{n \in \mathbf{N} \mid 1 \leq n \leq 26\}$ and $\text{Ran}(g) = \{0, 1\}$.

3. (a) $\text{Dom}(f) = (-\infty, \infty)$ and $\text{Ran}(f) = (-\infty, 0]$.
(b) $\text{Dom}(g) = (-\infty, \infty)$ and $\text{Ran}(g) = (-\infty, \infty)$.

(c) $\text{Dom}(h) = (-\infty, 5]$ and $\text{Ran}(h) = [0, \infty)$.

(d) $\text{Dom}(f) = (-\infty, -3) \cup (-3, \infty)$ and $\text{Ran}(f) = (-\infty, 0) \cup (0, \infty)$.

4. (a) $\text{Dom}(f) = \mathbf{R}$ and $\text{Ran}(f) = \{x \in \mathbf{R} \mid x \leq 0\}$.

   (b) $\text{Dom}(g) = \mathbf{R}$ and $\text{Ran}(g) = \mathbf{R}$.

   (c) $\text{Dom}(h) = \{x \in \mathbf{R} \mid x \leq 5\}$ and $\text{Ran}(h) = \{x \in \mathbf{R} \mid x \geq 0\}$.

   (d) $\text{Dom}(f) = \{x \in \mathbf{R} \mid x \neq -3\}$ and $\text{Ran}(f) = \{x \in \mathbf{R} \mid x \neq 0\}$.

5. (a) The range of f is $[0, \infty)$ and the range of g is $\{1, 4, 9, 16, 25, ...\}$, which can also be written as $\{n^2 \mid n \in \mathbf{N}\}$.

   (b) If we restrict the domain of f to $[0, \infty)$ the range remains unchanged. There are many other ways to restrict the domain without restricting the range, the most notable of which is to $(-\infty, 0]$.

6. (a) $\text{Dom}(f) = \mathbf{Z}$ and $\text{Ran}(f) = \{..., -9, -4, -1, 0\}$.

   (b) $\text{Dom}(g) = \mathbf{Z}$ and $\text{Ran}(g) = \{..., -3, -2, -1, 0, 1, 2, 3, ...\}$.

   (c) $\text{Dom}(h) = \{..., 2, 3, 4, 5\}$ and $\text{Ran}(h) = \{0, 1, \sqrt{2}, \sqrt{3}, 2, \sqrt{5}, ...\}$.

   (d) $\text{Dom}(f) = \{n \in \mathbf{Z} \mid n \neq -3\}$ and $\text{Ran}(f) = \{1, -1, \frac{1}{2}, -\frac{1}{2}, \frac{1}{3}, -\frac{1}{3}, ...\}$.

I.3 Appendix F Solutions

1. Example 2, $r = -1$; Example 3, $r = \frac{1}{3}$; Example 5, $r = \frac{1}{2}$

3. (a) $1, 2, 4, 7, 12, 20, 33, ...$

   (b) $1, 0, 1, 0, 1, 0, ...$

   (c) $1, \frac{3}{2}, \frac{7}{4}, \frac{15}{8}, \frac{31}{16}, ...$

   (d) $1, 4, 9, 16, 25, ...$

4. (a) $a_n = \frac{1}{2^n}, \qquad \lim_{n\to\infty} a_n = 0$

   (b) $s_1 = \frac{1}{2},\ s_2 = \frac{3}{4}, s_3 = \frac{7}{8},\ s_4 = \frac{15}{16}$

   (c) $s_n = \dfrac{2^n - 1}{2^n}, \quad \lim_{n\to\infty} s_n = 1$

   (d) $L - s_n = 1 - \dfrac{2^n - 1}{2^n} = \dfrac{1}{2^n}$

   (e) $n \geq 14$

5. $|r| = |\frac{1}{2}| < 1$, so the series converges. $S = \dfrac{\frac{1}{2}}{1 - \frac{1}{2}} = 1$

6. $|r| = |\frac{1}{3}| < 1$, so the series converges. $S = \dfrac{\frac{1}{3}}{1 - \frac{1}{3}} = \dfrac{1}{2}$

Index