Chapter 3

Probability

Probability forms the foundation of statistics, and you're probably already aware of many of the ideas presented in this chapter. However, formalization of probability concepts is likely new for most readers.

While this chapter provides a theoretical foundation for the ideas in later chapters and provides a path to a deeper understanding, mastery of the concepts introduced in this chapter is not required for applying the methods introduced in the rest of this book.

For videos, slides, and other resources, please visit
www.openintro.org/os

3.1 Defining probability

Statistics is based on probability, and while probability is not required for the applied techniques in this book, it may help you gain a deeper understanding of the methods and set a better foundation for future courses.

3.1.1 Introductory examples

Before we get into technical ideas, let's walk through some basic examples that may feel more familiar.

EXAMPLE 3.1

A "die", the singular of dice, is a cube with six faces numbered `1`, `2`, `3`, `4`, `5`, and `6`. What is the chance of getting `1` when rolling a die?

E

If the die is fair, then the chance of a `1` is as good as the chance of any other number. Since there are six outcomes, the chance must be 1-in-6 or, equivalently, $1/6$.

EXAMPLE 3.2

What is the chance of getting a `1` or `2` in the next roll?

E

`1` and `2` constitute two of the six equally likely possible outcomes, so the chance of getting one of these two outcomes must be $2/6 = 1/3$.

EXAMPLE 3.3

What is the chance of getting either `1`, `2`, `3`, `4`, `5`, or `6` on the next roll?

E

100%. The outcome must be one of these numbers.

EXAMPLE 3.4

What is the chance of not rolling a `2`?

E

Since the chance of rolling a `2` is $1/6$ or $16.\bar{6}\%$, the chance of not rolling a `2` must be $100\% - 16.\bar{6}\% = 83.\bar{3}\%$ or $5/6$.

Alternatively, we could have noticed that not rolling a `2` is the same as getting a `1`, `3`, `4`, `5`, or `6`, which makes up five of the six equally likely outcomes and has probability $5/6$.

EXAMPLE 3.5

Consider rolling two dice. If $1/6$ of the time the first die is a `1` and $1/6$ of those times the second die is a `1`, what is the chance of getting two `1`s?

E

If $16.\bar{6}\%$ of the time the first die is a `1` and $1/6$ of *those* times the second die is also a `1`, then the chance that both dice are `1` is $(1/6) \times (1/6)$ or $1/36$.

3.1.2 Probability

We use probability to build tools to describe and understand apparent randomness. We often frame probability in terms of a **random process** giving rise to an **outcome**.

Roll a die	$\rightarrow$	1, 2, 3, 4, 5, or 6
Flip a coin	$\rightarrow$	H or T

Rolling a die or flipping a coin is a seemingly random process and each gives rise to an outcome.

PROBABILITY

The **probability** of an outcome is the proportion of times the outcome would occur if we observed the random process an infinite number of times.

Probability is defined as a proportion, and it always takes values between 0 and 1 (inclusively). It may also be displayed as a percentage between 0% and 100%.

Probability can be illustrated by rolling a die many times. Let $\hat{p}_n$ be the proportion of outcomes that are 1 after the first n rolls. As the number of rolls increases, $\hat{p}_n$ will converge to the probability of rolling a 1, $p = 1/6$. Figure 3.1 shows this convergence for 100,000 die rolls. The tendency of $\hat{p}_n$ to stabilize around p is described by the **Law of Large Numbers**.

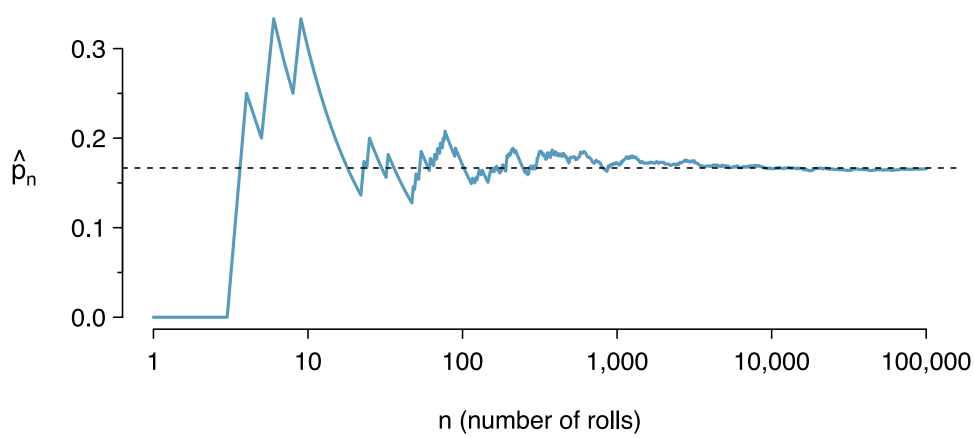


Figure 3.1: The fraction of die rolls that are 1 at each stage in a simulation. The proportion tends to get closer to the probability $1/6 \approx 0.167$ as the number of rolls increases.

LAW OF LARGE NUMBERS

As more observations are collected, the proportion $\hat{p}_n$ of occurrences with a particular outcome converges to the probability p of that outcome.

Occasionally the proportion will veer off from the probability and appear to defy the Law of Large Numbers, as $\hat{p}_n$ does many times in Figure 3.1. However, these deviations become smaller as the number of rolls increases.

Above we write p as the probability of rolling a 1. We can also write this probability as

$$P(\text{rolling a } 1)$$

As we become more comfortable with this notation, we will abbreviate it further. For instance, if it is clear that the process is "rolling a die", we could abbreviate $P(\text{rolling a } 1)$ as $P(1)$.

GUIDED PRACTICE 3.6

G

Random processes include rolling a die and flipping a coin. (a) Think of another random process. (b) Describe all the possible outcomes of that process. For instance, rolling a die is a random process with possible outcomes `1`, `2`, ..., `6`.[1]

What we think of as random processes are not necessarily random, but they may just be too difficult to understand exactly. The fourth example in the footnote solution to Guided Practice 3.6 suggests a roommate's behavior is a random process. However, even if a roommate's behavior is not truly random, modeling her behavior as a random process can still be useful.

3.1.3 Disjoint or mutually exclusive outcomes

Two outcomes are called **disjoint** or **mutually exclusive** if they cannot both happen. For instance, if we roll a die, the outcomes `1` and `2` are disjoint since they cannot both occur. On the other hand, the outcomes `1` and "rolling an odd number" are not disjoint since both occur if the outcome of the roll is a `1`. The terms *disjoint* and *mutually exclusive* are equivalent and interchangeable.

Calculating the probability of disjoint outcomes is easy. When rolling a die, the outcomes `1` and `2` are disjoint, and we compute the probability that one of these outcomes will occur by adding their separate probabilities:

$$P(\texttt{1} \text{ or } \texttt{2}) = P(\texttt{1}) + P(\texttt{2}) = 1/6 + 1/6 = 1/3$$

What about the probability of rolling a `1`, `2`, `3`, `4`, `5`, or `6`? Here again, all of the outcomes are disjoint so we add the probabilities:

$$\begin{aligned} &P(\texttt{1} \text{ or } \texttt{2} \text{ or } \texttt{3} \text{ or } \texttt{4} \text{ or } \texttt{5} \text{ or } \texttt{6}) \\ &\quad = P(\texttt{1}) + P(\texttt{2}) + P(\texttt{3}) + P(\texttt{4}) + P(\texttt{5}) + P(\texttt{6}) \\ &\quad = 1/6 + 1/6 + 1/6 + 1/6 + 1/6 + 1/6 = 1 \end{aligned}$$

The **Addition Rule** guarantees the accuracy of this approach when the outcomes are disjoint.

ADDITION RULE OF DISJOINT OUTCOMES

If A_1 and A_2 represent two disjoint outcomes, then the probability that one of them occurs is given by

$$P(A_1 \text{ or } A_2) = P(A_1) + P(A_2)$$

If there are many disjoint outcomes A_1, ..., A_k, then the probability that one of these outcomes will occur is

$$P(A_1) + P(A_2) + \cdots + P(A_k)$$

[1] Here are four examples. (i) Whether someone gets sick in the next month or not is an apparently random process with outcomes `sick` and `not`. (ii) We can *generate* a random process by randomly picking a person and measuring that person's height. The outcome of this process will be a positive number. (iii) Whether the stock market goes up or down next week is a seemingly random process with possible outcomes `up`, `down`, and `no_change`. Alternatively, we could have used the percent change in the stock market as a numerical outcome. (iv) Whether your roommate cleans her dishes tonight probably seems like a random process with possible outcomes `cleans_dishes` and `leaves_dishes`.

GUIDED PRACTICE 3.7

We are interested in the probability of rolling a `1`, `4`, or `5`. (a) Explain why the outcomes `1`, `4`, and `5` are disjoint. (b) Apply the Addition Rule for disjoint outcomes to determine P(`1` or `4` or `5`).[2]

GUIDED PRACTICE 3.8

In the **`loans`** data set in Chapter 2, the **`homeownership`** variable described whether the borrower rents, has a mortgage, or owns her property. Of the 10,000 borrowers, 3858 rented, 4789 had a mortgage, and 1353 owned their home.[3]

(a) Are the outcomes **`rent`**, **`mortgage`**, and **`own`** disjoint?

(b) Determine the proportion of loans with value **`mortgage`** and **`own`** separately.

(c) Use the Addition Rule for disjoint outcomes to compute the probability a randomly selected loan from the data set is for someone who has a mortgage or owns her home.

Data scientists rarely work with individual outcomes and instead consider *sets* or *collections* of outcomes. Let A represent the event where a die roll results in `1` or `2` and B represent the event that the die roll is a `4` or a `6`. We write A as the set of outcomes {`1`, `2`} and B = {`4`, `6`}. These sets are commonly called **events**. Because A and B have no elements in common, they are disjoint events. A and B are represented in Figure 3.2.

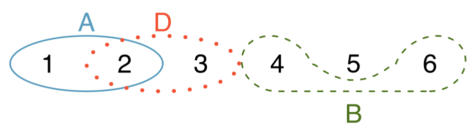


Figure 3.2: Three events, A, B, and D, consist of outcomes from rolling a die. A and B are disjoint since they do not have any outcomes in common.

The Addition Rule applies to both disjoint outcomes and disjoint events. The probability that one of the disjoint events A or B occurs is the sum of the separate probabilities:

$$P(A \text{ or } B) = P(A) + P(B) = 1/3 + 1/3 = 2/3$$

GUIDED PRACTICE 3.9

(a) Verify the probability of event A, $P(A)$, is 1/3 using the Addition Rule. (b) Do the same for event B.[4]

GUIDED PRACTICE 3.10

(a) Using Figure 3.2 as a reference, what outcomes are represented by event D? (b) Are events B and D disjoint? (c) Are events A and D disjoint?[5]

GUIDED PRACTICE 3.11

In Guided Practice 3.10, you confirmed B and D from Figure 3.2 are disjoint. Compute the probability that event B or event D occurs.[6]

---

[2](a) The random process is a die roll, and at most one of these outcomes can come up. This means they are disjoint outcomes. (b) $P(\texttt{1} \text{ or } \texttt{4} \text{ or } \texttt{5}) = P(\texttt{1}) + P(\texttt{4}) + P(\texttt{5}) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6} = \frac{1}{2}$

[3](a) Yes. Each loan is categorized in only one level of **`homeownership`**. (b) Mortgage: $\frac{4789}{10000} = 0.479$. Own: $\frac{1353}{10000} = 0.135$. (c) $P(\texttt{mortgage} \text{ or } \texttt{own}) = P(\texttt{mortgage}) + P(\texttt{own}) = 0.479 + 0.135 = 0.614$.

[4](a) $P(A) = P(\texttt{1} \text{ or } \texttt{2}) = P(\texttt{1}) + P(\texttt{2}) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3}$. (b) Similarly, $P(B) = 1/3$.

[5](a) Outcomes `2` and `3`. (b) Yes, events B and D are disjoint because they share no outcomes. (c) The events A and D share an outcome in common, `2`, and so are not disjoint.

[6]Since B and D are disjoint events, use the Addition Rule: $P(B \text{ or } D) = P(B) + P(D) = \frac{1}{3} + \frac{1}{3} = \frac{2}{3}$.

3.1.4 Probabilities when events are not disjoint

Let's consider calculations for two events that are not disjoint in the context of a regular deck of 52 cards, represented in Figure 3.3. If you are unfamiliar with the cards in a regular deck, please see the footnote.[7]

2♣	3♣	4♣	5♣	6♣	7♣	8♣	9♣	10♣	J♣	Q♣	K♣	A♣
2♢	3♢	4♢	5♢	6♢	7♢	8♢	9♢	10♢	J♢	Q♢	K♢	A♢
2♡	3♡	4♡	5♡	6♡	7♡	8♡	9♡	10♡	J♡	Q♡	K♡	A♡
2♠	3♠	4♠	5♠	6♠	7♠	8♠	9♠	10♠	J♠	Q♠	K♠	A♠

Figure 3.3: Representations of the 52 unique cards in a deck.

GUIDED PRACTICE 3.12

(a) What is the probability that a randomly selected card is a diamond? (b) What is the probability that a randomly selected card is a face card?[8]

Venn diagrams are useful when outcomes can be categorized as "in" or "out" for two or three variables, attributes, or random processes. The Venn diagram in Figure 3.4 uses a circle to represent diamonds and another to represent face cards. If a card is both a diamond and a face card, it falls into the intersection of the circles. If it is a diamond but not a face card, it will be in part of the left circle that is not in the right circle (and so on). The total number of cards that are diamonds is given by the total number of cards in the diamonds circle: $10 + 3 = 13$. The probabilities are also shown (e.g. $10/52 = 0.1923$).

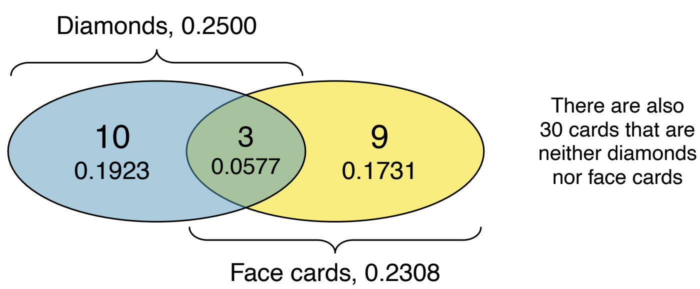


Figure 3.4: A Venn diagram for diamonds and face cards.

Let A represent the event that a randomly selected card is a diamond and B represent the event that it is a face card. How do we compute $P(A \text{ or } B)$? Events A and B are not disjoint – the cards $J\diamondsuit$, $Q\diamondsuit$, and $K\diamondsuit$ fall into both categories – so we cannot use the Addition Rule for disjoint events. Instead we use the Venn diagram. We start by adding the probabilities of the two events:

$$P(A) + P(B) = P(\diamondsuit) + P(\text{face card}) = 13/52 + 12/52$$

[7]The 52 cards are split into four **suits**: ♣ (club), ♢ (diamond), ♡ (heart), ♠ (spade). Each suit has its 13 cards labeled: `2`, `3`, ..., `10`, `J` (jack), `Q` (queen), `K` (king), and `A` (ace). Thus, each card is a unique combination of a suit and a label, e.g. 4♡ and J♣. The 12 cards represented by the jacks, queens, and kings are called **`face cards`**. The cards that are ♢ or ♡ are typically colored red while the other two suits are typically colored black.

[8](a) There are 52 cards and 13 diamonds. If the cards are thoroughly shuffled, each card has an equal chance of being drawn, so the probability that a randomly selected card is a diamond is $P(\diamondsuit) = \frac{13}{52} = 0.250$. (b) Likewise, there are 12 face cards, so $P(\text{face card}) = \frac{12}{52} = \frac{3}{13} = 0.231$.

However, the three cards that are in both events were counted twice, once in each probability. We must correct this double counting:

$$\begin{aligned} P(A \text{ or } B) &= P(\diamondsuit \text{ or face card}) \\ &= P(\diamondsuit) + P(\text{face card}) - P(\diamondsuit \text{ and face card}) \\ &= 13/52 + 12/52 - 3/52 \\ &= 22/52 = 11/26 \end{aligned}$$

This equation is an example of the **General Addition Rule**.

GENERAL ADDITION RULE

If A and B are any two events, disjoint or not, then the probability that at least one of them will occur is

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

where $P(A \text{ and } B)$ is the probability that both events occur.

TIP: "or" is inclusive

When we write "or" in statistics, we mean "and/or" unless we explicitly state otherwise. Thus, A or B occurs means A, B, or both A and B occur.

GUIDED PRACTICE 3.13

(a) If A and B are disjoint, describe why this implies $P(A \text{ and } B) = 0$. (b) Using part (a), verify that the General Addition Rule simplifies to the simpler Addition Rule for disjoint events if A and B are disjoint.[9]

GUIDED PRACTICE 3.14

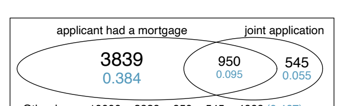
In the `loans` data set describing 10,000 loans, 1495 loans were from joint applications (e.g. a couple applied together), 4789 applicants had a mortgage, and 950 had both of these characteristics. Create a Venn diagram for this setup.[10]

GUIDED PRACTICE 3.15

(a) Use your Venn diagram from Guided Practice 3.14 to determine the probability a randomly drawn loan from the `loans` data set is from a joint application where the couple had a mortgage. (b) What is the probability that the loan had either of these attributes?[11]

[9] (a) If A and B are disjoint, A and B can never occur simultaneously. (b) If A and B are disjoint, then the last $P(A \text{ and } B)$ term of in the General Addition Rule formula is 0 (see part (a)) and we are left with the Addition Rule for disjoint events.

[10] Both the counts and corresponding probabilities (e.g. $3839/10000 = 0.384$) are shown. Notice that the number of loans represented in the left circle corresponds to $3839 + 950 = 4789$, and the number represented in the right circle is $950 + 545 = 1495$.



[11] (a) The solution is represented by the intersection of the two circles: 0.095. (b) This is the sum of the three disjoint probabilities shown in the circles: $0.384 + 0.095 + 0.055 = 0.534$ (off by 0.001 due to a rounding error).

3.1.5 Probability distributions

A **probability distribution** is a table of all disjoint outcomes and their associated probabilities. Figure 3.5 shows the probability distribution for the sum of two dice.

Dice sum	2	3	4	5	6	7	8	9	10	11	12
Probability	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Figure 3.5: Probability distribution for the sum of two dice.

RULES FOR PROBABILITY DISTRIBUTIONS

A probability distribution is a list of the possible outcomes with corresponding probabilities that satisfies three rules:

1. The outcomes listed must be disjoint.
2. Each probability must be between 0 and 1.
3. The probabilities must total 1.

GUIDED PRACTICE 3.16

G

Figure 3.6 suggests three distributions for household income in the United States. Only one is correct. Which one must it be? What is wrong with the other two?[12]

Income Range	\$0-25k	\$25k-50k	\$50k-100k	\$100k+
(a)	0.18	0.39	0.33	0.16
(b)	0.38	-0.27	0.52	0.37
(c)	0.28	0.27	0.29	0.16

Figure 3.6: Proposed distributions of US household incomes (Guided Practice 3.16).

Chapter 1 emphasized the importance of plotting data to provide quick summaries. Probability distributions can also be summarized in a bar plot. For instance, the distribution of US household incomes is shown in Figure 3.7 as a bar plot. The probability distribution for the sum of two dice is shown in Figure 3.5 and plotted in Figure 3.8.

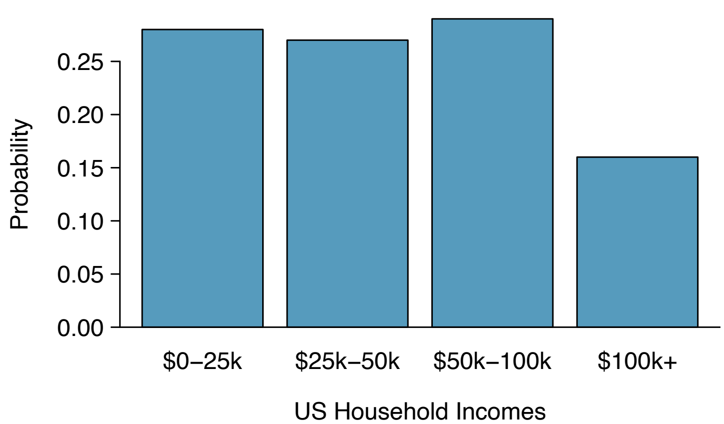


Figure 3.7: The probability distribution of US household income.

[12]The probabilities of (a) do not sum to 1. The second probability in (b) is negative. This leaves (c), which sure enough satisfies the requirements of a distribution. One of the three was said to be the actual distribution of US household incomes, so it must be (c).

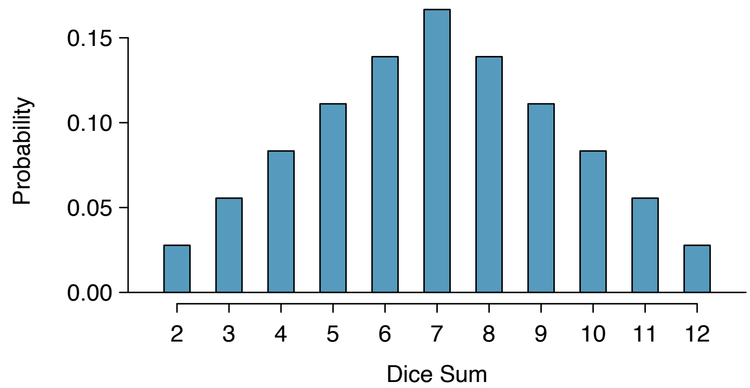


Figure 3.8: The probability distribution of the sum of two dice.

In these bar plots, the bar heights represent the probabilities of outcomes. If the outcomes are numerical and discrete, it is usually (visually) convenient to make a bar plot that resembles a histogram, as in the case of the sum of two dice. Another example of plotting the bars at their respective locations is shown in Figure 3.18 on page 115.

3.1.6 Complement of an event

Rolling a die produces a value in the set $\{1, 2, 3, 4, 5, 6\}$. This set of all possible outcomes is called the **sample space** (S) for rolling a die. We often use the sample space to examine the scenario where an event does not occur.

Let $D = \{2, 3\}$ represent the event that the outcome of a die roll is 2 or 3. Then the **complement** of D represents all outcomes in our sample space that are not in D, which is denoted by $D^c = \{1, 4, 5, 6\}$. That is, D^c is the set of all possible outcomes not already included in D. Figure 3.9 shows the relationship between D, D^c, and the sample space S.

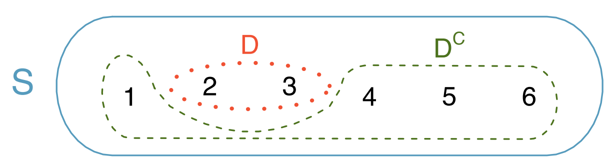


Figure 3.9: Event $D = \{2, 3\}$ and its complement, $D^c = \{1, 4, 5, 6\}$. S represents the sample space, which is the set of all possible outcomes.

G **GUIDED PRACTICE 3.17**

(a) Compute $P(D^c) = P(\text{rolling a } 1, 4, 5, \text{ or } 6)$. (b) What is $P(D) + P(D^c)$?[13]

G **GUIDED PRACTICE 3.18**

Events $A = \{1, 2\}$ and $B = \{4, 6\}$ are shown in Figure 3.2 on page 84. (a) Write out what A^c and B^c represent. (b) Compute $P(A^c)$ and $P(B^c)$. (c) Compute $P(A) + P(A^c)$ and $P(B) + P(B^c)$.[14]

[13] (a) The outcomes are disjoint and each has probability 1/6, so the total probability is $4/6 = 2/3$. (b) We can also see that $P(D) = \frac{1}{6} + \frac{1}{6} = 1/3$. Since D and D^c are disjoint, $P(D) + P(D^c) = 1$.

[14] Brief solutions: (a) $A^c = \{3, 4, 5, 6\}$ and $B^c = \{1, 2, 3, 5\}$. (b) Noting that each outcome is disjoint, add the individual outcome probabilities to get $P(A^c) = 2/3$ and $P(B^c) = 2/3$. (c) A and A^c are disjoint, and the same is true of B and B^c. Therefore, $P(A) + P(A^c) = 1$ and $P(B) + P(B^c) = 1$.

A complement of an event A is constructed to have two very important properties: (i) every possible outcome not in A is in A^c, and (ii) A and A^c are disjoint. Property (i) implies

$$P(A \text{ or } A^c) = 1$$

That is, if the outcome is not in A, it must be represented in A^c. We use the Addition Rule for disjoint events to apply Property (ii):

$$P(A \text{ or } A^c) = P(A) + P(A^c)$$

Combining the last two equations yields a very useful relationship between the probability of an event and its complement.

COMPLEMENT

The complement of event A is denoted A^c, and A^c represents all outcomes not in A. A and A^c are mathematically related:

$$P(A) + P(A^c) = 1, \quad \text{i.e.} \quad P(A) = 1 - P(A^c)$$

In simple examples, computing A or A^c is feasible in a few steps. However, using the complement can save a lot of time as problems grow in complexity.

GUIDED PRACTICE 3.19

G

Let A represent the event where we roll two dice and their total is less than 12. (a) What does the event A^c represent? (b) Determine $P(A^c)$ from Figure 3.5 on page 87. (c) Determine $P(A)$.[15]

GUIDED PRACTICE 3.20

Find the following probabilities for rolling two dice:[16]

G

(a) The sum of the dice is *not* 6.

(b) The sum is at least 4. That is, determine the probability of the event $B = \{4, 5, ..., 12\}$.

(c) The sum is no more than 10. That is, determine the probability of the event $D = \{2, 3, ..., 10\}$.

3.1.7 Independence

Just as variables and observations can be independent, random processes can be independent, too. Two processes are **independent** if knowing the outcome of one provides no useful information about the outcome of the other. For instance, flipping a coin and rolling a die are two independent processes – knowing the coin was heads does not help determine the outcome of a die roll. On the other hand, stock prices usually move up or down together, so they are not independent.

Example 3.5 provides a basic example of two independent processes: rolling two dice. We want to determine the probability that both will be 1. Suppose one of the dice is red and the other white. If the outcome of the red die is a 1, it provides no information about the outcome of the white die. We first encountered this same question in Example 3.5 (page 81), where we calculated the probability using the following reasoning: $1/6$ of the time the red die is a 1, and $1/6$ of *those* times the white die

[15](a) The complement of A: when the total is equal to 12. (b) $P(A^c) = 1/36$. (c) Use the probability of the complement from part (b), $P(A^c) = 1/36$, and the equation for the complement: $P(\text{less than } 12) = 1 - P(12) = 1 - 1/36 = 35/36$.

[16](a) First find $P(6) = 5/36$, then use the complement: $P(\text{not } 6) = 1 - P(6) = 31/36$.

(b) First find the complement, which requires much less effort: $P(2 \text{ or } 3) = 1/36 + 2/36 = 1/12$. Then calculate $P(B) = 1 - P(B^c) = 1 - 1/12 = 11/12$.

(c) As before, finding the complement is the clever way to determine $P(D)$. First find $P(D^c) = P(11 \text{ or } 12) = 2/36 + 1/36 = 1/12$. Then calculate $P(D) = 1 - P(D^c) = 11/12$.

will also be `1`. This is illustrated in Figure 3.10. Because the rolls are independent, the probabilities of the corresponding outcomes can be multiplied to get the final answer: $(1/6) \times (1/6) = 1/36$. This can be generalized to many independent processes.

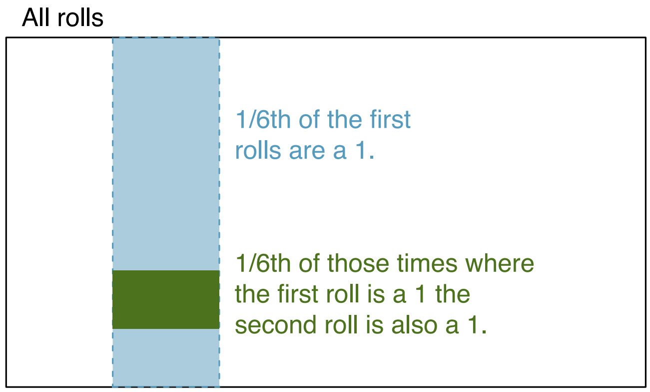


Figure 3.10: 1/6 of the time, the first roll is a `1`. Then 1/6 of *those* times, the second roll will also be a `1`.

EXAMPLE 3.21

What if there was also a blue die independent of the other two? What is the probability of rolling the three dice and getting all `1`s?

E

The same logic applies from Example 3.5. If 1/36 of the time the white and red dice are both `1`, then 1/6 of *those* times the blue die will also be `1`, so multiply:

$$\begin{aligned} P(white = \texttt{1} \text{ and } red = \texttt{1} \text{ and } blue = \texttt{1}) &= P(white = \texttt{1}) \times P(red = \texttt{1}) \times P(blue = \texttt{1}) \\ &= (1/6) \times (1/6) \times (1/6) = 1/216 \end{aligned}$$

Example 3.21 illustrates what is called the Multiplication Rule for independent processes.

MULTIPLICATION RULE FOR INDEPENDENT PROCESSES

If A and B represent events from two different and independent processes, then the probability that both A and B occur can be calculated as the product of their separate probabilities:

$$P(A \text{ and } B) = P(A) \times P(B)$$

Similarly, if there are k events A_1, ..., A_k from k independent processes, then the probability they all occur is

$$P(A_1) \times P(A_2) \times \cdots \times P(A_k)$$

GUIDED PRACTICE 3.22

G

About 9% of people are left-handed. Suppose 2 people are selected at random from the U.S. population. Because the sample size of 2 is very small relative to the population, it is reasonable to assume these two people are independent. (a) What is the probability that both are left-handed? (b) What is the probability that both are right-handed?[17]

[17](a) The probability the first person is left-handed is 0.09, which is the same for the second person. We apply the Multiplication Rule for independent processes to determine the probability that both will be left-handed: $0.09 \times 0.09 = 0.0081$.

(b) It is reasonable to assume the proportion of people who are ambidextrous (both right- and left-handed) is nearly 0, which results in $P(\text{right-handed}) = 1 - 0.09 = 0.91$. Using the same reasoning as in part (a), the probability that both will be right-handed is $0.91 \times 0.91 = 0.8281$.

GUIDED PRACTICE 3.23

Suppose 5 people are selected at random.[18]

(a) What is the probability that all are right-handed?

(b) What is the probability that all are left-handed?

(c) What is the probability that not all of the people are right-handed?

Suppose the variables `handedness` and `sex` are independent, i.e. knowing someone's `sex` provides no useful information about their `handedness` and vice-versa. Then we can compute whether a randomly selected person is right-handed and female[19] using the Multiplication Rule:

$$\begin{aligned} P(\text{right-handed and female}) &= P(\text{right-handed}) \times P(\text{female}) \\ &= 0.91 \times 0.50 = 0.455 \end{aligned}$$

GUIDED PRACTICE 3.24

Three people are selected at random.[20]

(a) What is the probability that the first person is male and right-handed?

(b) What is the probability that the first two people are male and right-handed?.

(c) What is the probability that the third person is female and left-handed?

(d) What is the probability that the first two people are male and right-handed and the third person is female and left-handed?

Sometimes we wonder if one outcome provides useful information about another outcome. The question we are asking is, are the occurrences of the two events independent? We say that two events A and B are independent if they satisfy $P(A \text{ and } B) = P(A) \times P(B)$.

EXAMPLE 3.25

If we shuffle up a deck of cards and draw one, is the event that the card is a heart independent of the event that the card is an ace?

The probability the card is a heart is $1/4$ and the probability that it is an ace is $1/13$. The probability the card is the ace of hearts is $1/52$. We check whether $P(A \text{ and } B) = P(A) \times P(B)$ is satisfied:

$$P(\heartsuit) \times P(\text{ace}) = \frac{1}{4} \times \frac{1}{13} = \frac{1}{52} = P(\heartsuit \text{ and ace})$$

Because the equation holds, the event that the card is a heart and the event that the card is an ace are independent events.

---

[18](a) The abbreviations `RH` and `LH` are used for right-handed and left-handed, respectively. Since each are independent, we apply the Multiplication Rule for independent processes:

$$\begin{aligned} P(\text{all five are } \texttt{RH}) &= P(\text{first} = \texttt{RH}, \text{second} = \texttt{RH}, \ldots, \text{fifth} = \texttt{RH}) \\ &= P(\text{first} = \texttt{RH}) \times P(\text{second} = \texttt{RH}) \times \cdots \times P(\text{fifth} = \texttt{RH}) \\ &= 0.91 \times 0.91 \times 0.91 \times 0.91 \times 0.91 = 0.624 \end{aligned}$$

(b) Using the same reasoning as in (a), $0.09 \times 0.09 \times 0.09 \times 0.09 \times 0.09 = 0.0000059$
(c) Use the complement, $P(\text{all five are } \texttt{RH})$, to answer this question:

$$P(\text{not all } \texttt{RH}) = 1 - P(\text{all } \texttt{RH}) = 1 - 0.624 = 0.376$$

[19]The actual proportion of the U.S. population that is `female` is about 50%, and so we use 0.5 for the probability of sampling a woman. However, this probability does differ in other countries.

[20]Brief answers are provided. (a) This can be written in probability notation as $P(\text{a randomly selected person is male and right-handed}) = 0.455$. (b) 0.207. (c) 0.045. (d) 0.0093.

Exercises

3.1 True or false. Determine if the statements below are true or false, and explain your reasoning.

(a) If a fair coin is tossed many times and the last eight tosses are all heads, then the chance that the next toss will be heads is somewhat less than 50%.

(b) Drawing a face card (jack, queen, or king) and drawing a red card from a full deck of playing cards are mutually exclusive events.

(c) Drawing a face card and drawing an ace from a full deck of playing cards are mutually exclusive events.

3.2 Roulette wheel. The game of roulette involves spinning a wheel with 38 slots: 18 red, 18 black, and 2 green. A ball is spun onto the wheel and will eventually land in a slot, where each slot has an equal chance of capturing the ball.

(a) You watch a roulette wheel spin 3 consecutive times and the ball lands on a red slot each time. What is the probability that the ball will land on a red slot on the next spin?

(b) You watch a roulette wheel spin 300 consecutive times and the ball lands on a red slot each time. What is the probability that the ball will land on a red slot on the next spin?

(c) Are you equally confident of your answers to parts (a) and (b)? Why or why not?

Photo by Håkan Dahlström
(http://flic.kr/p/93fEzp)

3.3 Four games, one winner. Below are four versions of the same game. Your archnemesis gets to pick the version of the game, and then you get to choose how many times to flip a coin: 10 times or 100 times. Identify how many coin flips you should choose for each version of the game. It costs \$1 to play each game. Explain your reasoning.

(a) If the proportion of heads is larger than 0.60, you win \$1.

(b) If the proportion of heads is larger than 0.40, you win \$1.

(c) If the proportion of heads is between 0.40 and 0.60, you win \$1.

(d) If the proportion of heads is smaller than 0.30, you win \$1.

3.4 Backgammon. Backgammon is a board game for two players in which the playing pieces are moved according to the roll of two dice. Players win by removing all of their pieces from the board, so it is usually good to roll high numbers. You are playing backgammon with a friend and you roll two 6s in your first roll and two 6s in your second roll. Your friend rolls two 3s in his first roll and again in his second row. Your friend claims that you are cheating, because rolling double 6s twice in a row is very unlikely. Using probability, show that your rolls were just as likely as his.

3.5 Coin flips. If you flip a fair coin 10 times, what is the probability of

(a) getting all tails?

(b) getting all heads?

(c) getting at least one tails?

3.6 Dice rolls. If you roll a pair of fair dice, what is the probability of

(a) getting a sum of 1?

(b) getting a sum of 5?

(c) getting a sum of 12?

3.7 Swing voters. A Pew Research survey asked 2,373 randomly sampled registered voters their political affiliation (Republican, Democrat, or Independent) and whether or not they identify as swing voters. 35% of respondents identified as Independent, 23% identified as swing voters, and 11% identified as both.[21]

(a) Are being Independent and being a swing voter disjoint, i.e. mutually exclusive?
(b) Draw a Venn diagram summarizing the variables and their associated probabilities.
(c) What percent of voters are Independent but not swing voters?
(d) What percent of voters are Independent or swing voters?
(e) What percent of voters are neither Independent nor swing voters?
(f) Is the event that someone is a swing voter independent of the event that someone is a political Independent?

3.8 Poverty and language. The American Community Survey is an ongoing survey that provides data every year to give communities the current information they need to plan investments and services. The 2010 American Community Survey estimates that 14.6% of Americans live below the poverty line, 20.7% speak a language other than English (foreign language) at home, and 4.2% fall into both categories.[22]

(a) Are living below the poverty line and speaking a foreign language at home disjoint?
(b) Draw a Venn diagram summarizing the variables and their associated probabilities.
(c) What percent of Americans live below the poverty line and only speak English at home?
(d) What percent of Americans live below the poverty line or speak a foreign language at home?
(e) What percent of Americans live above the poverty line and only speak English at home?
(f) Is the event that someone lives below the poverty line independent of the event that the person speaks a foreign language at home?

3.9 Disjoint vs. independent. In parts (a) and (b), identify whether the events are disjoint, independent, or neither (events cannot be both disjoint and independent).

(a) You and a randomly selected student from your class both earn A's in this course.
(b) You and your class study partner both earn A's in this course.
(c) If two events can occur at the same time, must they be dependent?

3.10 Guessing on an exam. In a multiple choice exam, there are 5 questions and 4 choices for each question (a, b, c, d). Nancy has not studied for the exam at all and decides to randomly guess the answers. What is the probability that:

(a) the first question she gets right is the 5^{th} question?
(b) she gets all of the questions right?
(c) she gets at least one question right?

[21] Pew Research Center, With Voters Focused on Economy, Obama Lead Narrows, data collected between April 4-15, 2012.

[22] U.S. Census Bureau, 2010 American Community Survey 1-Year Estimates, Characteristics of People by Language Spoken at Home.

3.11 Educational attainment of couples. The table below shows the distribution of education level attained by US residents by gender based on data collected in the 2010 American Community Survey.[23]

		Gender	
		Male	Female
Highest education attained	Less than 9th grade	0.07	0.13
	9th to 12th grade, no diploma	0.10	0.09
	HS graduate (or equivalent)	0.30	0.20
	Some college, no degree	0.22	0.24
	Associate's degree	0.06	0.08
	Bachelor's degree	0.16	0.17
	Graduate or professional degree	0.09	0.09
	Total	1.00	1.00

(a) What is the probability that a randomly chosen man has at least a Bachelor's degree?

(b) What is the probability that a randomly chosen woman has at least a Bachelor's degree?

(c) What is the probability that a man and a woman getting married both have at least a Bachelor's degree? Note any assumptions you must make to answer this question.

(d) If you made an assumption in part (c), do you think it was reasonable? If you didn't make an assumption, double check your earlier answer and then return to this part.

3.12 School absences. Data collected at elementary schools in DeKalb County, GA suggest that each year roughly 25% of students miss exactly one day of school, 15% miss 2 days, and 28% miss 3 or more days due to sickness.[24]

(a) What is the probability that a student chosen at random doesn't miss any days of school due to sickness this year?

(b) What is the probability that a student chosen at random misses no more than one day?

(c) What is the probability that a student chosen at random misses at least one day?

(d) If a parent has two kids at a DeKalb County elementary school, what is the probability that neither kid will miss any school? Note any assumption you must make to answer this question.

(e) If a parent has two kids at a DeKalb County elementary school, what is the probability that both kids will miss some school, i.e. at least one day? Note any assumption you make.

(f) If you made an assumption in part (d) or (e), do you think it was reasonable? If you didn't make any assumptions, double check your earlier answers.

[23] U.S. Census Bureau, 2010 American Community Survey 1-Year Estimates, Educational Attainment.

[24] S.S. Mizan et al. "Absence, Extended Absence, and Repeat Tardiness Related to Asthma Status among Elementary School Children". In: *Journal of Asthma* 48.3 (2011), pp. 228–234.

3.2 Conditional probability

There can be rich relationships between two or more variables that are useful to understand. For example a car insurance company will consider information about a person's driving history to assess the risk that they will be responsible for an accident. These types of relationships are the realm of conditional probabilities.

3.2.1 Exploring probabilities with a contingency table

The `photo_classify` data set represents a classifier a sample of 1822 photos from a photo sharing website. Data scientists have been working to improve a classifier for whether the photo is about fashion or not, and these 1822 photos represent a test for their classifier. Each photo gets two classifications: the first is called `mach_learn` and gives a classification from a machine learning (ML) system of either `pred_fashion` or `pred_not`. Each of these 1822 photos have also been classified carefully by a team of people, which we take to be the source of truth; this variable is called `truth` and takes values `fashion` and `not`. Figure 3.11 summarizes the results.

		truth		
		fashion	not	Total
mach_learn	pred_fashion	197	22	219
	pred_not	112	1491	1603
	Total	309	1513	1822

Figure 3.11: Contingency table summarizing the `photo_classify` data set.

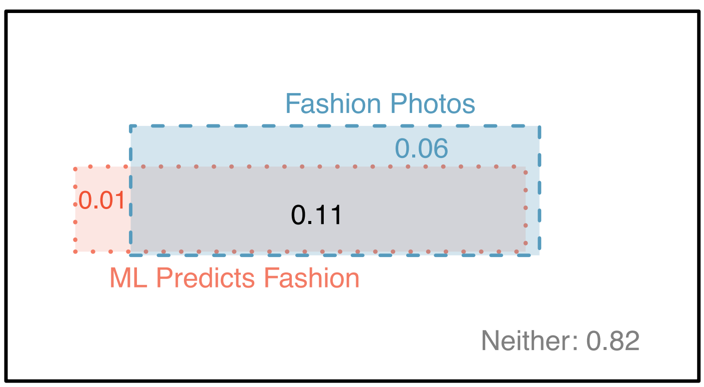


Figure 3.12: A Venn diagram using boxes for the `photo_classify` data set.

EXAMPLE 3.26

If a photo is actually about fashion, what is the chance the ML classifier correctly identified the photo as being about fashion?

We can estimate this probability using the data. Of the 309 fashion photos, the ML algorithm correctly classified 197 of the photos:

$$P(\texttt{mach\_learn} \text{ is } \texttt{pred\_fashion} \text{ given } \texttt{truth} \text{ is } \texttt{fashion}) = \frac{197}{309} = 0.638$$

EXAMPLE 3.27

We sample a photo from the data set and learn the ML algorithm predicted this photo was not about fashion. What is the probability that it was incorrect and the photo is about fashion?

E

If the ML classifier suggests a photo is not about fashion, then it comes from the second row in the data set. Of these 1603 photos, 112 were actually about fashion:

$$P(\texttt{truth} \text{ is } \texttt{fashion} \text{ given } \texttt{mach\_learn} \text{ is } \texttt{pred\_not}) = \frac{112}{1603} = 0.070$$

3.2.2 Marginal and joint probabilities

Figure 3.11 includes row and column totals for each variable separately in the `photo_classify` data set. These totals represent **marginal probabilities** for the sample, which are the probabilities based on a single variable without regard to any other variables. For instance, a probability based solely on the `mach_learn` variable is a marginal probability:

$$P(\texttt{mach\_learn} \text{ is } \texttt{pred\_fashion}) = \frac{219}{1822} = 0.12$$

A probability of outcomes for two or more variables or processes is called a **joint probability**:

$$P(\texttt{mach\_learn} \text{ is } \texttt{pred\_fashion} \text{ and } \texttt{truth} \text{ is } \texttt{fashion}) = \frac{197}{1822} = 0.11$$

It is common to substitute a comma for "and" in a joint probability, although using either the word "and" or a comma is acceptable:

$$P(\texttt{mach\_learn} \text{ is } \texttt{pred\_fashion}, \texttt{truth} \text{ is } \texttt{fashion})$$

means the same thing as

$$P(\texttt{mach\_learn} \text{ is } \texttt{pred\_fashion} \text{ and } \texttt{truth} \text{ is } \texttt{fashion})$$

MARGINAL AND JOINT PROBABILITIES

If a probability is based on a single variable, it is a *marginal probability.* The probability of outcomes for two or more variables or processes is called a *joint probability.*

We use **table proportions** to summarize joint probabilities for the `photo_classify` sample. These proportions are computed by dividing each count in Figure 3.11 by the table's total, 1822, to obtain the proportions in Figure 3.13. The joint probability distribution of the `mach_learn` and `truth` variables is shown in Figure 3.14.

	`truth: fashion`	`truth: not`	Total
`mach_learn: pred_fashion`	0.1081	0.0121	0.1202
`mach_learn: pred_not`	0.0615	0.8183	0.8798
Total	0.1696	0.8304	1.00

Figure 3.13: Probability table summarizing the `photo_classify` data set.

Joint outcome	Probability
`mach_learn` is `pred_fashion` and `truth` is `fashion`	0.1081
`mach_learn` is `pred_fashion` and `truth` is `not`	0.0121
`mach_learn` is `pred_not` and `truth` is `fashion`	0.0615
`mach_learn` is `pred_not` and `truth` is `not`	0.8183
Total	1.0000

Figure 3.14: Joint probability distribution for the `photo_classify` data set.

GUIDED PRACTICE 3.28

Verify Figure 3.14 represents a probability distribution: events are disjoint, all probabilities are non-negative, and the probabilities sum to 1.[25]

We can compute marginal probabilities using joint probabilities in simple cases. For example, the probability a randomly selected photo from the data set is about fashion is found by summing the outcomes where `truth` takes value `fashion`:

$$\begin{aligned}P(\underline{\texttt{truth}\text{ is }\texttt{fashion}}) &= P(\texttt{mach\_learn}\text{ is }\texttt{pred\_fashion}\text{ and }\underline{\texttt{truth}\text{ is }\texttt{fashion}}) \\ &\qquad + P(\texttt{mach\_learn}\text{ is }\texttt{pred\_not}\text{ and }\underline{\texttt{truth}\text{ is }\texttt{fashion}}) \\ &= 0.1081 + 0.0615 \\ &= 0.1696\end{aligned}$$

3.2.3 Defining conditional probability

The ML classifier predicts whether a photo is about fashion, even if it is not perfect. We would like to better understand how to use information from a variable like `mach_learn` to improve our probability estimation of a second variable, which in this example is `truth`.

The probability that a random photo from the data set is about fashion is about 0.17. If we knew the machine learning classifier predicted the photo was about fashion, could we get a better estimate of the probability the photo is actually about fashion? Absolutely. To do so, we limit our view to only those 219 cases where the ML classifier predicted that the photo was about fashion and look at the fraction where the photo was actually about fashion:

$$P(\texttt{truth}\text{ is }\texttt{fashion}\text{ given }\texttt{mach\_learn}\text{ is }\texttt{pred\_fashion}) = \frac{197}{219} = 0.900$$

We call this a **conditional probability** because we computed the probability under a condition: the ML classifier prediction said the photo was about fashion.

There are two parts to a conditional probability, the **outcome of interest** and the **condition**. It is useful to think of the condition as information we know to be true, and this information usually can be described as a known outcome or event. We generally separate the text inside our probability notation into the outcome of interest and the condition with a vertical bar:

$$\begin{aligned}&P(\texttt{truth}\text{ is }\texttt{fashion}\text{ given }\texttt{mach\_learn}\text{ is }\texttt{pred\_fashion}) \\ &= P(\texttt{truth}\text{ is }\texttt{fashion} \mid \texttt{mach\_learn}\text{ is }\texttt{pred\_fashion}) = \frac{197}{219} = 0.900\end{aligned}$$

The vertical bar "|" is read as *given*.

[25] Each of the four outcome combination are disjoint, all probabilities are indeed non-negative, and the sum of the probabilities is $0.1081 + 0.0121 + 0.0615 + 0.8183 = 1.00$.

In the last equation, we computed the probability a photo was about fashion based on the condition that the ML algorithm predicted it was about fashion as a fraction:

$$\begin{aligned} &P(\texttt{truth} \text{ is } \texttt{fashion} \mid \texttt{mach\_learn} \text{ is } \texttt{pred\_fashion}) \\ &\quad = \frac{\#\text{ cases where } \texttt{truth} \text{ is } \texttt{fashion} \text{ and } \texttt{mach\_learn} \text{ is } \texttt{pred\_fashion}}{\#\text{ cases where } \texttt{mach\_learn} \text{ is } \texttt{pred\_fashion}} \\ &\quad = \frac{197}{219} = 0.900 \end{aligned}$$

We considered only those cases that met the condition, `mach_learn` is `pred_fashion`, and then we computed the ratio of those cases that satisfied our outcome of interest, photo was actually about fashion.

Frequently, marginal and joint probabilities are provided instead of count data. For example, disease rates are commonly listed in percentages rather than in a count format. We would like to be able to compute conditional probabilities even when no counts are available, and we use the last equation as a template to understand this technique.

We considered only those cases that satisfied the condition, where the ML algorithm predicted fashion. Of these cases, the conditional probability was the fraction representing the outcome of interest, that the photo was about fashion. Suppose we were provided only the information in Figure 3.13, i.e. only probability data. Then if we took a sample of 1000 photos, we would anticipate about 12.0% or $0.120 \times 1000 = 120$ would be predicted to be about fashion (`mach_learn` is `pred_fashion`). Similarly, we would expect about 10.8% or $0.108 \times 1000 = 108$ to meet both the information criteria and represent our outcome of interest. Then the conditional probability can be computed as

$$\begin{aligned} &P(\texttt{truth} \text{ is } \texttt{fashion} \mid \texttt{mach\_learn} \text{ is } \texttt{pred\_fashion}) \\ &\quad = \frac{\#\ (\texttt{truth} \text{ is } \texttt{fashion} \text{ and } \texttt{mach\_learn} \text{ is } \texttt{pred\_fashion})}{\#\ (\texttt{mach\_learn} \text{ is } \texttt{pred\_fashion})} \\ &\quad = \frac{108}{120} = \frac{0.108}{0.120} = 0.90 \end{aligned}$$

Here we are examining exactly the fraction of two probabilities, 0.108 and 0.120, which we can write as

$$P(\texttt{truth} \text{ is } \texttt{fashion} \text{ and } \texttt{mach\_learn} \text{ is } \texttt{pred\_fashion}) \quad \text{and} \quad P(\texttt{mach\_learn} \text{ is } \texttt{pred\_fashion}).$$

The fraction of these probabilities is an example of the general formula for conditional probability.

CONDITIONAL PROBABILITY

The conditional probability of outcome A given condition B is computed as the following:

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$$

GUIDED PRACTICE 3.29

G

(a) Write out the following statement in conditional probability notation: "*The probability that the ML prediction was correct, if the photo was about fashion*". Here the condition is now based on the photo's `truth` status, not the ML algorithm.

(b) Determine the probability from part (a). Table 3.13 on page 96 may be helpful.[26]

[26](a) If the photo is about fashion and the ML algorithm prediction was correct, then the ML algorithm my have a value of `pred_fashion`:

$$P(\texttt{mach\_learn} \text{ is } \texttt{pred\_fashion} \mid \texttt{truth} \text{ is } \texttt{fashion})$$

(b) The equation for conditional probability indicates we should first find
$P(\texttt{mach\_learn} \text{ is } \texttt{pred\_fashion} \text{ and } \texttt{truth} \text{ is } \texttt{fashion}) = 0.1081$ and $P(\texttt{truth} \text{ is } \texttt{fashion}) = 0.1696$.
Then the ratio represents the conditional probability: $0.1081/0.1696 = 0.6374$.

GUIDED PRACTICE 3.30

(a) Determine the probability that the algorithm is incorrect if it is known the photo is about fashion.

(b) Using the answers from part (a) and Guided Practice 3.29(b), compute

$$P(\texttt{mach\_learn is pred\_fashion} \mid \texttt{truth is fashion}) + P(\texttt{mach\_learn is pred\_not} \mid \texttt{truth is fashion})$$

(c) Provide an intuitive argument to explain why the sum in (b) is 1.[27]

3.2.4 Smallpox in Boston, 1721

The `smallpox` data set provides a sample of 6,224 individuals from the year 1721 who were exposed to smallpox in Boston. Doctors at the time believed that inoculation, which involves exposing a person to the disease in a controlled form, could reduce the likelihood of death.

Each case represents one person with two variables: `inoculated` and `result`. The variable `inoculated` takes two levels: `yes` or `no`, indicating whether the person was inoculated or not. The variable `result` has outcomes `lived` or `died`. These data are summarized in Tables 3.15 and 3.16.

		inoculated		
		yes	no	Total
result	lived	238	5136	5374
	died	6	844	850
	Total	244	5980	6224

Figure 3.15: Contingency table for the `smallpox` data set.

		inoculated		
		yes	no	Total
result	lived	0.0382	0.8252	0.8634
	died	0.0010	0.1356	0.1366
	Total	0.0392	0.9608	1.0000

Figure 3.16: Table proportions for the `smallpox` data, computed by dividing each count by the table total, 6224.

GUIDED PRACTICE 3.31

Write out, in formal notation, the probability a randomly selected person who was not inoculated died from smallpox, and find this probability.[28]

GUIDED PRACTICE 3.32

Determine the probability that an inoculated person died from smallpox. How does this result compare with the result of Guided Practice 3.31?[29]

[27](a) This probability is $\frac{P(\texttt{mach\_learn is pred\_not, truth is fashion})}{P(\texttt{truth is fashion})} = \frac{0.0615}{0.1696} = 0.3626$. (b) The total equals 1. (c) Under the condition the photo is about fashion, the ML algorithm must have either predicted it was about fashion or predicted it was not about fashion. The complement still works for conditional probabilities, provided the probabilities are conditioned on the same information.

[28]$P(\texttt{result} = \texttt{died} \mid \texttt{inoculated} = \texttt{no}) = \frac{P(\texttt{result} = \texttt{died and inoculated} = \texttt{no})}{P(\texttt{inoculated} = \texttt{no})} = \frac{0.1356}{0.9608} = 0.1411$.

[29]$P(\texttt{result} = \texttt{died} \mid \texttt{inoculated} = \texttt{yes}) = \frac{P(\texttt{result} = \texttt{died and inoculated} = \texttt{yes})}{P(\texttt{inoculated} = \texttt{yes})} = \frac{0.0010}{0.0392} = 0.0255$ (if we avoided rounding errors, we'd get $6/244 = 0.0246$). The death rate for individuals who were inoculated is only about 1 in 40 while the death rate is about 1 in 7 for those who were not inoculated.

G

GUIDED PRACTICE 3.33

The people of Boston self-selected whether or not to be inoculated. (a) Is this study observational or was this an experiment? (b) Can we infer any causal connection using these data? (c) What are some potential confounding variables that might influence whether someone `lived` or `died` and also affect whether that person was inoculated?[30]

3.2.5 General multiplication rule

Section 3.1.7 introduced the Multiplication Rule for independent processes. Here we provide the **General Multiplication Rule** for events that might not be independent.

GENERAL MULTIPLICATION RULE

If A and B represent two outcomes or events, then

$$P(A \text{ and } B) = P(A|B) \times P(B)$$

It is useful to think of A as the outcome of interest and B as the condition.

This General Multiplication Rule is simply a rearrangement of the conditional probability equation.

E

EXAMPLE 3.34

Consider the `smallpox` data set. Suppose we are given only two pieces of information: 96.08% of residents were not inoculated, and 85.88% of the residents who were not inoculated ended up surviving. How could we compute the probability that a resident was not inoculated and lived?

We will compute our answer using the General Multiplication Rule and then verify it using Figure 3.16. We want to determine

$$P(\texttt{result} = \texttt{lived} \text{ and } \texttt{inoculated} = \texttt{no})$$

and we are given that

$$P(\texttt{result} = \texttt{lived} \mid \texttt{inoculated} = \texttt{no}) = 0.8588 \qquad P(\texttt{inoculated} = \texttt{no}) = 0.9608$$

Among the 96.08% of people who were not inoculated, 85.88% survived:

$$P(\texttt{result} = \texttt{lived} \text{ and } \texttt{inoculated} = \texttt{no}) = 0.8588 \times 0.9608 = 0.8251$$

This is equivalent to the General Multiplication Rule. We can confirm this probability in Figure 3.16 at the intersection of `no` and `lived` (with a small rounding error).

G

GUIDED PRACTICE 3.35

Use $P(\texttt{inoculated} = \texttt{yes}) = 0.0392$ and $P(\texttt{result} = \texttt{lived} \mid \texttt{inoculated} = \texttt{yes}) = 0.9754$ to determine the probability that a person was both inoculated and lived.[31]

G

GUIDED PRACTICE 3.36

If 97.54% of the inoculated people lived, what proportion of inoculated people must have died?[32]

---

[30]Brief answers: (a) Observational. (b) No, we cannot infer causation from this observational study. (c) Accessibility to the latest and best medical care. There are other valid answers for part (c).

[31]The answer is 0.0382, which can be verified using Figure 3.16.

[32]There were only two possible outcomes: `lived` or `died`. This means that 100% - 97.54% = 2.46% of the people who were inoculated died.

SUM OF CONDITIONAL PROBABILITIES

Let A_1, ..., A_k represent all the disjoint outcomes for a variable or process. Then if B is an event, possibly for another variable or process, we have:

$$P(A_1|B) + \cdots + P(A_k|B) = 1$$

The rule for complements also holds when an event and its complement are conditioned on the same information:

$$P(A|B) = 1 - P(A^c|B)$$

GUIDED PRACTICE 3.37

G Based on the probabilities computed above, does it appear that inoculation is effective at reducing the risk of death from smallpox?[33]

3.2.6 Independence considerations in conditional probability

If two events are independent, then knowing the outcome of one should provide no information about the other. We can show this is mathematically true using conditional probabilities.

GUIDED PRACTICE 3.38

Let X and Y represent the outcomes of rolling two dice.[34]

G
(a) What is the probability that the first die, X, is 1?
(b) What is the probability that both X and Y are 1?
(c) Use the formula for conditional probability to compute $P(Y = 1 \mid X = 1)$.
(d) What is $P(Y = 1)$? Is this different from the answer from part (c)? Explain.

We can show in Guided Practice 3.38(c) that the conditioning information has no influence by using the Multiplication Rule for independence processes:

$$\begin{aligned} P(Y = 1 \mid X = 1) &= \frac{P(Y = 1 \text{ and } X = 1)}{P(X = 1)} \\ &= \frac{P(Y = 1) \times P(X = 1)}{P(X = 1)} \\ &= P(Y = 1) \end{aligned}$$

GUIDED PRACTICE 3.39

G Ron is watching a roulette table in a casino and notices that the last five outcomes were `black`. He figures that the chances of getting `black` six times in a row is very small (about 1/64) and puts his paycheck on red. What is wrong with his reasoning?[35]

[33] The samples are large relative to the difference in death rates for the "inoculated" and "not inoculated" groups, so it seems there is an association between `inoculated` and `outcome`. However, as noted in the solution to Guided Practice 3.33, this is an observational study and we cannot be sure if there is a causal connection. (Further research has shown that inoculation is effective at reducing death rates.)

[34] Brief solutions: (a) 1/6. (b) 1/36. (c) $\frac{P(Y=1 \text{ and } X=1)}{P(X=1)} = \frac{1/36}{1/6} = 1/6$. (d) The probability is the same as in part (c): $P(Y = 1) = 1/6$. The probability that $Y = 1$ was unchanged by knowledge about X, which makes sense as X and Y are independent.

[35] He has forgotten that the next roulette spin is independent of the previous spins. Casinos do employ this practice, posting the last several outcomes of many betting games to trick unsuspecting gamblers into believing the odds are in their favor. This is called the **gambler's fallacy**.

3.2.7 Tree diagrams

Tree diagrams are a tool to organize outcomes and probabilities around the structure of the data. They are most useful when two or more processes occur in a sequence and each process is conditioned on its predecessors.

The `smallpox` data fit this description. We see the population as split by `inoculation`: `yes` and `no`. Following this split, survival rates were observed for each group. This structure is reflected in the **tree diagram** shown in Figure 3.17. The first branch for `inoculation` is said to be the **primary** branch while the other branches are **secondary**.

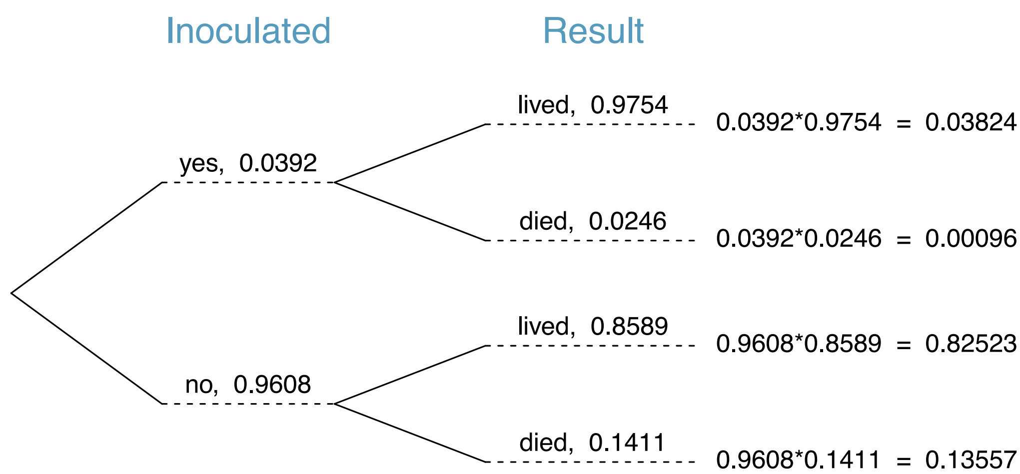


Figure 3.17: A tree diagram of the `smallpox` data set.

Tree diagrams are annotated with marginal and conditional probabilities, as shown in Figure 3.17. This tree diagram splits the smallpox data by `inoculation` into the `yes` and `no` groups with respective marginal probabilities 0.0392 and 0.9608. The secondary branches are conditioned on the first, so we assign conditional probabilities to these branches. For example, the top branch in Figure 3.17 is the probability that `result` = `lived` conditioned on the information that `inoculated` = `yes`. We may (and usually do) construct joint probabilities at the end of each branch in our tree by multiplying the numbers we come across as we move from left to right. These joint probabilities are computed using the General Multiplication Rule:

$$
\begin{aligned}
&P(\texttt{inoculated} = \texttt{yes} \text{ and } \texttt{result} = \texttt{lived}) \\
&\quad = P(\texttt{inoculated} = \texttt{yes}) \times P(\texttt{result} = \texttt{lived}|\texttt{inoculated} = \texttt{yes}) \\
&\quad = 0.0392 \times 0.9754 = 0.0382
\end{aligned}
$$

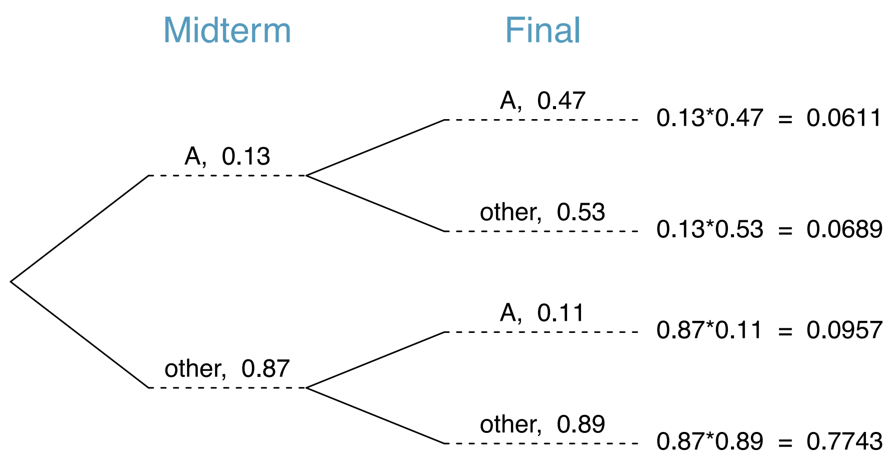
EXAMPLE 3.40

Consider the midterm and final for a statistics class. Suppose 13% of students earned an `A` on the midterm. Of those students who earned an `A` on the midterm, 47% received an `A` on the final, and 11% of the students who earned lower than an `A` on the midterm received an `A` on the final. You randomly pick up a final exam and notice the student received an `A`. What is the probability that this student earned an `A` on the midterm?

The end-goal is to find $P(\texttt{midterm} = \texttt{A} | \texttt{final} = \texttt{A})$. To calculate this conditional probability, we need the following probabilities:

$$P(\texttt{midterm} = \texttt{A} \text{ and } \texttt{final} = \texttt{A}) \qquad \text{and} \qquad P(\texttt{final} = \texttt{A})$$

However, this information is not provided, and it is not obvious how to calculate these probabilities. Since we aren't sure how to proceed, it is useful to organize the information into a tree diagram:



E

When constructing a tree diagram, variables provided with marginal probabilities are often used to create the tree's primary branches; in this case, the marginal probabilities are provided for midterm grades. The final grades, which correspond to the conditional probabilities provided, will be shown on the secondary branches.

With the tree diagram constructed, we may compute the required probabilities:

$$\begin{aligned}
&P(\texttt{midterm} = \texttt{A} \text{ and } \texttt{final} = \texttt{A}) = 0.0611 \\
&P(\underline{\texttt{final} = \texttt{A}}) \\
&\quad = P(\texttt{midterm} = \texttt{other} \text{ and } \underline{\texttt{final} = \texttt{A}}) + P(\texttt{midterm} = \texttt{A} \text{ and } \underline{\texttt{final} = \texttt{A}}) \\
&\quad = 0.0957 + 0.0611 = 0.1568
\end{aligned}$$

The marginal probability, $P(\texttt{final} = \texttt{A})$, was calculated by adding up all the joint probabilities on the right side of the tree that correspond to $\texttt{final} = \texttt{A}$. We may now finally take the ratio of the two probabilities:

$$\begin{aligned}
P(\texttt{midterm} = \texttt{A} | \texttt{final} = \texttt{A}) &= \frac{P(\texttt{midterm} = \texttt{A} \text{ and } \texttt{final} = \texttt{A})}{P(\texttt{final} = \texttt{A})} \\
&= \frac{0.0611}{0.1568} = 0.3897
\end{aligned}$$

The probability the student also earned an A on the midterm is about 0.39.

GUIDED PRACTICE 3.41

After an introductory statistics course, 78% of students can successfully construct tree diagrams. Of those who can construct tree diagrams, 97% passed, while only 57% of those students who could not construct tree diagrams passed. (a) Organize this information into a tree diagram. (b) What is the probability that a randomly selected student passed? (c) Compute the probability a student is able to construct a tree diagram if it is known that she passed.[36]

3.2.8 Bayes' Theorem

In many instances, we are given a conditional probability of the form

$$P(\text{statement about variable 1} \mid \text{statement about variable 2})$$

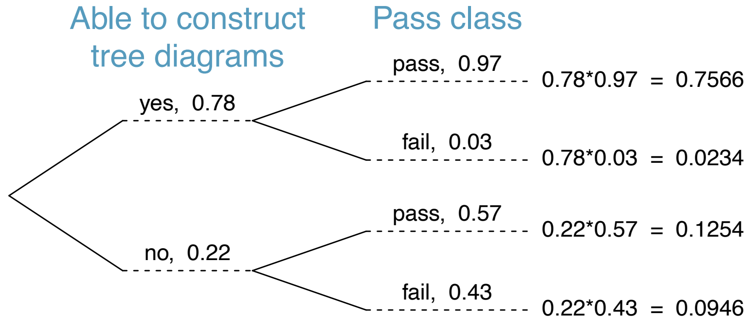
but we would really like to know the inverted conditional probability:

$$P(\text{statement about variable 2} \mid \text{statement about variable 1})$$

Tree diagrams can be used to find the second conditional probability when given the first. However, sometimes it is not possible to draw the scenario in a tree diagram. In these cases, we can apply a very useful and general formula: Bayes' Theorem.

We first take a critical look at an example of inverting conditional probabilities where we still apply a tree diagram.

[36](a) The tree diagram is shown to the right.
(b) Identify which two joint probabilities represent students who passed, and add them: $P(\text{passed}) = 0.7566 + 0.1254 = 0.8820$.
(c) $P(\text{construct tree diagram} \mid \text{passed}) = \frac{0.7566}{0.8820} = 0.8578$.



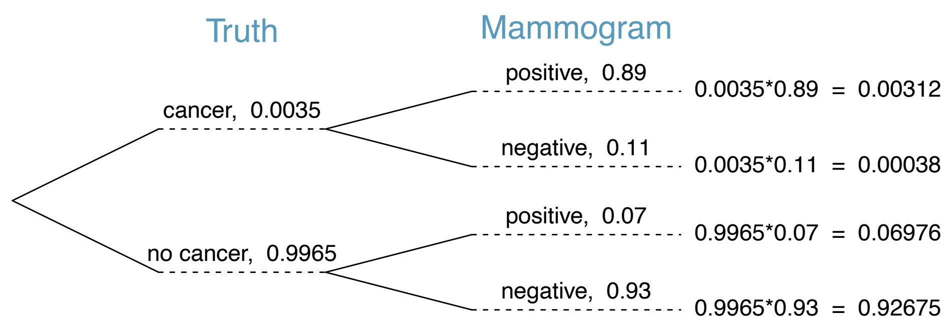
EXAMPLE 3.42

In Canada, about 0.35% of women over 40 will develop breast cancer in any given year. A common screening test for cancer is the mammogram, but this test is not perfect. In about 11% of patients with breast cancer, the test gives a **false negative**: it indicates a woman does not have breast cancer when she does have breast cancer. Similarly, the test gives a **false positive** in 7% of patients who do not have breast cancer: it indicates these patients have breast cancer when they actually do not. If we tested a random woman over 40 for breast cancer using a mammogram and the test came back positive – that is, the test suggested the patient has cancer – what is the probability that the patient actually has breast cancer?

Notice that we are given sufficient information to quickly compute the probability of testing positive if a woman has breast cancer ($1.00 - 0.11 = 0.89$). However, we seek the inverted probability of cancer given a positive test result. (Watch out for the non-intuitive medical language: a *positive* test result suggests the possible presence of cancer in a mammogram screening.) This inverted probability may be broken into two pieces:

$$P(\text{has BC} \mid \text{mammogram}^+) = \frac{P(\text{has BC and mammogram}^+)}{P(\text{mammogram}^+)}$$

where "has BC" is an abbreviation for the patient having breast cancer and "mammogram$^+$" means the mammogram screening was positive. We can construct a tree diagram for these probabilities:



The probability the patient has breast cancer and the mammogram is positive is

$$\begin{aligned} P(\text{has BC and mammogram}^+) &= P(\text{mammogram}^+ \mid \text{has BC})P(\text{has BC}) \\ &= 0.89 \times 0.0035 = 0.00312 \end{aligned}$$

The probability of a positive test result is the sum of the two corresponding scenarios:

$$\begin{aligned} P(\underline{\text{mammogram}^+}) &= P(\underline{\text{mammogram}^+} \text{ and has BC}) \\ &\qquad + P(\underline{\text{mammogram}^+} \text{ and no BC}) \\ &= P(\text{has BC})P(\text{mammogram}^+ \mid \text{has BC}) \\ &\qquad + P(\text{no BC})P(\text{mammogram}^+ \mid \text{no BC}) \\ &= 0.0035 \times 0.89 + 0.9965 \times 0.07 = 0.07288 \end{aligned}$$

Then if the mammogram screening is positive for a patient, the probability the patient has breast cancer is

$$\begin{aligned} P(\text{has BC} \mid \text{mammogram}^+) &= \frac{P(\text{has BC and mammogram}^+)}{P(\text{mammogram}^+)} \\ &= \frac{0.00312}{0.07288} \approx 0.0428 \end{aligned}$$

That is, even if a patient has a positive mammogram screening, there is still only a 4% chance that she has breast cancer.

Example 3.42 highlights why doctors often run more tests regardless of a first positive test result. When a medical condition is rare, a single positive test isn't generally definitive.

Consider again the last equation of Example 3.42. Using the tree diagram, we can see that the numerator (the top of the fraction) is equal to the following product:

$$P(\text{has BC and mammogram}^+) = P(\text{mammogram}^+ \mid \text{has BC})P(\text{has BC})$$

The denominator – the probability the screening was positive – is equal to the sum of probabilities for each positive screening scenario:

$$P(\underline{\text{mammogram}^+}) = P(\underline{\text{mammogram}^+} \text{ and no BC}) + P(\underline{\text{mammogram}^+} \text{ and has BC})$$

In the example, each of the probabilities on the right side was broken down into a product of a conditional probability and marginal probability using the tree diagram.

$$\begin{aligned} P(\text{mammogram}^+) &= P(\text{mammogram}^+ \text{ and no BC}) + P(\text{mammogram}^+ \text{ and has BC}) \\ &= P(\text{mammogram}^+ \mid \text{no BC})P(\text{no BC}) \\ &\qquad + P(\text{mammogram}^+ \mid \text{has BC})P(\text{has BC}) \end{aligned}$$

We can see an application of Bayes' Theorem by substituting the resulting probability expressions into the numerator and denominator of the original conditional probability.

$$\begin{aligned} &P(\text{has BC} \mid \text{mammogram}^+) \\ &\quad = \frac{P(\text{mammogram}^+ \mid \text{has BC})P(\text{has BC})}{P(\text{mammogram}^+ \mid \text{no BC})P(\text{no BC}) + P(\text{mammogram}^+ \mid \text{has BC})P(\text{has BC})} \end{aligned}$$

BAYES' THEOREM: INVERTING PROBABILITIES

Consider the following conditional probability for variable 1 and variable 2:

$$P(\text{outcome } A_1 \text{ of variable 1} \mid \text{outcome } B \text{ of variable 2})$$

Bayes' Theorem states that this conditional probability can be identified as the following fraction:

$$\frac{P(B|A_1)P(A_1)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots + P(B|A_k)P(A_k)}$$

where A_2, A_3, ..., and A_k represent all other possible outcomes of the first variable.

Bayes' Theorem is a generalization of what we have done using tree diagrams. The numerator identifies the probability of getting both A_1 and B. The denominator is the marginal probability of getting B. This bottom component of the fraction appears long and complicated since we have to add up probabilities from all of the different ways to get B. We always completed this step when using tree diagrams. However, we usually did it in a separate step so it didn't seem as complex.
To apply Bayes' Theorem correctly, there are two preparatory steps:

(1) First identify the marginal probabilities of each possible outcome of the first variable: $P(A_1)$, $P(A_2)$, ..., $P(A_k)$.

(2) Then identify the probability of the outcome B, conditioned on each possible scenario for the first variable: $P(B|A_1)$, $P(B|A_2)$, ..., $P(B|A_k)$.

Once each of these probabilities are identified, they can be applied directly within the formula. Bayes' Theorem tends to be a good option when there are so many scenarios that drawing a tree diagram would be complex.

GUIDED PRACTICE 3.43

Jose visits campus every Thursday evening. However, some days the parking garage is full, often due to college events. There are academic events on 35% of evenings, sporting events on 20% of evenings, and no events on 45% of evenings. When there is an academic event, the garage fills up about 25% of the time, and it fills up 70% of evenings with sporting events. On evenings when there are no events, it only fills up about 5% of the time. If Jose comes to campus and finds the garage full, what is the probability that there is a sporting event? Use a tree diagram to solve this problem.[37]

EXAMPLE 3.44

Here we solve the same problem presented in Guided Practice 3.43, except this time we use Bayes' Theorem.

The outcome of interest is whether there is a sporting event (call this A_1), and the condition is that the lot is full (B). Let A_2 represent an academic event and A_3 represent there being no event on campus. Then the given probabilities can be written as

$$P(A_1) = 0.2 \qquad P(A_2) = 0.35 \qquad P(A_3) = 0.45$$
$$P(B|A_1) = 0.7 \qquad P(B|A_2) = 0.25 \qquad P(B|A_3) = 0.05$$

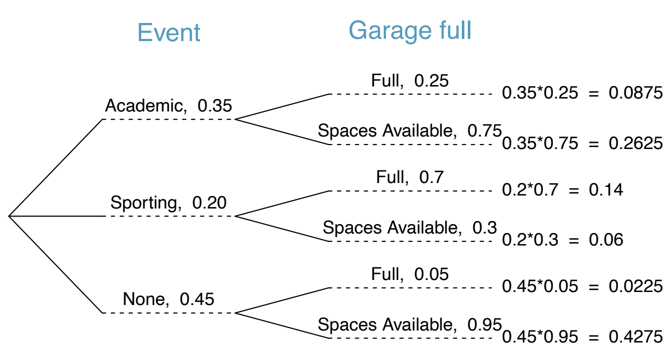
Bayes' Theorem can be used to compute the probability of a sporting event (A_1) under the condition that the parking lot is full (B):

$$\begin{aligned} P(A_1|B) &= \frac{P(B|A_1)P(A_1)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3)} \\ &= \frac{(0.7)(0.2)}{(0.7)(0.2) + (0.25)(0.35) + (0.05)(0.45)} \\ &= 0.56 \end{aligned}$$

Based on the information that the garage is full, there is a 56% probability that a sporting event is being held on campus that evening.

---

[37]The tree diagram, with three primary branches, is shown to the right. Next, we identify two probabilities from the tree diagram. (1) The probability that there is a sporting event and the garage is full: 0.14. (2) The probability the garage is full: $0.0875 + 0.14 + 0.0225 = 0.25$. Then the solution is the ratio of these probabilities: $\frac{0.14}{0.25} = 0.56$. If the garage is full, there is a 56% probability that there is a sporting event.



GUIDED PRACTICE 3.45

Use the information in the previous exercise and example to verify the probability that there is an academic event conditioned on the parking lot being full is 0.35.[38]

GUIDED PRACTICE 3.46

In Guided Practice 3.43 and 3.45, you found that if the parking lot is full, the probability there is a sporting event is 0.56 and the probability there is an academic event is 0.35. Using this information, compute $P(\text{no event} \mid \text{the lot is full})$.[39]

The last several exercises offered a way to update our belief about whether there is a sporting event, academic event, or no event going on at the school based on the information that the parking lot was full. This strategy of *updating beliefs* using Bayes' Theorem is actually the foundation of an entire section of statistics called **Bayesian statistics**. While Bayesian statistics is very important and useful, we will not have time to cover much more of it in this book.

[38]Short answer:

$$\begin{aligned}
P(A_2|B) &= \frac{P(B|A_2)P(A_2)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3)} \\
&= \frac{(0.25)(0.35)}{(0.7)(0.2) + (0.25)(0.35) + (0.05)(0.45)} \\
&= 0.35
\end{aligned}$$

[39]Each probability is conditioned on the same information that the garage is full, so the complement may be used: $1.00 - 0.56 - 0.35 = 0.09$.

Exercises

3.13 Joint and conditional probabilities. P(A) = 0.3, P(B) = 0.7

(a) Can you compute P(A and B) if you only know P(A) and P(B)?

(b) Assuming that events A and B arise from independent random processes,

  i. what is P(A and B)?
  ii. what is P(A or B)?
  iii. what is P(A|B)?

(c) If we are given that P(A and B) = 0.1, are the random variables giving rise to events A and B independent?

(d) If we are given that P(A and B) = 0.1, what is P(A|B)?

3.14 PB & J. Suppose 80% of people like peanut butter, 89% like jelly, and 78% like both. Given that a randomly sampled person likes peanut butter, what's the probability that he also likes jelly?

3.15 Global warming. A Pew Research poll asked 1,306 Americans "From what you've read and heard, is there solid evidence that the average temperature on earth has been getting warmer over the past few decades, or not?". The table below shows the distribution of responses by party and ideology, where the counts have been replaced with relative frequencies.[40]

		Response			
		Earth is warming	Not warming	Don't Know Refuse	Total
Party and Ideology	Conservative Republican	0.11	0.20	0.02	0.33
	Mod/Lib Republican	0.06	0.06	0.01	0.13
	Mod/Cons Democrat	0.25	0.07	0.02	0.34
	Liberal Democrat	0.18	0.01	0.01	0.20
	Total	0.60	0.34	0.06	1.00

(a) Are believing that the earth is warming and being a liberal Democrat mutually exclusive?

(b) What is the probability that a randomly chosen respondent believes the earth is warming or is a liberal Democrat?

(c) What is the probability that a randomly chosen respondent believes the earth is warming given that he is a liberal Democrat?

(d) What is the probability that a randomly chosen respondent believes the earth is warming given that he is a conservative Republican?

(e) Does it appear that whether or not a respondent believes the earth is warming is independent of their party and ideology? Explain your reasoning.

(f) What is the probability that a randomly chosen respondent is a moderate/liberal Republican given that he does not believe that the earth is warming?

[40] Pew Research Center, Majority of Republicans No Longer See Evidence of Global Warming, data collected on October 27, 2010.

3.16 Health coverage, relative frequencies. The Behavioral Risk Factor Surveillance System (BRFSS) is an annual telephone survey designed to identify risk factors in the adult population and report emerging health trends. The following table displays the distribution of health status of respondents to this survey (excellent, very good, good, fair, poor) and whether or not they have health insurance.

		Health Status					
		Excellent	Very good	Good	Fair	Poor	Total
Health	No	0.0230	0.0364	0.0427	0.0192	0.0050	0.1262
Coverage	Yes	0.2099	0.3123	0.2410	0.0817	0.0289	0.8738
	Total	0.2329	0.3486	0.2838	0.1009	0.0338	1.0000

(a) Are being in excellent health and having health coverage mutually exclusive?

(b) What is the probability that a randomly chosen individual has excellent health?

(c) What is the probability that a randomly chosen individual has excellent health given that he has health coverage?

(d) What is the probability that a randomly chosen individual has excellent health given that he doesn't have health coverage?

(e) Do having excellent health and having health coverage appear to be independent?

3.17 Burger preferences. A 2010 SurveyUSA poll asked 500 Los Angeles residents, "What is the best hamburger place in Southern California? Five Guys Burgers? In-N-Out Burger? Fat Burger? Tommy's Hamburgers? Umami Burger? Or somewhere else?" The distribution of responses by gender is shown below.[41]

		Gender		
		Male	Female	Total
	Five Guys Burgers	5	6	11
	In-N-Out Burger	162	181	343
Best	Fat Burger	10	12	22
hamburger	Tommy's Hamburgers	27	27	54
place	Umami Burger	5	1	6
	Other	26	20	46
	Not Sure	13	5	18
	Total	248	252	500

(a) Are being female and liking Five Guys Burgers mutually exclusive?

(b) What is the probability that a randomly chosen male likes In-N-Out the best?

(c) What is the probability that a randomly chosen female likes In-N-Out the best?

(d) What is the probability that a man and a woman who are dating both like In-N-Out the best? Note any assumption you make and evaluate whether you think that assumption is reasonable.

(e) What is the probability that a randomly chosen person likes Umami best or that person is female?

[41] SurveyUSA, Results of SurveyUSA News Poll #17718, data collected on December 2, 2010.

3.18 Assortative mating. Assortative mating is a nonrandom mating pattern where individuals with similar genotypes and/or phenotypes mate with one another more frequently than what would be expected under a random mating pattern. Researchers studying this topic collected data on eye colors of 204 Scandinavian men and their female partners. The table below summarizes the results. For simplicity, we only include heterosexual relationships in this exercise.[42]

		Partner (female)			
		Blue	Brown	Green	Total
Self (male)	Blue	78	23	13	114
	Brown	19	23	12	54
	Green	11	9	16	36
	Total	108	55	41	204

(a) What is the probability that a randomly chosen male respondent or his partner has blue eyes?

(b) What is the probability that a randomly chosen male respondent with blue eyes has a partner with blue eyes?

(c) What is the probability that a randomly chosen male respondent with brown eyes has a partner with blue eyes? What about the probability of a randomly chosen male respondent with green eyes having a partner with blue eyes?

(d) Does it appear that the eye colors of male respondents and their partners are independent? Explain your reasoning.

3.19 Drawing box plots. After an introductory statistics course, 80% of students can successfully construct box plots. Of those who can construct box plots, 86% passed, while only 65% of those students who could not construct box plots passed.

(a) Construct a tree diagram of this scenario.

(b) Calculate the probability that a student is able to construct a box plot if it is known that he passed.

3.20 Predisposition for thrombosis. A genetic test is used to determine if people have a predisposition for *thrombosis*, which is the formation of a blood clot inside a blood vessel that obstructs the flow of blood through the circulatory system. It is believed that 3% of people actually have this predisposition. The genetic test is 99% accurate if a person actually has the predisposition, meaning that the probability of a positive test result when a person actually has the predisposition is 0.99. The test is 98% accurate if a person does not have the predisposition. What is the probability that a randomly selected person who tests positive for the predisposition by the test actually has the predisposition?

3.21 It's never lupus. Lupus is a medical phenomenon where antibodies that are supposed to attack foreign cells to prevent infections instead see plasma proteins as foreign bodies, leading to a high risk of blood clotting. It is believed that 2% of the population suffer from this disease. The test is 98% accurate if a person actually has the disease. The test is 74% accurate if a person does not have the disease. There is a line from the Fox television show *House* that is often used after a patient tests positive for lupus: "It's never lupus." Do you think there is truth to this statement? Use appropriate probabilities to support your answer.

3.22 Exit poll. Edison Research gathered exit poll results from several sources for the Wisconsin recall election of Scott Walker. They found that 53% of the respondents voted in favor of Scott Walker. Additionally, they estimated that of those who did vote in favor for Scott Walker, 37% had a college degree, while 44% of those who voted against Scott Walker had a college degree. Suppose we randomly sampled a person who participated in the exit poll and found that he had a college degree. What is the probability that he voted in favor of Scott Walker?[43]

[42] B. Laeng et al. "Why do blue-eyed men prefer women with the same eye color?" In: *Behavioral Ecology and Sociobiology* 61.3 (2007), pp. 371–384.

[43] New York Times, Wisconsin recall exit polls.

3.3 Sampling from a small population

When we sample observations from a population, usually we're only sampling a small fraction of the possible individuals or cases. However, sometimes our sample size is large enough or the population is small enough that we sample more than 10% of a population[44] *without replacement* (meaning we do not have a chance of sampling the same cases twice). Sampling such a notable fraction of a population can be important for how we analyze the sample.

E

EXAMPLE 3.47

Professors sometimes select a student at random to answer a question. If each student has an equal chance of being selected and there are 15 people in your class, what is the chance that she will pick you for the next question?

If there are 15 people to ask and none are skipping class, then the probability is 1/15, or about 0.067.

E

EXAMPLE 3.48

If the professor asks 3 questions, what is the probability that you will not be selected? Assume that she will not pick the same person twice in a given lecture.

For the first question, she will pick someone else with probability 14/15. When she asks the second question, she only has 14 people who have not yet been asked. Thus, if you were not picked on the first question, the probability you are again not picked is 13/14. Similarly, the probability you are again not picked on the third question is 12/13, and the probability of not being picked for any of the three questions is

$$
\begin{aligned}
&P(\text{not picked in 3 questions}) \\
&\quad = P(\texttt{Q1} = \texttt{not\_picked}, \texttt{Q2} = \texttt{not\_picked}, \texttt{Q3} = \texttt{not\_picked}.) \\
&\quad = \frac{14}{15} \times \frac{13}{14} \times \frac{12}{13} = \frac{12}{15} = 0.80
\end{aligned}
$$

G

GUIDED PRACTICE 3.49

What rule permitted us to multiply the probabilities in Example 3.48?[45]

[44] The 10% guideline is a rule of thumb cutoff for when these considerations become more important.

[45] The three probabilities we computed were actually one marginal probability, $P(\texttt{Q1}{=}\texttt{not\_picked})$, and two conditional probabilities:

$$
\begin{aligned}
&P(\texttt{Q2} = \texttt{not\_picked} \mid \texttt{Q1} = \texttt{not\_picked}) \\
&P(\texttt{Q3} = \texttt{not\_picked} \mid \texttt{Q1} = \texttt{not\_picked}, \texttt{Q2} = \texttt{not\_picked})
\end{aligned}
$$

Using the General Multiplication Rule, the product of these three probabilities is the probability of not being picked in 3 questions.

EXAMPLE 3.50

Suppose the professor randomly picks without regard to who she already selected, i.e. students can be picked more than once. What is the probability that you will not be picked for any of the three questions?

Each pick is independent, and the probability of not being picked for any individual question is 14/15. Thus, we can use the Multiplication Rule for independent processes.

E

$$\begin{aligned} &P(\text{not picked in 3 questions}) \\ &\quad = P(\texttt{Q1} = \texttt{not\_picked}, \texttt{Q2} = \texttt{not\_picked}, \texttt{Q3} = \texttt{not\_picked.}) \\ &\quad = \frac{14}{15} \times \frac{14}{15} \times \frac{14}{15} = 0.813 \end{aligned}$$

You have a slightly higher chance of not being picked compared to when she picked a new person for each question. However, you now may be picked more than once.

GUIDED PRACTICE 3.51

G

Under the setup of Example 3.50, what is the probability of being picked to answer all three questions?[46]

If we sample from a small population **without replacement**, we no longer have independence between our observations. In Example 3.48, the probability of not being picked for the second question was conditioned on the event that you were not picked for the first question. In Example 3.50, the professor sampled her students **with replacement**: she repeatedly sampled the entire class without regard to who she already picked.

GUIDED PRACTICE 3.52

G

Your department is holding a raffle. They sell 30 tickets and offer seven prizes. (a) They place the tickets in a hat and draw one for each prize. The tickets are sampled without replacement, i.e. the selected tickets are not placed back in the hat. What is the probability of winning a prize if you buy one ticket? (b) What if the tickets are sampled with replacement?[47]

GUIDED PRACTICE 3.53

G

Compare your answers in Guided Practice 3.52. How much influence does the sampling method have on your chances of winning a prize?[48]

Had we repeated Guided Practice 3.52 with 300 tickets instead of 30, we would have found something interesting: the results would be nearly identical. The probability would be 0.0233 without replacement and 0.0231 with replacement. When the sample size is only a small fraction of the population (under 10%), observations are nearly independent even when sampling without replacement.

---

[46] $P(\text{being picked to answer all three questions}) = \left(\frac{1}{15}\right)^3 = 0.00030$.

[47] (a) First determine the probability of not winning. The tickets are sampled without replacement, which means the probability you do not win on the first draw is 29/30, 28/29 for the second, ..., and 23/24 for the seventh. The probability you win no prize is the product of these separate probabilities: 23/30. That is, the probability of winning a prize is $1 - 23/30 = 7/30 = 0.233$. (b) When the tickets are sampled with replacement, there are seven independent draws. Again we first find the probability of not winning a prize: $(29/30)^7 = 0.789$. Thus, the probability of winning (at least) one prize when drawing with replacement is 0.211.

[48] There is about a 10% larger chance of winning a prize when using sampling without replacement. However, at most one prize may be won under this sampling procedure.

Exercises

3.23 Marbles in an urn. Imagine you have an urn containing 5 red, 3 blue, and 2 orange marbles in it.

(a) What is the probability that the first marble you draw is blue?

(b) Suppose you drew a blue marble in the first draw. If drawing with replacement, what is the probability of drawing a blue marble in the second draw?

(c) Suppose you instead drew an orange marble in the first draw. If drawing with replacement, what is the probability of drawing a blue marble in the second draw?

(d) If drawing with replacement, what is the probability of drawing two blue marbles in a row?

(e) When drawing with replacement, are the draws independent? Explain.

3.24 Socks in a drawer. In your sock drawer you have 4 blue, 5 gray, and 3 black socks. Half asleep one morning you grab 2 socks at random and put them on. Find the probability you end up wearing

(a) 2 blue socks

(b) no gray socks

(c) at least 1 black sock

(d) a green sock

(e) matching socks

3.25 Chips in a bag. Imagine you have a bag containing 5 red, 3 blue, and 2 orange chips.

(a) Suppose you draw a chip and it is blue. If drawing without replacement, what is the probability the next is also blue?

(b) Suppose you draw a chip and it is orange, and then you draw a second chip without replacement. What is the probability this second chip is blue?

(c) If drawing without replacement, what is the probability of drawing two blue chips in a row?

(d) When drawing without replacement, are the draws independent? Explain.

3.26 Books on a bookshelf. The table below shows the distribution of books on a bookcase based on whether they are nonfiction or fiction and hardcover or paperback.

		Format		
		Hardcover	Paperback	Total
Type	Fiction	13	59	72
	Nonfiction	15	8	23
	Total	28	67	95

(a) Find the probability of drawing a hardcover book first then a paperback fiction book second when drawing without replacement.

(b) Determine the probability of drawing a fiction book first and then a hardcover book second, when drawing without replacement.

(c) Calculate the probability of the scenario in part (b), except this time complete the calculations under the scenario where the first book is placed back on the bookcase before randomly drawing the second book.

(d) The final answers to parts (b) and (c) are very similar. Explain why this is the case.

3.27 Student outfits. In a classroom with 24 students, 7 students are wearing jeans, 4 are wearing shorts, 8 are wearing skirts, and the rest are wearing leggings. If we randomly select 3 students without replacement, what is the probability that one of the selected students is wearing leggings and the other two are wearing jeans? Note that these are mutually exclusive clothing options.

3.28 The birthday problem. Suppose we pick three people at random. For each of the following questions, ignore the special case where someone might be born on February 29th, and assume that births are evenly distributed throughout the year.

(a) What is the probability that the first two people share a birthday?

(b) What is the probability that at least two people share a birthday?

3.4 Random variables

It's often useful to model a process using what's called a **random variable**. Such a model allows us to apply a mathematical framework and statistical principles for better understanding and predicting outcomes in the real world.

E

EXAMPLE 3.54

Two books are assigned for a statistics class: a textbook and its corresponding study guide. The university bookstore determined 20% of enrolled students do not buy either book, 55% buy the textbook only, and 25% buy both books, and these percentages are relatively constant from one term to another. If there are 100 students enrolled, how many books should the bookstore expect to sell to this class?

Around 20 students will not buy either book (0 books total), about 55 will buy one book (55 books total), and approximately 25 will buy two books (totaling 50 books for these 25 students). The bookstore should expect to sell about 105 books for this class.

G

GUIDED PRACTICE 3.55

Would you be surprised if the bookstore sold slightly more or less than 105 books?[49]

E

EXAMPLE 3.56

The textbook costs \$137 and the study guide \$33. How much revenue should the bookstore expect from this class of 100 students?

About 55 students will just buy a textbook, providing revenue of

$$\$137 \times 55 = \$7,535$$

The roughly 25 students who buy both the textbook and the study guide would pay a total of

$$(\$137 + \$33) \times 25 = \$170 \times 25 = \$4,250$$

Thus, the bookstore should expect to generate about $\$7,535 + \$4,250 = \$11,785$ from these 100 students for this one class. However, there might be some *sampling variability* so the actual amount may differ by a little bit.

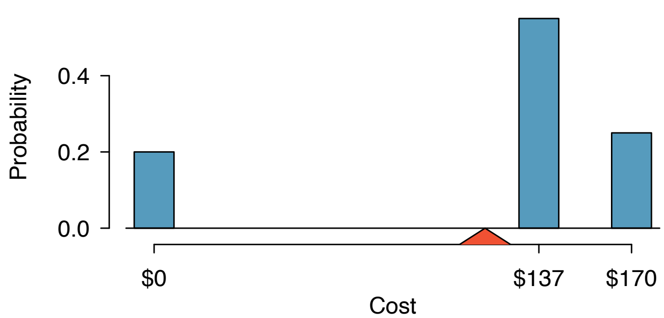


Figure 3.18: Probability distribution for the bookstore's revenue from one student. The triangle represents the average revenue per student.

[49]If they sell a little more or a little less, this should not be a surprise. Hopefully Chapter 1 helped make clear that there is natural variability in observed data. For example, if we would flip a coin 100 times, it will not usually come up heads exactly half the time, but it will probably be close.

EXAMPLE 3.57

What is the average revenue per student for this course?

The expected total revenue is \$11,785, and there are 100 students. Therefore the expected revenue per student is $\$11,785/100 = \$117.85$.

3.4.1 Expectation

We call a variable or process with a numerical outcome a **random variable**, and we usually represent this random variable with a capital letter such as X, Y, or Z. The amount of money a single student will spend on her statistics books is a random variable, and we represent it by X.

RANDOM VARIABLE

A random process or variable with a numerical outcome.

The possible outcomes of X are labeled with a corresponding lower case letter x and subscripts. For example, we write $x_1 = \$0$, $x_2 = \$137$, and $x_3 = \$170$, which occur with probabilities 0.20, 0.55, and 0.25. The distribution of X is summarized in Figure 3.18 and Figure 3.19.

i	1	2	3	Total
x_i	\$0	\$137	\$170	–
$P(X = x_i)$	0.20	0.55	0.25	1.00

Figure 3.19: The probability distribution for the random variable X, representing the bookstore's revenue from a single student.

We computed the average outcome of X as \$117.85 in Example 3.57. We call this average the **expected value** of X, denoted by $E(X)$. The expected value of a random variable is computed by adding each outcome weighted by its probability:

$$\begin{aligned} E(X) &= 0 \times P(X = 0) + 137 \times P(X = 137) + 170 \times P(X = 170) \\ &= 0 \times 0.20 + 137 \times 0.55 + 170 \times 0.25 = 117.85 \end{aligned}$$

EXPECTED VALUE OF A DISCRETE RANDOM VARIABLE

If X takes outcomes x_1, ..., x_k with probabilities $P(X = x_1)$, ..., $P(X = x_k)$, the expected value of X is the sum of each outcome multiplied by its corresponding probability:

$$\begin{aligned} E(X) &= x_1 \times P(X = x_1) + \cdots + x_k \times P(X = x_k) \\ &= \sum_{i=1}^{k} x_i P(X = x_i) \end{aligned}$$

The Greek letter μ may be used in place of the notation $E(X)$.

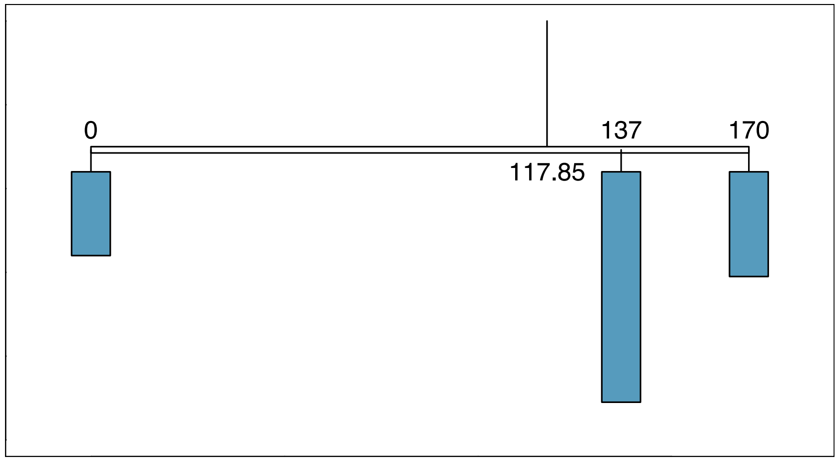


Figure 3.20: A weight system representing the probability distribution for X. The string holds the distribution at the mean to keep the system balanced.

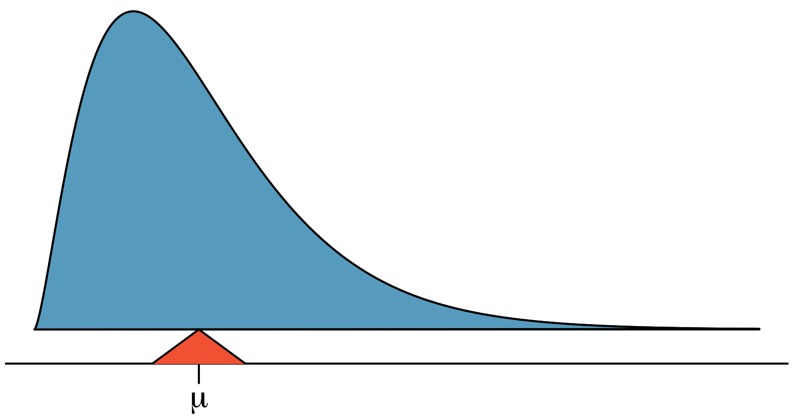


Figure 3.21: A continuous distribution can also be balanced at its mean.

The expected value for a random variable represents the average outcome. For example, $E(X) = 117.85$ represents the average amount the bookstore expects to make from a single student, which we could also write as $\mu = 117.85$.

It is also possible to compute the expected value of a continuous random variable (see Section 3.5). However, it requires a little calculus and we save it for a later class.[50]

In physics, the expectation holds the same meaning as the center of gravity. The distribution can be represented by a series of weights at each outcome, and the mean represents the balancing point. This is represented in Figures 3.18 and 3.20. The idea of a center of gravity also expands to continuous probability distributions. Figure 3.21 shows a continuous probability distribution balanced atop a wedge placed at the mean.

[50] $\mu = \int xf(x)dx$ where $f(x)$ represents a function for the density curve.

3.4.2 Variability in random variables

Suppose you ran the university bookstore. Besides how much revenue you expect to generate, you might also want to know the volatility (variability) in your revenue.

The variance and standard deviation can be used to describe the variability of a random variable. Section 2.1.4 introduced a method for finding the variance and standard deviation for a data set. We first computed deviations from the mean ($x_i - \mu$), squared those deviations, and took an average to get the variance. In the case of a random variable, we again compute squared deviations. However, we take their sum weighted by their corresponding probabilities, just like we did for the expectation. This weighted sum of squared deviations equals the variance, and we calculate the standard deviation by taking the square root of the variance, just as we did in Section 2.1.4.

GENERAL VARIANCE FORMULA

If X takes outcomes x_1, ..., x_k with probabilities $P(X = x_1)$, ..., $P(X = x_k)$ and expected value $\mu = E(X)$, then the variance of X, denoted by $Var(X)$ or the symbol σ^2, is

$$\begin{aligned}\sigma^2 &= (x_1 - \mu)^2 \times P(X = x_1) + \cdots \\ &\qquad \cdots + (x_k - \mu)^2 \times P(X = x_k) \\ &= \sum_{j=1}^{k} (x_j - \mu)^2 P(X = x_j)\end{aligned}$$

The standard deviation of X, labeled σ, is the square root of the variance.

EXAMPLE 3.58

Compute the expected value, variance, and standard deviation of X, the revenue of a single statistics student for the bookstore.

It is useful to construct a table that holds computations for each outcome separately, then add up the results.

i	1	2	3	Total
x_i	\$0	\$137	\$170	
$P(X = x_i)$	0.20	0.55	0.25	
$x_i \times P(X = x_i)$	0	75.35	42.50	117.85

E

Thus, the expected value is $\mu = 117.85$, which we computed earlier. The variance can be constructed by extending this table:

i	1	2	3	Total
x_i	\$0	\$137	\$170	
$P(X = x_i)$	0.20	0.55	0.25	
$x_i \times P(X = x_i)$	0	75.35	42.50	117.85
$x_i - \mu$	-117.85	19.15	52.15	
$(x_i - \mu)^2$	13888.62	366.72	2719.62	
$(x_i - \mu)^2 \times P(X = x_i)$	2777.7	201.7	679.9	3659.3

The variance of X is $\sigma^2 = 3659.3$, which means the standard deviation is $\sigma = \sqrt{3659.3} = \$60.49$.

GUIDED PRACTICE 3.59

The bookstore also offers a chemistry textbook for \$159 and a book supplement for \$41. From past experience, they know about 25% of chemistry students just buy the textbook while 60% buy both the textbook and supplement.[51]

G

(a) What proportion of students don't buy either book? Assume no students buy the supplement without the textbook.

(b) Let Y represent the revenue from a single student. Write out the probability distribution of Y, i.e. a table for each outcome and its associated probability.

(c) Compute the expected revenue from a single chemistry student.

(d) Find the standard deviation to describe the variability associated with the revenue from a single student.

3.4.3 Linear combinations of random variables

So far, we have thought of each variable as being a complete story in and of itself. Sometimes it is more appropriate to use a combination of variables. For instance, the amount of time a person spends commuting to work each week can be broken down into several daily commutes. Similarly, the total gain or loss in a stock portfolio is the sum of the gains and losses in its components.

EXAMPLE 3.60

John travels to work five days a week. We will use X_1 to represent his travel time on Monday, X_2 to represent his travel time on Tuesday, and so on. Write an equation using X_1, ..., X_5 that represents his travel time for the week, denoted by W.

E

His total weekly travel time is the sum of the five daily values:

$$W = X_1 + X_2 + X_3 + X_4 + X_5$$

Breaking the weekly travel time W into pieces provides a framework for understanding each source of randomness and is useful for modeling W.

[51](a) 100% - 25% - 60% = 15% of students do not buy any books for the class. Part (b) is represented by the first two lines in the table below. The expectation for part (c) is given as the total on the line $y_i \times P(Y = y_i)$. The result of part (d) is the square-root of the variance listed on in the total on the last line: $\sigma = \sqrt{Var(Y)} = \$69.28$.

i (scenario)	1 (`noBook`)	2 (`textbook`)	3 (`both`)	Total
y_i	0.00	159.00	200.00	
$P(Y = y_i)$	0.15	0.25	0.60	
$y_i \times P(Y = y_i)$	0.00	39.75	120.00	$E(Y) = 159.75$
$y_i - E(Y)$	-159.75	-0.75	40.25	
$(y_i - E(Y))^2$	25520.06	0.56	1620.06	
$(y_i - E(Y))^2 \times P(Y)$	3828.0	0.1	972.0	$Var(Y) \approx 4800$

E

EXAMPLE 3.61

It takes John an average of 18 minutes each day to commute to work. What would you expect his average commute time to be for the week?

We were told that the average (i.e. expected value) of the commute time is 18 minutes per day: $E(X_i) = 18$. To get the expected time for the sum of the five days, we can add up the expected time for each individual day:

$$\begin{aligned} E(W) &= E(X_1 + X_2 + X_3 + X_4 + X_5) \\ &= E(X_1) + E(X_2) + E(X_3) + E(X_4) + E(X_5) \\ &= 18 + 18 + 18 + 18 + 18 = 90 \text{ minutes} \end{aligned}$$

The expectation of the total time is equal to the sum of the expected individual times. More generally, the expectation of a sum of random variables is always the sum of the expectation for each random variable.

G

GUIDED PRACTICE 3.62

Elena is selling a TV at a cash auction and also intends to buy a toaster oven in the auction. If X represents the profit for selling the TV and Y represents the cost of the toaster oven, write an equation that represents the net change in Elena's cash.[52]

G

GUIDED PRACTICE 3.63

Based on past auctions, Elena figures she should expect to make about \$175 on the TV and pay about \$23 for the toaster oven. In total, how much should she expect to make or spend?[53]

G

GUIDED PRACTICE 3.64

Would you be surprised if John's weekly commute wasn't exactly 90 minutes or if Elena didn't make exactly \$152? Explain.[54]

Two important concepts concerning combinations of random variables have so far been introduced. First, a final value can sometimes be described as the sum of its parts in an equation. Second, intuition suggests that putting the individual average values into this equation gives the average value we would expect in total. This second point needs clarification – it is guaranteed to be true in what are called *linear combinations of random variables*.

A **linear combination** of two random variables X and Y is a fancy phrase to describe a combination

$$aX + bY$$

where a and b are some fixed and known numbers. For John's commute time, there were five random variables – one for each work day – and each random variable could be written as having a fixed coefficient of 1:

$$1X_1 + 1X_2 + 1X_3 + 1X_4 + 1X_5$$

For Elena's net gain or loss, the X random variable had a coefficient of +1 and the Y random variable had a coefficient of -1.

---

[52]She will make X dollars on the TV but spend Y dollars on the toaster oven: $X - Y$.

[53]$E(X - Y) = E(X) - E(Y) = 175 - 23 = \$152$. She should expect to make about \$152.

[54]No, since there is probably some variability. For example, the traffic will vary from one day to next, and auction prices will vary depending on the quality of the merchandise and the interest of the attendees.

When considering the average of a linear combination of random variables, it is safe to plug in the mean of each random variable and then compute the final result. For a few examples of nonlinear combinations of random variables – cases where we cannot simply plug in the means – see the footnote.[55]

LINEAR COMBINATIONS OF RANDOM VARIABLES AND THE AVERAGE RESULT

If X and Y are random variables, then a linear combination of the random variables is given by

$$aX + bY$$

where a and b are some fixed numbers. To compute the average value of a linear combination of random variables, plug in the average of each individual random variable and compute the result:

$$a \times E(X) + b \times E(Y)$$

Recall that the expected value is the same as the mean, e.g. $E(X) = \mu_X$.

EXAMPLE 3.65

E

Leonard has invested \$6000 in Caterpillar Inc (stock ticker: CAT) and \$2000 in Exxon Mobil Corp (XOM). If X represents the change in Caterpillar's stock next month and Y represents the change in Exxon Mobil's stock next month, write an equation that describes how much money will be made or lost in Leonard's stocks for the month.

For simplicity, we will suppose X and Y are not in percents but are in decimal form (e.g. if Caterpillar's stock increases 1%, then $X = 0.01$; or if it loses 1%, then $X = -0.01$). Then we can write an equation for Leonard's gain as

$$\$6000 \times X + \$2000 \times Y$$

If we plug in the change in the stock value for X and Y, this equation gives the change in value of Leonard's stock portfolio for the month. A positive value represents a gain, and a negative value represents a loss.

GUIDED PRACTICE 3.66

G

Caterpillar stock has recently been rising at 2.0% and Exxon Mobil's at 0.2% per month, respectively. Compute the expected change in Leonard's stock portfolio for next month.[56]

GUIDED PRACTICE 3.67

G

You should have found that Leonard expects a positive gain in Guided Practice 3.66. However, would you be surprised if he actually had a loss this month?[57]

---

[55] If X and Y are random variables, consider the following combinations: X^{1+Y}, $X \times Y$, X/Y. In such cases, plugging in the average value for each random variable and computing the result will not generally lead to an accurate average value for the end result.

[56] $E(\$6000 \times X + \$2000 \times Y) = \$6000 \times 0.020 + \$2000 \times 0.002 = \$124$.

[57] No. While stocks tend to rise over time, they are often volatile in the short term.

3.4.4 Variability in linear combinations of random variables

Quantifying the average outcome from a linear combination of random variables is helpful, but it is also important to have some sense of the uncertainty associated with the total outcome of that combination of random variables. The expected net gain or loss of Leonard's stock portfolio was considered in Guided Practice 3.66. However, there was no quantitative discussion of the volatility of this portfolio. For instance, while the average monthly gain might be about \$124 according to the data, that gain is not guaranteed. Figure 3.22 shows the monthly changes in a portfolio like Leonard's during a three year period. The gains and losses vary widely, and quantifying these fluctuations is important when investing in stocks.

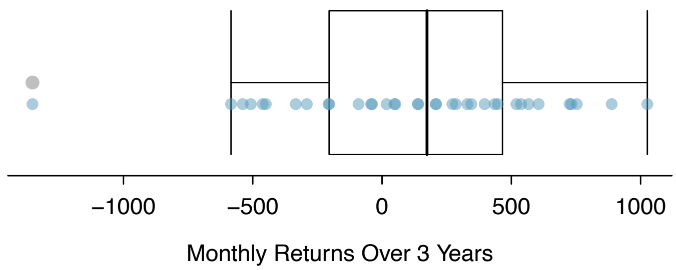


Figure 3.22: The change in a portfolio like Leonard's for 36 months, where \$6000 is in Caterpillar's stock and \$2000 is in Exxon Mobil's.

Just as we have done in many previous cases, we use the variance and standard deviation to describe the uncertainty associated with Leonard's monthly returns. To do so, the variances of each stock's monthly return will be useful, and these are shown in Figure 3.23. The stocks' returns are nearly independent.

Here we use an equation from probability theory to describe the uncertainty of Leonard's monthly returns; we leave the proof of this method to a dedicated probability course. The variance of a linear combination of random variables can be computed by plugging in the variances of the individual random variables and squaring the coefficients of the random variables:

$$Var(aX + bY) = a^2 \times Var(X) + b^2 \times Var(Y)$$

It is important to note that this equality assumes the random variables are independent; if independence doesn't hold, then a modification to this equation would be required that we leave as a topic for a future course to cover. This equation can be used to compute the variance of Leonard's monthly return:

$$\begin{aligned} Var(6000 \times X + 2000 \times Y) &= 6000^2 \times Var(X) + 2000^2 \times Var(Y) \\ &= 36,000,000 \times 0.0057 + 4,000,000 \times 0.0021 \\ &\approx 213,600 \end{aligned}$$

The standard deviation is computed as the square root of the variance: $\sqrt{213,600} = \$463$. While an average monthly return of \$124 on an \$8000 investment is nothing to scoff at, the monthly returns are so volatile that Leonard should not expect this income to be very stable.

	Mean ($\bar{x}$)	Standard deviation (s)	Variance (s^2)
CAT	0.0204	0.0757	0.0057
XOM	0.0025	0.0455	0.0021

Figure 3.23: The mean, standard deviation, and variance of the CAT and XOM stocks. These statistics were estimated from historical stock data, so notation used for sample statistics has been used.

VARIABILITY OF LINEAR COMBINATIONS OF RANDOM VARIABLES

The variance of a linear combination of random variables may be computed by squaring the constants, substituting in the variances for the random variables, and computing the result:

$$Var(aX + bY) = a^2 \times Var(X) + b^2 \times Var(Y)$$

This equation is valid as long as the random variables are independent of each other. The standard deviation of the linear combination may be found by taking the square root of the variance.

EXAMPLE 3.68

Suppose John's daily commute has a standard deviation of 4 minutes. What is the uncertainty in his total commute time for the week?

The expression for John's commute time was

$$X_1 + X_2 + X_3 + X_4 + X_5$$

Each coefficient is 1, and the variance of each day's time is $4^2 = 16$. Thus, the variance of the total weekly commute time is

$$\begin{aligned} \text{variance} &= 1^2 \times 16 + 1^2 \times 16 + 1^2 \times 16 + 1^2 \times 16 + 1^2 \times 16 = 5 \times 16 = 80 \\ \text{standard deviation} &= \sqrt{\text{variance}} = \sqrt{80} = 8.94 \end{aligned}$$

The standard deviation for John's weekly work commute time is about 9 minutes.

GUIDED PRACTICE 3.69

The computation in Example 3.68 relied on an important assumption: the commute time for each day is independent of the time on other days of that week. Do you think this is valid? Explain.[58]

GUIDED PRACTICE 3.70

Consider Elena's two auctions from Guided Practice 3.62 on page 120. Suppose these auctions are approximately independent and the variability in auction prices associated with the TV and toaster oven can be described using standard deviations of \$25 and \$8. Compute the standard deviation of Elena's net gain.[59]

Consider again Guided Practice 3.70. The negative coefficient for Y in the linear combination was eliminated when we squared the coefficients. This generally holds true: negatives in a linear combination will have no impact on the variability computed for a linear combination, but they do impact the expected value computations.

---

[58]One concern is whether traffic patterns tend to have a weekly cycle (e.g. Fridays may be worse than other days). If that is the case, and John drives, then the assumption is probably not reasonable. However, if John walks to work, then his commute is probably not affected by any weekly traffic cycle.

[59]The equation for Elena can be written as

$$(1) \times X + (-1) \times Y$$

The variances of X and Y are 625 and 64. We square the coefficients and plug in the variances:

$$(1)^2 \times Var(X) + (-1)^2 \times Var(Y) = 1 \times 625 + 1 \times 64 = 689$$

The variance of the linear combination is 689, and the standard deviation is the square root of 689: about \$26.25.

Exercises

3.29 College smokers. At a university, 13% of students smoke.

(a) Calculate the expected number of smokers in a random sample of 100 students from this university.

(b) The university gym opens at 9 am on Saturday mornings. One Saturday morning at 8:55 am there are 27 students outside the gym waiting for it to open. Should you use the same approach from part (a) to calculate the expected number of smokers among these 27 students?

3.30 Ace of clubs wins. Consider the following card game with a well-shuffled deck of cards. If you draw a red card, you win nothing. If you get a spade, you win $5. For any club, you win $10 plus an extra $20 for the ace of clubs.

(a) Create a probability model for the amount you win at this game. Also, find the expected winnings for a single game and the standard deviation of the winnings.

(b) What is the maximum amount you would be willing to pay to play this game? Explain your reasoning.

3.31 Hearts win. In a new card game, you start with a well-shuffled full deck and draw 3 cards without replacement. If you draw 3 hearts, you win $50. If you draw 3 black cards, you win $25. For any other draws, you win nothing.

(a) Create a probability model for the amount you win at this game, and find the expected winnings. Also compute the standard deviation of this distribution.

(b) If the game costs $5 to play, what would be the expected value and standard deviation of the net profit (or loss)? *(Hint: profit = winnings − cost;* $X - 5$*)*

(c) If the game costs $5 to play, should you play this game? Explain.

3.32 Is it worth it? Andy is always looking for ways to make money fast. Lately, he has been trying to make money by gambling. Here is the game he is considering playing: The game costs $2 to play. He draws a card from a deck. If he gets a number card (2-10), he wins nothing. For any face card (jack, queen or king), he wins $3. For any ace, he wins $5, and he wins an *extra* $20 if he draws the ace of clubs.

(a) Create a probability model and find Andy's expected profit per game.

(b) Would you recommend this game to Andy as a good way to make money? Explain.

3.33 Portfolio return. A portfolio's value increases by 18% during a financial boom and by 9% during normal times. It decreases by 12% during a recession. What is the expected return on this portfolio if each scenario is equally likely?

3.34 Baggage fees. An airline charges the following baggage fees: $25 for the first bag and $35 for the second. Suppose 54% of passengers have no checked luggage, 34% have one piece of checked luggage and 12% have two pieces. We suppose a negligible portion of people check more than two bags.

(a) Build a probability model, compute the average revenue per passenger, and compute the corresponding standard deviation.

(b) About how much revenue should the airline expect for a flight of 120 passengers? With what standard deviation? Note any assumptions you make and if you think they are justified.

3.35 American roulette. The game of American roulette involves spinning a wheel with 38 slots: 18 red, 18 black, and 2 green. A ball is spun onto the wheel and will eventually land in a slot, where each slot has an equal chance of capturing the ball. Gamblers can place bets on red or black. If the ball lands on their color, they double their money. If it lands on another color, they lose their money. Suppose you bet $1 on red. What's the expected value and standard deviation of your winnings?

3.36 European roulette. The game of European roulette involves spinning a wheel with 37 slots: 18 red, 18 black, and 1 green. A ball is spun onto the wheel and will eventually land in a slot, where each slot has an equal chance of capturing the ball. Gamblers can place bets on red or black. If the ball lands on their color, they double their money. If it lands on another color, they lose their money.

(a) Suppose you play roulette and bet $3 on a single round. What is the expected value and standard deviation of your total winnings?

(b) Suppose you bet $1 in three different rounds. What is the expected value and standard deviation of your total winnings?

(c) How do your answers to parts (a) and (b) compare? What does this say about the riskiness of the two games?

3.5 Continuous distributions

So far in this chapter we've discussed cases where the outcome of a variable is discrete. In this section, we consider a context where the outcome is a continuous numerical variable.

E

EXAMPLE 3.71

Figure 3.24 shows a few different hollow histograms for the heights of US adults. How does changing the number of bins allow you to make different interpretations of the data?

Adding more bins provides greater detail. This sample is extremely large, which is why much smaller bins still work well. Usually we do not use so many bins with smaller sample sizes since small counts per bin mean the bin heights are very volatile.

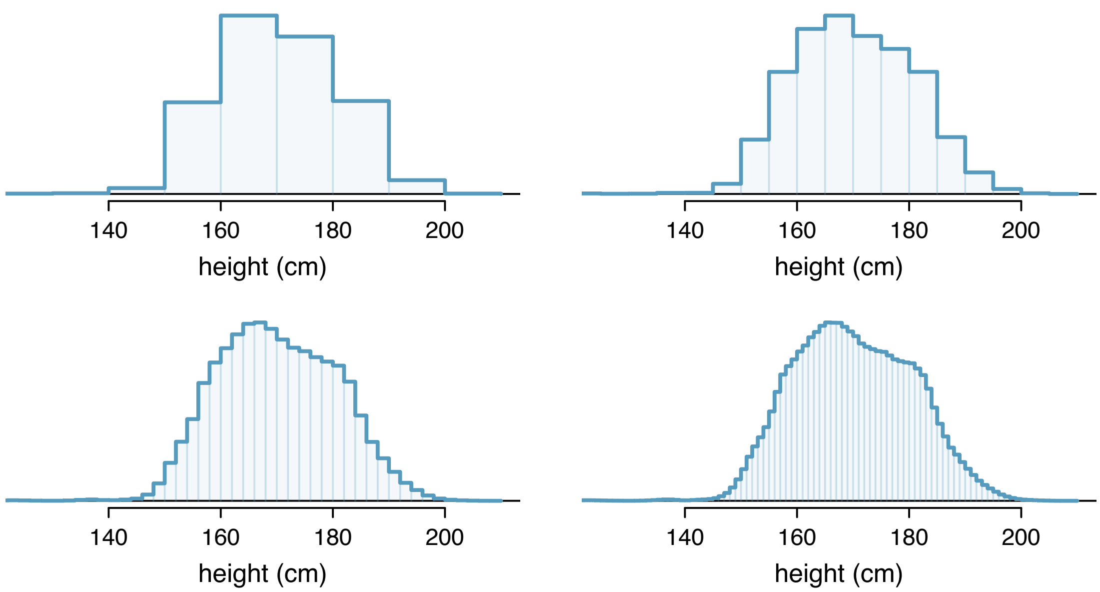


Figure 3.24: Four hollow histograms of US adults heights with varying bin widths.

E

EXAMPLE 3.72

What proportion of the sample is between 180 cm and 185 cm tall (about 5'11" to 6'1")?

We can add up the heights of the bins in the range 180 cm and 185 and divide by the sample size. For instance, this can be done with the two shaded bins shown in Figure 3.25. The two bins in this region have counts of 195,307 and 156,239 people, resulting in the following estimate of the probability:

$$\frac{195307 + 156239}{3{,}000{,}000} = 0.1172$$

This fraction is the same as the proportion of the histogram's area that falls in the range 180 to 185 cm.

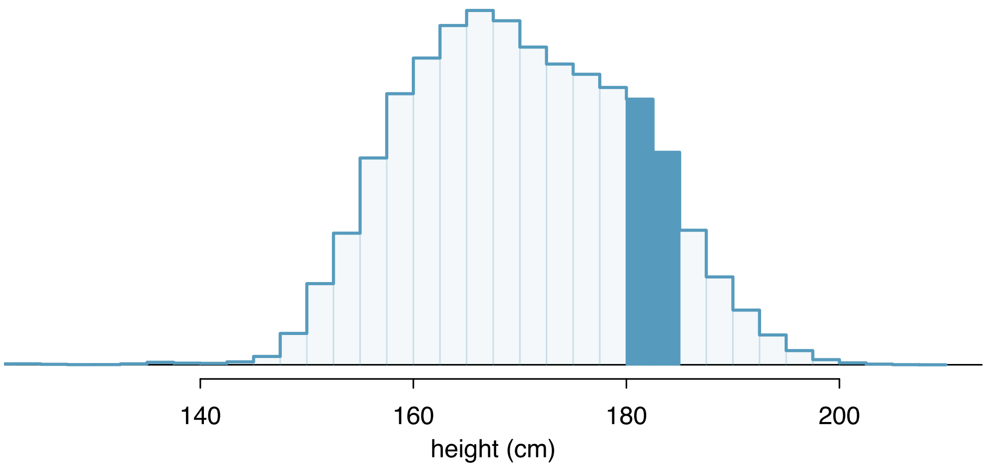


Figure 3.25: A histogram with bin sizes of 2.5 cm. The shaded region represents individuals with heights between 180 and 185 cm.

3.5.1 From histograms to continuous distributions

Examine the transition from a boxy hollow histogram in the top-left of Figure 3.24 to the much smoother plot in the lower-right. In this last plot, the bins are so slim that the hollow histogram is starting to resemble a smooth curve. This suggests the population height as a *continuous* numerical variable might best be explained by a curve that represents the outline of extremely slim bins.

This smooth curve represents a **probability density function** (also called a **density** or **distribution**), and such a curve is shown in Figure 3.26 overlaid on a histogram of the sample. A density has a special property: the total area under the density's curve is 1.

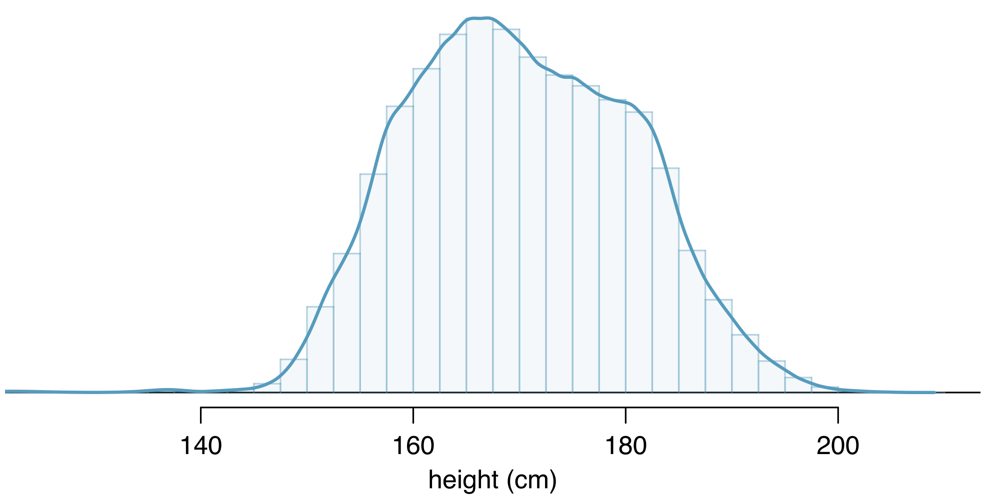


Figure 3.26: The continuous probability distribution of heights for US adults.

3.5.2 Probabilities from continuous distributions

We computed the proportion of individuals with heights 180 to 185 cm in Example 3.72 as a fraction:

$$\frac{\text{number of people between 180 and 185}}{\text{total sample size}}$$

We found the number of people with heights between 180 and 185 cm by determining the fraction of the histogram's area in this region. Similarly, we can use the area in the shaded region under the curve to find a probability (with the help of a computer):

$$P(\texttt{height} \text{ between } 180 \text{ and } 185) = \text{area between } 180 \text{ and } 185 = 0.1157$$

The probability that a randomly selected person is between 180 and 185 cm is 0.1157. This is very close to the estimate from Example 3.72: 0.1172.

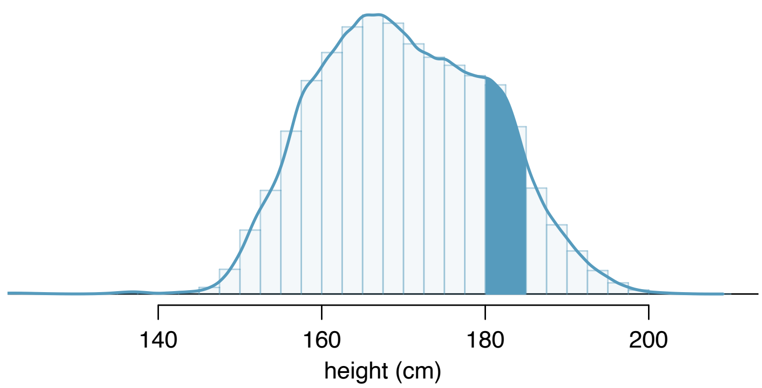


Figure 3.27: Density for heights in the US adult population with the area between 180 and 185 cm shaded. Compare this plot with Figure 3.25.

GUIDED PRACTICE 3.73

Three US adults are randomly selected. The probability a single adult is between 180 and 185 cm is 0.1157.[60]

(a) What is the probability that all three are between 180 and 185 cm tall?

(b) What is the probability that none are between 180 and 185 cm?

EXAMPLE 3.74

What is the probability that a randomly selected person is **exactly** 180 cm? Assume you can measure perfectly.

This probability is zero. A person might be close to 180 cm, but not exactly 180 cm tall. This also makes sense with the definition of probability as area; there is no area captured between 180 cm and 180 cm.

GUIDED PRACTICE 3.75

Suppose a person's height is rounded to the nearest centimeter. Is there a chance that a random person's **measured** height will be 180 cm?[61]

---

[60] Brief answers: (a) $0.1157 \times 0.1157 \times 0.1157 = 0.0015$. (b) $(1 - 0.1157)^3 = 0.692$

[61] This has positive probability. Anyone between 179.5 cm and 180.5 cm will have a *measured* height of 180 cm. This is probably a more realistic scenario to encounter in practice versus Example 3.74.

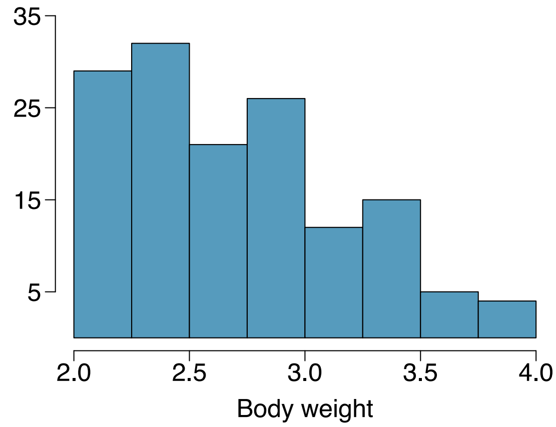
Exercises

3.37 Cat weights. The histogram shown below represents the weights (in kg) of 47 female and 97 male cats.[62]

(a) What fraction of these cats weigh less than 2.5 kg?

(b) What fraction of these cats weigh between 2.5 and 2.75 kg?

(c) What fraction of these cats weigh between 2.75 and 3.5 kg?



3.38 Income and gender. The relative frequency table below displays the distribution of annual total personal income (in 2009 inflation-adjusted dollars) for a representative sample of 96,420,486 Americans. These data come from the American Community Survey for 2005-2009. This sample is comprised of 59% males and 41% females.[63]

(a) Describe the distribution of total personal income.

(b) What is the probability that a randomly chosen US resident makes less than $50,000 per year?

(c) What is the probability that a randomly chosen US resident makes less than $50,000 per year and is female? Note any assumptions you make.

(d) The same data source indicates that 71.8% of females make less than $50,000 per year. Use this value to determine whether or not the assumption you made in part (c) is valid.

Income	*Total*
$1 to $9,999 or loss	2.2%
$10,000 to $14,999	4.7%
$15,000 to $24,999	15.8%
$25,000 to $34,999	18.3%
$35,000 to $49,999	21.2%
$50,000 to $64,999	13.9%
$65,000 to $74,999	5.8%
$75,000 to $99,999	8.4%
$100,000 or more	9.7%

[62] W. N. Venables and B. D. Ripley. *Modern Applied Statistics with S*. Fourth Edition. www.stats.ox.ac.uk/pub/MASS4. New York: Springer, 2002.

[63] U.S. Census Bureau, 2005-2009 American Community Survey.

Chapter exercises

3.39 Grade distributions. Each row in the table below is a proposed grade distribution for a class. Identify each as a valid or invalid probability distribution, and explain your reasoning.

	Grades				
	A	B	C	D	F
(a)	0.3	0.3	0.3	0.2	0.1
(b)	0	0	1	0	0
(c)	0.3	0.3	0.3	0	0
(d)	0.3	0.5	0.2	0.1	-0.1
(e)	0.2	0.4	0.2	0.1	0.1
(f)	0	-0.1	1.1	0	0

3.40 Health coverage, frequencies. The Behavioral Risk Factor Surveillance System (BRFSS) is an annual telephone survey designed to identify risk factors in the adult population and report emerging health trends. The following table summarizes two variables for the respondents: health status and health coverage, which describes whether each respondent had health insurance.[64]

		Health Status					
		Excellent	Very good	Good	Fair	Poor	Total
Health	No	459	727	854	385	99	2,524
Coverage	Yes	4,198	6,245	4,821	1,634	578	17,476
	Total	4,657	6,972	5,675	2,019	677	20,000

(a) If we draw one individual at random, what is the probability that the respondent has excellent health and doesn't have health coverage?

(b) If we draw one individual at random, what is the probability that the respondent has excellent health or doesn't have health coverage?

3.41 HIV in Swaziland. Swaziland has the highest HIV prevalence in the world: 25.9% of this country's population is infected with HIV.[65] The ELISA test is one of the first and most accurate tests for HIV. For those who carry HIV, the ELISA test is 99.7% accurate. For those who do not carry HIV, the test is 92.6% accurate. If an individual from Swaziland has tested positive, what is the probability that he carries HIV?

3.42 Twins. About 30% of human twins are identical, and the rest are fraternal. Identical twins are necessarily the same sex – half are males and the other half are females. One-quarter of fraternal twins are both male, one-quarter both female, and one-half are mixes: one male, one female. You have just become a parent of twins and are told they are both girls. Given this information, what is the probability that they are identical?

3.43 Cost of breakfast. Sally gets a cup of coffee and a muffin every day for breakfast from one of the many coffee shops in her neighborhood. She picks a coffee shop each morning at random and independently of previous days. The average price of a cup of coffee is \$1.40 with a standard deviation of 30¢ (\$0.30), the average price of a muffin is \$2.50 with a standard deviation of 15¢, and the two prices are independent of each other.

(a) What is the mean and standard deviation of the amount she spends on breakfast daily?

(b) What is the mean and standard deviation of the amount she spends on breakfast weekly (7 days)?

[64] Office of Surveillance, Epidemiology, and Laboratory Services Behavioral Risk Factor Surveillance System, BRFSS 2010 Survey Data.

[65] Source: CIA Factbook, Country Comparison: HIV/AIDS - Adult Prevalence Rate.

3.44 Scooping ice cream. Ice cream usually comes in 1.5 quart boxes (48 fluid ounces), and ice cream scoops hold about 2 ounces. However, there is some variability in the amount of ice cream in a box as well as the amount of ice cream scooped out. We represent the amount of ice cream in the box as X and the amount scooped out as Y. Suppose these random variables have the following means, standard deviations, and variances:

	mean	SD	variance
X	48	1	1
Y	2	0.25	0.0625

(a) An entire box of ice cream, plus 3 scoops from a second box is served at a party. How much ice cream do you expect to have been served at this party? What is the standard deviation of the amount of ice cream served?

(b) How much ice cream would you expect to be left in the box after scooping out one scoop of ice cream? That is, find the expected value of $X - Y$. What is the standard deviation of the amount left in the box?

(c) Using the context of this exercise, explain why we add variances when we subtract one random variable from another.

3.45 Variance of a mean, Part I. Suppose we have independent observations X_1 and X_2 from a distribution with mean μ and standard deviation σ. What is the variance of the mean of the two values: $\frac{X_1+X_2}{2}$?

3.46 Variance of a mean, Part II. Suppose we have 3 independent observations X_1, X_2, X_3 from a distribution with mean μ and standard deviation σ. What is the variance of the mean of these 3 values: $\frac{X_1+X_2+X_3}{3}$?

3.47 Variance of a mean, Part III. Suppose we have n independent observations X_1, X_2, ..., X_n from a distribution with mean μ and standard deviation σ. What is the variance of the mean of these n values: $\frac{X_1+X_2+\cdots+X_n}{n}$?

Chapter 4

Distributions of random variables

In this chapter, we discuss statistical distributions that frequently arise in the context of data analysis or statistical inference. We start with the normal distribution in the first section, which is used frequently in later chapters of this book. The remaining sections will occasionally be referenced but may be considered optional for the content in this book.

For videos, slides, and other resources, please visit
www.openintro.org/os

4.1 Normal distribution

Among all the distributions we see in practice, one is overwhelmingly the most common. The symmetric, unimodal, bell curve is ubiquitous throughout statistics. Indeed it is so common, that people often know it as the **normal curve** or **normal distribution**,[1] shown in Figure 4.1. Variables such as SAT scores and heights of US adult males closely follow the normal distribution.

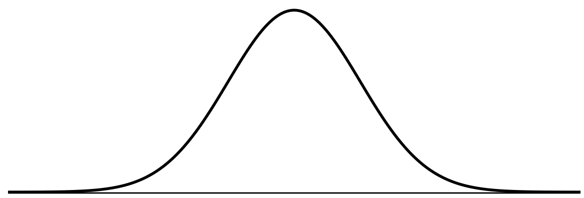


Figure 4.1: A normal curve.

NORMAL DISTRIBUTION FACTS

Many variables are nearly normal, but none are exactly normal. Thus the normal distribution, while not perfect for any single problem, is very useful for a variety of problems. We will use it in data exploration and to solve important problems in statistics.

4.1.1 Normal distribution model

The **normal distribution** always describes a symmetric, unimodal, bell-shaped curve. However, these curves can look different depending on the details of the model. Specifically, the normal distribution model can be adjusted using two parameters: mean and standard deviation. As you can probably guess, changing the mean shifts the bell curve to the left or right, while changing the standard deviation stretches or constricts the curve. Figure 4.2 shows the normal distribution with mean 0 and standard deviation 1 in the left panel and the normal distributions with mean 19 and standard deviation 4 in the right panel. Figure 4.3 shows these distributions on the same axis.

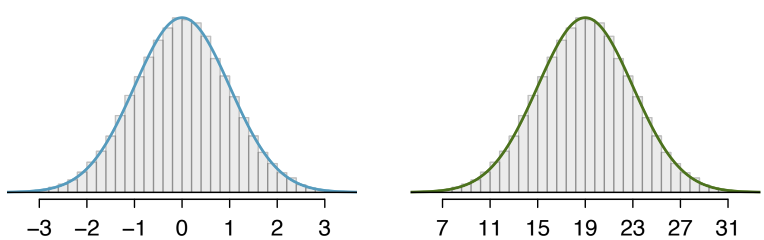


Figure 4.2: Both curves represent the normal distribution. However, they differ in their center and spread.

If a normal distribution has mean μ and standard deviation σ, we may write the distribution as $N(\mu, \sigma)$. The two distributions in Figure 4.3 may be written as

$$N(\mu = 0, \sigma = 1) \quad \text{and} \quad N(\mu = 19, \sigma = 4)$$

Because the mean and standard deviation describe a normal distribution exactly, they are called the distribution's **parameters**. The normal distribution with mean $\mu = 0$ and standard deviation $\sigma = 1$ is called the **standard normal distribution**.

[1] It is also introduced as the Gaussian distribution after Frederic Gauss, the first person to formalize its mathematical expression.

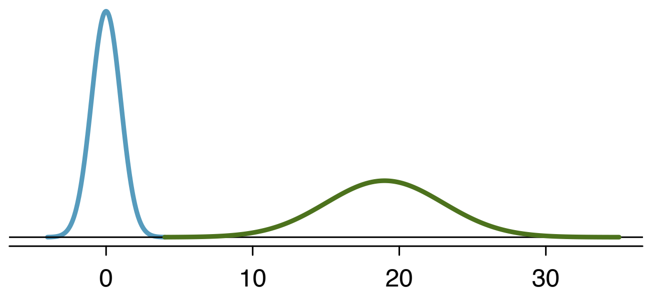


Figure 4.3: The normal distributions shown in Figure 4.2 but plotted together and on the same scale.

GUIDED PRACTICE 4.1

Write down the short-hand for a normal distribution with[2]
(a) mean 5 and standard deviation 3,
(b) mean -100 and standard deviation 10, and
(c) mean 2 and standard deviation 9.

4.1.2 Standardizing with Z-scores

We often want to put data onto a standardized scale, which can make comparisons more reasonable.

EXAMPLE 4.2

Table 4.4 shows the mean and standard deviation for total scores on the SAT and ACT. The distribution of SAT and ACT scores are both nearly normal. Suppose Ann scored 1300 on her SAT and Tom scored 24 on his ACT. Who performed better?

We use the standard deviation as a guide. Ann is 1 standard deviation above average on the SAT: $1100 + 200 = 1300$. Tom is 0.5 standard deviations above the mean on the ACT: $21 + 0.5 \times 6 = 24$. In Figure 4.5, we can see that Ann tends to do better with respect to everyone else than Tom did, so her score was better.

	SAT	ACT
Mean	1100	21
SD	200	6

Figure 4.4: Mean and standard deviation for the SAT and ACT.

Example 4.2 used a standardization technique called a Z-score, a method most commonly employed for nearly normal observations but that may be used with any distribution. The **Z-score** of an observation is defined as the number of standard deviations it falls above or below the mean. If the observation is one standard deviation above the mean, its Z-score is 1. If it is 1.5 standard deviations *below* the mean, then its Z-score is -1.5. If x is an observation from a distribution $N(\mu, \sigma)$, we define the Z-score mathematically as

$$Z = \frac{x - \mu}{\sigma}$$

Using $\mu_{SAT} = 1100$, $\sigma_{SAT} = 200$, and $x_{\text{Ann}} = 1300$, we find Ann's Z-score:

$$Z_{\text{Ann}} = \frac{x_{\text{Ann}} - \mu_{\text{SAT}}}{\sigma_{\text{SAT}}} = \frac{1300 - 1100}{200} = 1$$

[2](a) $N(\mu = 5, \sigma = 3)$. (b) $N(\mu = -100, \sigma = 10)$. (c) $N(\mu = 2, \sigma = 9)$.

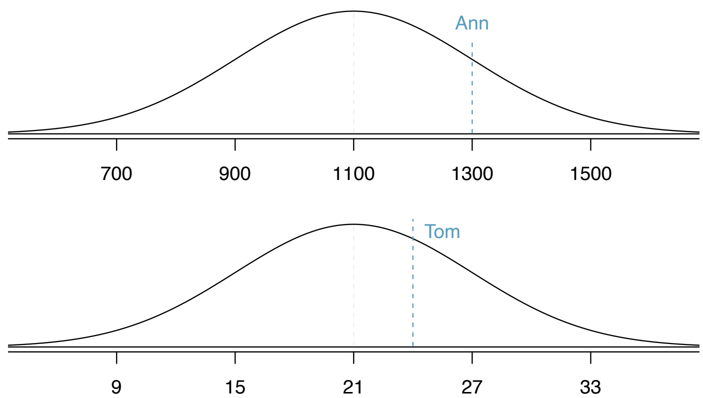


Figure 4.5: Ann's and Tom's scores shown against the SAT and ACT distributions.

THE Z-SCORE

The Z-score of an observation is the number of standard deviations it falls above or below the mean. We compute the Z-score for an observation x that follows a distribution with mean μ and standard deviation σ using

$$Z = \frac{x - \mu}{\sigma}$$

G

GUIDED PRACTICE 4.3

Use Tom's ACT score, 24, along with the ACT mean and standard deviation to find his Z-score.[3]

Observations above the mean always have positive Z-scores, while those below the mean always have negative Z-scores. If an observation is equal to the mean, such as an SAT score of 1100, then the Z-score is 0.

G

GUIDED PRACTICE 4.4

Let X represent a random variable from $N(\mu = 3, \sigma = 2)$, and suppose we observe $x = 5.19$.
(a) Find the Z-score of x.
(b) Use the Z-score to determine how many standard deviations above or below the mean x falls.[4]

G

GUIDED PRACTICE 4.5

Head lengths of brushtail possums follow a normal distribution with mean 92.6 mm and standard deviation 3.6 mm. Compute the Z-scores for possums with head lengths of 95.4 mm and 85.8 mm.[5]

We can use Z-scores to roughly identify which observations are more unusual than others. An observation x_1 is said to be more unusual than another observation x_2 if the absolute value of its Z-score is larger than the absolute value of the other observation's Z-score: $|Z_1| > |Z_2|$. This technique is especially insightful when a distribution is symmetric.

G

GUIDED PRACTICE 4.6

Which of the observations in Guided Practice 4.5 is more unusual?[6]

---

[3] $Z_{Tom} = \frac{x_{\text{Tom}} - \mu_{\text{ACT}}}{\sigma_{\text{ACT}}} = \frac{24-21}{6} = 0.5$

[4] (a) Its Z-score is given by $Z = \frac{x-\mu}{\sigma} = \frac{5.19-3}{2} = 2.19/2 = 1.095$. (b) The observation x is 1.095 standard deviations *above* the mean. We know it must be above the mean since Z is positive.

[5] For $x_1 = 95.4$ mm: $Z_1 = \frac{x_1-\mu}{\sigma} = \frac{95.4-92.6}{3.6} = 0.78$. For $x_2 = 85.8$ mm: $Z_2 = \frac{85.8-92.6}{3.6} = -1.89$.

[6] Because the *absolute value* of Z-score for the second observation is larger than that of the first, the second observation has a more unusual head length.

4.1.3 Finding tail areas

It's very useful in statistics to be able to identify tail areas of distributions. For instance, what fraction of people have an SAT score below Ann's score of 1300? This is the same as the **percentile** Ann is at, which is the percentage of cases that have lower scores than Ann. We can visualize such a tail area like the curve and shading shown in Figure 4.6.

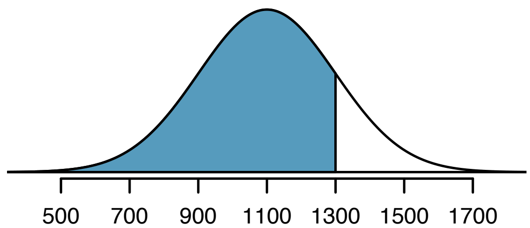


Figure 4.6: The area to the left of Z represents the fraction of people who scored lower than Ann.

There are many techniques for doing this, and we'll discuss three of the options.

1. The most common approach in practice is to use statistical software. For example, in the program **R**, we could find the area shown in Figure 4.6 using the following command, which takes in the Z-score and returns the lower tail area:
```
> pnorm(1)
[1] 0.8413447
```
According to this calculation, the region shaded that is below 1300 represents the proportion 0.841 (84.1%) of SAT test takers who had Z-scores below $Z = 1$. More generally, we can also specify the cutoff explicitly if we also note the mean and standard deviation:
```
> pnorm(1300, mean = 1100, sd = 200)
[1] 0.8413447
```
There are many other software options, such as Python or SAS; even spreadsheet programs such as Excel and Google Sheets support these calculations.

2. A common strategy in classrooms is to use a graphing calculator, such as a TI or Casio calculator. These calculators require a series of button presses that are less concisely described. You can find instructions on using these calculators for finding tail areas of a normal distribution in the OpenIntro video library:

www.openintro.org/videos

3. The last option for finding tail areas is to use what's called a **probability table**; these are occasionally used in classrooms but rarely in practice. Appendix C.1 contains such a table and a guide for how to use it.

We will solve normal distribution problems in this section by always first finding the Z-score. The reason is that we will encounter close parallels called test statistics beginning in Chapter 5; these are, in many instances, an equivalent of a Z-score.

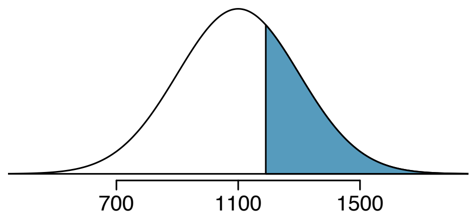
4.1.4 Normal probability examples

Cumulative SAT scores are approximated well by a normal model, $N(\mu = 1100, \sigma = 200)$.

EXAMPLE 4.7

Shannon is a randomly selected SAT taker, and nothing is known about Shannon's SAT aptitude. What is the probability Shannon scores at least 1190 on her SATs?

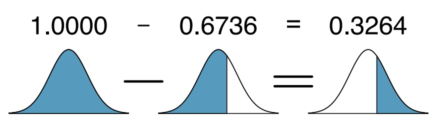
First, always draw and label a picture of the normal distribution. (Drawings need not be exact to be useful.) We are interested in the chance she scores above 1190, so we shade this upper tail:



The picture shows the mean and the values at 2 standard deviations above and below the mean. The simplest way to find the shaded area under the curve makes use of the Z-score of the cutoff value. With $\mu = 1100$, $\sigma = 200$, and the cutoff value $x = 1190$, the Z-score is computed as

$$Z = \frac{x - \mu}{\sigma} = \frac{1190 - 1100}{200} = \frac{90}{200} = 0.45$$

Using statistical software (or another preferred method), we can area left of $Z = 0.45$ as 0.6736. To find the area *above* $Z = 0.45$, we compute one minus the area of the lower tail:



The probability Shannon scores at least 1190 on the SAT is 0.3264.

ALWAYS DRAW A PICTURE FIRST, AND FIND THE Z-SCORE SECOND

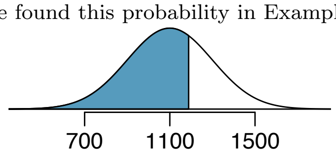
For any normal probability situation, *always always always* draw and label the normal curve and shade the area of interest first. The picture will provide an estimate of the probability. After drawing a figure to represent the situation, identify the Z-score for the value of interest.

GUIDED PRACTICE 4.8

If the probability of Shannon scoring at least 1190 is 0.3264, then what is the probability she scores less than 1190? Draw the normal curve representing this exercise, shading the lower region instead of the upper one.[7]

---

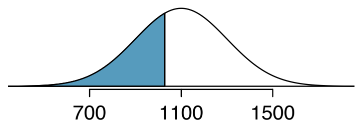
[7] We found this probability in Example 4.7: 0.6736.



EXAMPLE 4.9

Edward earned a 1030 on his SAT. What is his percentile?

First, a picture is needed. Edward's percentile is the proportion of people who do not get as high as a 1030. These are the scores to the left of 1030.



Identifying the mean $\mu = 1100$, the standard deviation $\sigma = 200$, and the cutoff for the tail area $x = 1030$ makes it easy to compute the Z-score:

$$Z = \frac{x - \mu}{\sigma} = \frac{1030 - 1100}{200} = -0.35$$

Using statistical software, we get a tail area of 0.3632. Edward is at the 36^{th} percentile.

GUIDED PRACTICE 4.10

Use the results of Example 4.9 to compute the proportion of SAT takers who did better than Edward. Also draw a new picture.[8]

FINDING AREAS TO THE RIGHT

Many software programs return the area to the left when given a Z-score. If you would like the area to the right, first find the area to the left and then subtract this amount from one.

GUIDED PRACTICE 4.11

Stuart earned an SAT score of 1500. Draw a picture for each part.
(a) What is his percentile?
(b) What percent of SAT takers did better than Stuart?[9]

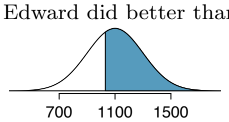
Based on a sample of 100 men, the heights of male adults in the US is nearly normal with mean 70.0" and standard deviation 3.3".

GUIDED PRACTICE 4.12

Mike is 5'7" and Jose is 6'4", and they both live in the US.
(a) What is Mike's height percentile?
(b) What is Jose's height percentile?
Also draw one picture for each part.[10]

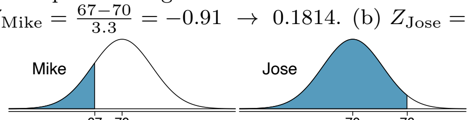
---

[8] If Edward did better than 36% of SAT takers, then about 64% must have done better than him.



[9] We leave the drawings to you. (a) $Z = \frac{1500-1100}{200} = 2 \rightarrow 0.9772$. (b) $1 - 0.9772 = 0.0228$.

[10] First put the heights into inches: 67 and 76 inches. Figures are shown below.
(a) $Z_{\text{Mike}} = \frac{67-70}{3.3} = -0.91 \rightarrow 0.1814$. (b) $Z_{\text{Jose}} = \frac{76-70}{3.3} = 1.82 \rightarrow 0.9656$.

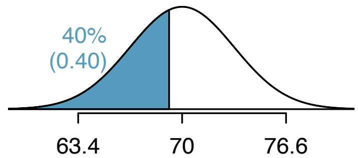


The last several problems have focused on finding the percentile (or upper tail) for a particular observation. What if you would like to know the observation corresponding to a particular percentile?

EXAMPLE 4.13

Erik's height is at the 40^{th} percentile. How tall is he?

As always, first draw the picture.



E

In this case, the lower tail probability is known (0.40), which can be shaded on the diagram. We want to find the observation that corresponds to this value. As a first step in this direction, we determine the Z-score associated with the 40^{th} percentile. Using software, we can obtain the corresponding Z-score of about -0.25.

Knowing $Z_{Erik} = -0.25$ and the population parameters $\mu = 70$ and $\sigma = 3.3$ inches, the Z-score formula can be set up to determine Erik's unknown height, labeled x_{Erik}:

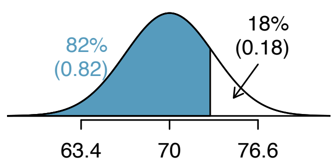
$$-0.25 = Z_{\text{Erik}} = \frac{x_{\text{Erik}} - \mu}{\sigma} = \frac{x_{\text{Erik}} - 70}{3.3}$$

Solving for x_{Erik} yields a height of 69.18 inches. That is, Erik is about 5'9".

EXAMPLE 4.14

What is the adult male height at the 82^{nd} percentile?

Again, we draw the figure first.



E

Next, we want to find the Z-score at the 82^{nd} percentile, which will be a positive value and can be found using software as $Z = 0.92$. Finally, the height x is found using the Z-score formula with the known mean μ, standard deviation σ, and Z-score $Z = 0.92$:

$$0.92 = Z = \frac{x - \mu}{\sigma} = \frac{x - 70}{3.3}$$

This yields 73.04 inches or about 6'1" as the height at the 82^{nd} percentile.

GUIDED PRACTICE 4.15

G

The SAT scores follow $N(1100, 200)$.[11]
(a) What is the 95^{th} percentile for SAT scores?
(b) What is the 97.5^{th} percentile for SAT scores?

[11] Short answers: (a) $Z_{95} = 1.65 \rightarrow 1430$ SAT score. (b) $Z_{97.5} = 1.96 \rightarrow 1492$ SAT score.

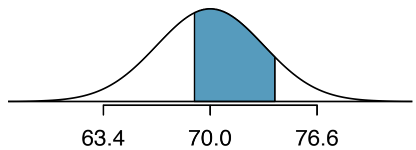
GUIDED PRACTICE 4.16

Adult male heights follow $N(70.0", 3.3")$.[12]
(a) What is the probability that a randomly selected male adult is at least 6'2" (74 inches)?
(b) What is the probability that a male adult is shorter than 5'9" (69 inches)?

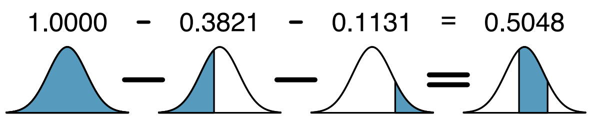
EXAMPLE 4.17

What is the probability that a random adult male is between 5'9" and 6'2"?

These heights correspond to 69 inches and 74 inches. First, draw the figure. The area of interest is no longer an upper or lower tail.



The total area under the curve is 1. If we find the area of the two tails that are not shaded (from Guided Practice 4.16, these areas are 0.3821 and 0.1131), then we can find the middle area:



That is, the probability of being between 5'9" and 6'2" is 0.5048.

GUIDED PRACTICE 4.18

SAT scores follow $N(1100, 200)$. What percent of SAT takers get between 1100 and 1400?[13]

GUIDED PRACTICE 4.19

Adult male heights follow $N(70.0", 3.3")$. What percent of adult males are between 5'5" and 5'7"?[14]

---

[12] Short answers: (a) $Z = 1.21 \rightarrow 0.8869$, then subtract this value from 1 to get 0.1131. (b) $Z = -0.30 \rightarrow 0.3821$.

[13] This is an abbreviated solution. (Be sure to draw a figure!) First find the percent who get below 1100 and the percent that get above 1400: $Z_{1100} = 0.00 \rightarrow 0.5000$ (area below), $Z_{1400} = 1.5 \rightarrow 0.0668$ (area above). Final answer: $1.0000 - 0.5000 - 0.0668 = 0.4332$.

[14] 5'5" is 65 inches ($Z = -1.52$). 5'7" is 67 inches ($Z = -0.91$). Numerical solution: $1.000 - 0.0643 - 0.8186 = 0.1171$, i.e. 11.71%.

4.1.5 68-95-99.7 rule

Here, we present a useful rule of thumb for the probability of falling within 1, 2, and 3 standard deviations of the mean in the normal distribution. This will be useful in a wide range of practical settings, especially when trying to make a quick estimate without a calculator or Z-table.

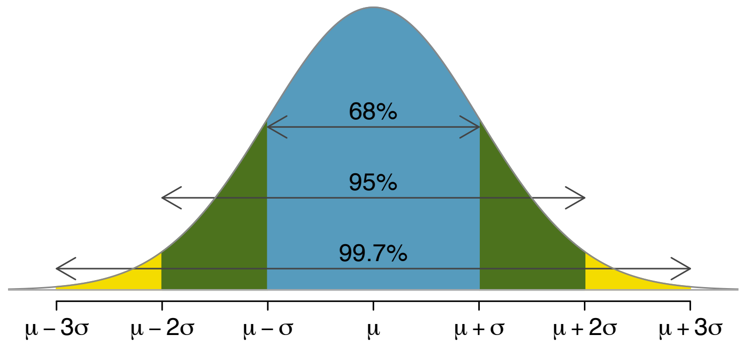


Figure 4.7: Probabilities for falling within 1, 2, and 3 standard deviations of the mean in a normal distribution.

G

GUIDED PRACTICE 4.20

Use software, a calculator, or a probability table to confirm that about 68%, 95%, and 99.7% of observations fall within 1, 2, and 3, standard deviations of the mean in the normal distribution, respectively. For instance, first find the area that falls between $Z = -1$ and $Z = 1$, which should have an area of about 0.68. Similarly there should be an area of about 0.95 between $Z = -2$ and $Z = 2$.[15]

It is possible for a normal random variable to fall 4, 5, or even more standard deviations from the mean. However, these occurrences are very rare if the data are nearly normal. The probability of being further than 4 standard deviations from the mean is about 1-in-15,000. For 5 and 6 standard deviations, it is about 1-in-2 million and 1-in-500 million, respectively.

G

GUIDED PRACTICE 4.21

SAT scores closely follow the normal model with mean $\mu = 1100$ and standard deviation $\sigma = 200$.[16]
(a) About what percent of test takers score 700 to 1500?
(b) What percent score between 1100 and 1500?

[15]First draw the pictures. Using software, we get 0.6827 within 1 standard deviation, 0.9545 within 2 standard deviations, and 0.9973 within 3 standard deviations.

[16](a) 700 and 1500 represent two standard deviations below and above the mean, which means about 95% of test takers will score between 700 and 1500. (b) We found that 700 to 1500 represents about 95% of test takers. These test takers would be evenly split by the center of the distribution, 1100, so $\frac{95\%}{2} = 47.5\%$ of all test takers score between 1100 and 1500.

Exercises

4.1 Area under the curve, Part I. What percent of a standard normal distribution $N(\mu = 0, \sigma = 1)$ is found in each region? Be sure to draw a graph.

(a) $Z < -1.35$ (b) $Z > 1.48$ (c) $-0.4 < Z < 1.5$ (d) $|Z| > 2$

4.2 Area under the curve, Part II. What percent of a standard normal distribution $N(\mu = 0, \sigma = 1)$ is found in each region? Be sure to draw a graph.

(a) $Z > -1.13$ (b) $Z < 0.18$ (c) $Z > 8$ (d) $|Z| < 0.5$

4.3 GRE scores, Part I. Sophia who took the Graduate Record Examination (GRE) scored 160 on the Verbal Reasoning section and 157 on the Quantitative Reasoning section. The mean score for Verbal Reasoning section for all test takers was 151 with a standard deviation of 7, and the mean score for the Quantitative Reasoning was 153 with a standard deviation of 7.67. Suppose that both distributions are nearly normal.

(a) Write down the short-hand for these two normal distributions.
(b) What is Sophia's Z-score on the Verbal Reasoning section? On the Quantitative Reasoning section? Draw a standard normal distribution curve and mark these two Z-scores.
(c) What do these Z-scores tell you?
(d) Relative to others, which section did she do better on?
(e) Find her percentile scores for the two exams.
(f) What percent of the test takers did better than her on the Verbal Reasoning section? On the Quantitative Reasoning section?
(g) Explain why simply comparing raw scores from the two sections could lead to an incorrect conclusion as to which section a student did better on.
(h) If the distributions of the scores on these exams are not nearly normal, would your answers to parts (b) - (f) change? Explain your reasoning.

4.4 Triathlon times, Part I. In triathlons, it is common for racers to be placed into age and gender groups. Friends Leo and Mary both completed the Hermosa Beach Triathlon, where Leo competed in the *Men, Ages 30 - 34* group while Mary competed in the *Women, Ages 25 - 29* group. Leo completed the race in 1:22:28 (4948 seconds), while Mary completed the race in 1:31:53 (5513 seconds). Obviously Leo finished faster, but they are curious about how they did within their respective groups. Can you help them? Here is some information on the performance of their groups:

- The finishing times of the *Men, Ages 30 - 34* group has a mean of 4313 seconds with a standard deviation of 583 seconds.
- The finishing times of the *Women, Ages 25 - 29* group has a mean of 5261 seconds with a standard deviation of 807 seconds.
- The distributions of finishing times for both groups are approximately Normal.

Remember: a better performance corresponds to a faster finish.

(a) Write down the short-hand for these two normal distributions.
(b) What are the Z-scores for Leo's and Mary's finishing times? What do these Z-scores tell you?
(c) Did Leo or Mary rank better in their respective groups? Explain your reasoning.
(d) What percent of the triathletes did Leo finish faster than in his group?
(e) What percent of the triathletes did Mary finish faster than in her group?
(f) If the distributions of finishing times are not nearly normal, would your answers to parts (b) - (e) change? Explain your reasoning.

4.5 GRE scores, Part II. In Exercise 4.3 we saw two distributions for GRE scores: $N(\mu = 151, \sigma = 7)$ for the verbal part of the exam and $N(\mu = 153, \sigma = 7.67)$ for the quantitative part. Use this information to compute each of the following:

(a) The score of a student who scored in the 80^{th} percentile on the Quantitative Reasoning section.
(b) The score of a student who scored worse than 70% of the test takers in the Verbal Reasoning section.

4.6 Triathlon times, Part II. In Exercise 4.4 we saw two distributions for triathlon times: $N(\mu = 4313, \sigma = 583)$ for *Men, Ages 30 - 34* and $N(\mu = 5261, \sigma = 807)$ for the *Women, Ages 25 - 29* group. Times are listed in seconds. Use this information to compute each of the following:

(a) The cutoff time for the fastest 5% of athletes in the men's group, i.e. those who took the shortest 5% of time to finish.

(b) The cutoff time for the slowest 10% of athletes in the women's group.

4.7 LA weather, Part I. The average daily high temperature in June in LA is 77°F with a standard deviation of 5°F. Suppose that the temperatures in June closely follow a normal distribution.

(a) What is the probability of observing an 83°F temperature or higher in LA during a randomly chosen day in June?

(b) How cool are the coldest 10% of the days (days with lowest average high temperature) during June in LA?

4.8 CAPM. The Capital Asset Pricing Model (CAPM) is a financial model that assumes returns on a portfolio are normally distributed. Suppose a portfolio has an average annual return of 14.7% (i.e. an average gain of 14.7%) with a standard deviation of 33%. A return of 0% means the value of the portfolio doesn't change, a negative return means that the portfolio loses money, and a positive return means that the portfolio gains money.

(a) What percent of years does this portfolio lose money, i.e. have a return less than 0%?

(b) What is the cutoff for the highest 15% of annual returns with this portfolio?

4.9 LA weather, Part II. Exercise 4.7 states that average daily high temperature in June in LA is 77°F with a standard deviation of 5°F, and it can be assumed that they to follow a normal distribution. We use the following equation to convert °F (Fahrenheit) to °C (Celsius):

$$C = (F - 32) \times \frac{5}{9}.$$

(a) Write the probability model for the distribution of temperature in °C in June in LA.

(b) What is the probability of observing a 28°C (which roughly corresponds to 83°F) temperature or higher in June in LA? Calculate using the °C model from part (a).

(c) Did you get the same answer or different answers in part (b) of this question and part (a) of Exercise 4.7? Are you surprised? Explain.

(d) Estimate the IQR of the temperatures (in °C) in June in LA.

4.10 Find the SD. Cholesterol levels for women aged 20 to 34 follow an approximately normal distribution with mean 185 milligrams per deciliter (mg/dl). Women with cholesterol levels above 220 mg/dl are considered to have high cholesterol and about 18.5% of women fall into this category. What is the standard deviation of the distribution of cholesterol levels for women aged 20 to 34?

4.2 Geometric distribution

How long should we expect to flip a coin until it turns up `heads`? Or how many times should we expect to roll a die until we get a `1`? These questions can be answered using the geometric distribution. We first formalize each trial – such as a single coin flip or die toss – using the Bernoulli distribution, and then we combine these with our tools from probability (Chapter 3) to construct the geometric distribution.

4.2.1 Bernoulli distribution

Many health insurance plans in the United States have a deductible, where the insured individual is responsible for costs up to the deductible, and then the costs above the deductible are shared between the individual and insurance company for the remainder of the year.

Suppose a health insurance company found that 70% of the people they insure stay below their deductible in any given year. Each of these people can be thought of as a **trial**. We label a person a **success** if her healthcare costs do not exceed the deductible. We label a person a **failure** if she does exceed her deductible in the year. Because 70% of the individuals will not hit their deductible, we denote the **probability of a success** as $p = 0.7$. The probability of a failure is sometimes denoted with $q = 1 - p$, which would be 0.3 for the insurance example.

When an individual trial only has two possible outcomes, often labeled as `success` or `failure`, it is called a **Bernoulli random variable**. We chose to label a person who does not hit her deductible as a "success" and all others as "failures". However, we could just as easily have reversed these labels. The mathematical framework we will build does not depend on which outcome is labeled a success and which a failure, as long as we are consistent.

Bernoulli random variables are often denoted as `1` for a success and `0` for a failure. In addition to being convenient in entering data, it is also mathematically handy. Suppose we observe ten trials:

`1 1 1 0 1 0 0 1 1 0`

Then the **sample proportion**, $\hat{p}$, is the sample mean of these observations:

$$\hat{p} = \frac{\text{\# of successes}}{\text{\# of trials}} = \frac{1+1+1+0+1+0+0+1+1+0}{10} = 0.6$$

This mathematical inquiry of Bernoulli random variables can be extended even further. Because `0` and `1` are numerical outcomes, we can define the mean and standard deviation of a Bernoulli random variable. (See Exercises 4.15 and 4.16.)

BERNOULLI RANDOM VARIABLE

If X is a random variable that takes value 1 with probability of success p and 0 with probability $1 - p$, then X is a Bernoulli random variable with mean and standard deviation

$$\mu = p \qquad\qquad \sigma = \sqrt{p(1-p)}$$

In general, it is useful to think about a Bernoulli random variable as a random process with only two outcomes: a success or failure. Then we build our mathematical framework using the numerical labels `1` and `0` for successes and failures, respectively.

4.2.2 Geometric distribution

The **geometric distribution** is used to describe how many trials it takes to observe a success. Let's first look at an example.

EXAMPLE 4.22

Suppose we are working at the insurance company and need to find a case where the person did not exceed her (or his) deductible as a case study. If the probability a person will not exceed her deductible is 0.7 and we are drawing people at random, what are the chances that the first person will not have exceeded her deductible, i.e. be a success? The second person? The third? What about we pull $n-1$ cases before we find the first success, i.e. the first success is the n^{th} person? (If the first success is the fifth person, then we say $n = 5$.)

E

The probability of stopping after the first person is just the chance the first person will not hit her (or his) deductible: 0.7. The probability the second person is the first to hit her deductible is

$$\begin{aligned}&P(\text{second person is the first to hit deductible})\\ &\quad = P(\text{the first won't, the second will}) = (0.3)(0.7) = 0.21\end{aligned}$$

Likewise, the probability it will be the third case is $(0.3)(0.3)(0.7) = 0.063$.

If the first success is on the n^{th} person, then there are $n-1$ failures and finally 1 success, which corresponds to the probability $(0.3)^{n-1}(0.7)$. This is the same as $(1-0.7)^{n-1}(0.7)$.

Example 4.22 illustrates what the **geometric distribution**, which describes the waiting time until a success for **independent and identically distributed (iid)** Bernoulli random variables. In this case, the *independence* aspect just means the individuals in the example don't affect each other, and *identical* means they each have the same probability of success.

The geometric distribution from Example 4.22 is shown in Figure 4.8. In general, the probabilities for a geometric distribution decrease **exponentially** fast.

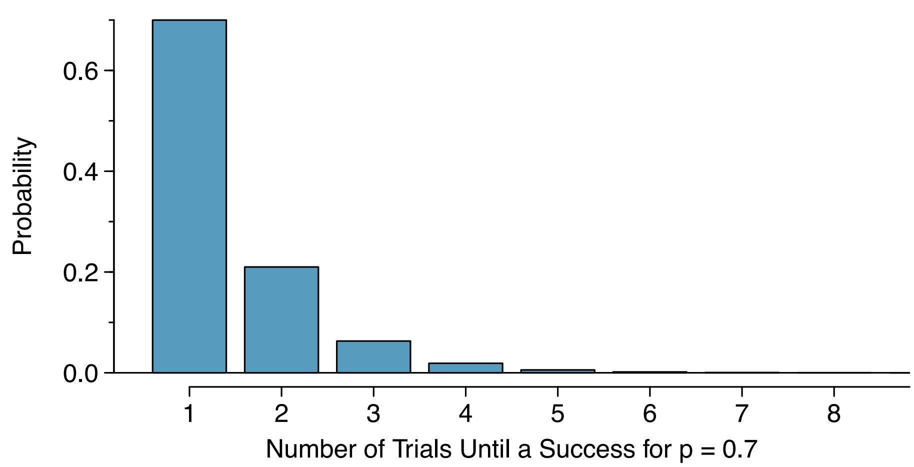


Figure 4.8: The geometric distribution when the probability of success is $p = 0.7$.

While this text will not derive the formulas for the mean (expected) number of trials needed to find the first success or the standard deviation or variance of this distribution, we present general formulas for each.

GEOMETRIC DISTRIBUTION

If the probability of a success in one trial is p and the probability of a failure is $1-p$, then the probability of finding the first success in the n^{th} trial is given by

$$(1-p)^{n-1}p$$

The mean (i.e. expected value), variance, and standard deviation of this wait time are given by

$$\mu = \frac{1}{p} \qquad \sigma^2 = \frac{1-p}{p^2} \qquad \sigma = \sqrt{\frac{1-p}{p^2}}$$

It is no accident that we use the symbol μ for both the mean and expected value. The mean and the expected value are one and the same.

It takes, on average, $1/p$ trials to get a success under the geometric distribution. This mathematical result is consistent with what we would expect intuitively. If the probability of a success is high (e.g. 0.8), then we don't usually wait very long for a success: $1/0.8 = 1.25$ trials on average. If the probability of a success is low (e.g. 0.1), then we would expect to view many trials before we see a success: $1/0.1 = 10$ trials.

G

GUIDED PRACTICE 4.23

The probability that a particular case would not exceed their deductible is said to be 0.7. If we were to examine cases until we found one that where the person did not hit her deductible, how many cases should we expect to check?[17]

E

EXAMPLE 4.24

What is the chance that we would find the first success within the first 3 cases?

This is the chance it is the first ($n = 1$), second ($n = 2$), or third ($n = 3$) case is the first success, which are three disjoint outcomes. Because the individuals in the sample are randomly sampled from a large population, they are independent. We compute the probability of each case and add the separate results:

$$\begin{aligned}
&P(n = 1, 2, \text{ or } 3) \\
&\quad = P(n = 1) + P(n = 2) + P(n = 3) \\
&\quad = (0.3)^{1-1}(0.7) + (0.3)^{2-1}(0.7) + (0.3)^{3-1}(0.7) \\
&\quad = 0.973
\end{aligned}$$

There is a probability of 0.973 that we would find a successful case within 3 cases.

G

GUIDED PRACTICE 4.25

Determine a more clever way to solve Example 4.24. Show that you get the same result.[18]

[17] We would expect to see about $1/0.7 \approx 1.43$ individuals to find the first success.

[18] First find the probability of the complement: $P(\text{no success in first 3 trials}) = 0.3^3 = 0.027$. Next, compute one minus this probability: $1 - P(\text{no success in 3 trials}) = 1 - 0.027 = 0.973$.

EXAMPLE 4.26

Suppose a car insurer has determined that 88% of its drivers will not exceed their deductible in a given year. If someone at the company were to randomly draw driver files until they found one that had not exceeded their deductible, what is the expected number of drivers the insurance employee must check? What is the standard deviation of the number of driver files that must be drawn?

In this example, a success is again when someone will not exceed the insurance deductible, which has probability $p = 0.88$. The expected number of people to be checked is $1/p = 1/0.88 = 1.14$ and the standard deviation is $\sqrt{(1-p)/p^2} = 0.39$.

GUIDED PRACTICE 4.27

Using the results from Example 4.26, $\mu = 1.14$ and $\sigma = 0.39$, would it be appropriate to use the normal model to find what proportion of experiments would end in 3 or fewer trials?[19]

The independence assumption is crucial to the geometric distribution's accurate description of a scenario. Mathematically, we can see that to construct the probability of the success on the n^{th} trial, we had to use the Multiplication Rule for Independent Processes. It is no simple task to generalize the geometric model for dependent trials.

[19] No. The geometric distribution is always right skewed and can never be well-approximated by the normal model.

Exercises

4.11 Is it Bernoulli? Determine if each trial can be considered an independent Bernoulli trial for the following situations.

(a) Cards dealt in a hand of poker.

(b) Outcome of each roll of a die.

4.12 With and without replacement. In the following situations assume that half of the specified population is male and the other half is female.

(a) Suppose you're sampling from a room with 10 people. What is the probability of sampling two females in a row when sampling with replacement? What is the probability when sampling without replacement?

(b) Now suppose you're sampling from a stadium with 10,000 people. What is the probability of sampling two females in a row when sampling with replacement? What is the probability when sampling without replacement?

(c) We often treat individuals who are sampled from a large population as independent. Using your findings from parts (a) and (b), explain whether or not this assumption is reasonable.

4.13 Eye color, Part I. A husband and wife both have brown eyes but carry genes that make it possible for their children to have brown eyes (probability 0.75), blue eyes (0.125), or green eyes (0.125).

(a) What is the probability the first blue-eyed child they have is their third child? Assume that the eye colors of the children are independent of each other.

(b) On average, how many children would such a pair of parents have before having a blue-eyed child? What is the standard deviation of the number of children they would expect to have until the first blue-eyed child?

4.14 Defective rate. A machine that produces a special type of transistor (a component of computers) has a 2% defective rate. The production is considered a random process where each transistor is independent of the others.

(a) What is the probability that the 10^{th} transistor produced is the first with a defect?

(b) What is the probability that the machine produces no defective transistors in a batch of 100?

(c) On average, how many transistors would you expect to be produced before the first with a defect? What is the standard deviation?

(d) Another machine that also produces transistors has a 5% defective rate where each transistor is produced independent of the others. On average how many transistors would you expect to be produced with this machine before the first with a defect? What is the standard deviation?

(e) Based on your answers to parts (c) and (d), how does increasing the probability of an event affect the mean and standard deviation of the wait time until success?

4.15 Bernoulli, the mean. Use the probability rules from Section 3.4 to derive the mean of a Bernoulli random variable, i.e. a random variable X that takes value 1 with probability p and value 0 with probability $1 - p$. That is, compute the expected value of a generic Bernoulli random variable.

4.16 Bernoulli, the standard deviation. Use the probability rules from Section 3.4 to derive the standard deviation of a Bernoulli random variable, i.e. a random variable X that takes value 1 with probability p and value 0 with probability $1 - p$. That is, compute the square root of the variance of a generic Bernoulli random variable.

4.3 Binomial distribution

The **binomial distribution** is used to describe the number of successes in a fixed number of trials. This is different from the geometric distribution, which described the number of trials we must wait before we observe a success.

4.3.1 The binomial distribution

Let's again imagine ourselves back at the insurance agency where 70% of individuals do not exceed their deductible.

E

EXAMPLE 4.28

Suppose the insurance agency is considering a random sample of four individuals they insure. What is the chance exactly one of them will exceed the deductible and the other three will not? Let's call the four people Ariana (A), Brittany (B), Carlton (C), and Damian (D) for convenience.

Let's consider a scenario where one person exceeds the deductible:

$$\begin{aligned}
&P(A = \texttt{exceed},\ B = \texttt{not},\ C = \texttt{not},\ D = \texttt{not}) \\
&\quad = P(A = \texttt{exceed})\ P(B = \texttt{not})\ P(C = \texttt{not})\ P(D = \texttt{not}) \\
&\quad = (0.3)(0.7)(0.7)(0.7) \\
&\quad = (0.7)^3(0.3)^1 \\
&\quad = 0.103
\end{aligned}$$

But there are three other scenarios: Brittany, Carlton, or Damian could have been the one to exceed the deductible. In each of these cases, the probability is again $(0.7)^3(0.3)^1$. These four scenarios exhaust all the possible ways that exactly one of these four people could have exceeded the deductible, so the total probability is $4 \times (0.7)^3(0.3)^1 = 0.412$.

G

GUIDED PRACTICE 4.29

Verify that the scenario where Brittany is the only one exceed the deductible has probability $(0.7)^3(0.3)^1$. [20]

The scenario outlined in Example 4.28 is an example of a binomial distribution scenario. The **binomial distribution** describes the probability of having exactly k successes in n independent Bernoulli trials with probability of a success p (in Example 4.28, $n = 4$, $k = 3$, $p = 0.7$). We would like to determine the probabilities associated with the binomial distribution more generally, i.e. we want a formula where we can use n, k, and p to obtain the probability. To do this, we reexamine each part of Example 4.28.

There were four individuals who could have been the one to exceed the deductible, and each of these four scenarios had the same probability. Thus, we could identify the final probability as

$$[\#\text{ of scenarios}] \times P(\text{single scenario})$$

The first component of this equation is the number of ways to arrange the $k = 3$ successes among the $n = 4$ trials. The second component is the probability of any of the four (equally probable) scenarios.

[20] $P(A = \texttt{not},\ B = \texttt{exceed},\ C = \texttt{not},\ D = \texttt{not}) = (0.7)(0.3)(0.7)(0.7) = (0.7)^3(0.3)^1$.

Consider $P(\text{single scenario})$ under the general case of k successes and $n-k$ failures in the n trials. In any such scenario, we apply the Multiplication Rule for independent events:

$$p^k(1-p)^{n-k}$$

This is our general formula for $P(\text{single scenario})$.

Secondly, we introduce a general formula for the number of ways to choose k successes in n trials, i.e. arrange k successes and $n-k$ failures:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

The quantity $\binom{n}{k}$ is read **n choose k**.[21] The exclamation point notation (e.g. $k!$) denotes a **factorial** expression.

$$\begin{aligned}
&0! = 1\\
&1! = 1\\
&2! = 2 \times 1 = 2\\
&3! = 3 \times 2 \times 1 = 6\\
&4! = 4 \times 3 \times 2 \times 1 = 24\\
&\vdots\\
&n! = n \times (n-1) \times \ldots \times 3 \times 2 \times 1
\end{aligned}$$

Using the formula, we can compute the number of ways to choose $k=3$ successes in $n=4$ trials:

$$\binom{4}{3} = \frac{4!}{3!(4-3)!} = \frac{4!}{3!1!} = \frac{4 \times 3 \times 2 \times 1}{(3 \times 2 \times 1)(1)} = 4$$

This result is exactly what we found by carefully thinking of each possible scenario in Example 4.28.

Substituting n choose k for the number of scenarios and $p^k(1-p)^{n-k}$ for the single scenario probability yields the general binomial formula.

BINOMIAL DISTRIBUTION

Suppose the probability of a single trial being a success is p. Then the probability of observing exactly k successes in n independent trials is given by

$$\binom{n}{k}p^k(1-p)^{n-k} = \frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}$$

The mean, variance, and standard deviation of the number of observed successes are

$$\mu = np \qquad \sigma^2 = np(1-p) \qquad \sigma = \sqrt{np(1-p)}$$

IS IT BINOMIAL? FOUR CONDITIONS TO CHECK.

(1) The trials are independent.
(2) The number of trials, n, is fixed.
(3) Each trial outcome can be classified as a *success* or *failure*.
(4) The probability of a success, p, is the same for each trial.

---

[21] Other notation for n choose k includes ${}_nC_k$, C_n^k, and $C(n,k)$.

EXAMPLE 4.30

What is the probability that 3 of 8 randomly selected individuals will have exceeded the insurance deductible, i.e. that 5 of 8 will not exceed the deductible? Recall that 70% of individuals will not exceed the deductible.

We would like to apply the binomial model, so we check the conditions. The number of trials is fixed ($n = 8$) (condition 2) and each trial outcome can be classified as a success or failure (condition 3). Because the sample is random, the trials are independent (condition 1) and the probability of a success is the same for each trial (condition 4).

E

In the outcome of interest, there are $k = 5$ successes in $n = 8$ trials (recall that a success is an individual who does *not* exceed the deductible, and the probability of a success is $p = 0.7$. So the probability that 5 of 8 will not exceed the deductible and 3 will exceed the deductible is given by

$$\begin{aligned} \binom{8}{5}(0.7)^5(1-0.7)^{8-5} &= \frac{8!}{5!(8-5)!}(0.7)^5(1-0.7)^{8-5} \\ &= \frac{8!}{5!3!}(0.7)^5(0.3)^3 \end{aligned}$$

Dealing with the factorial part:

$$\frac{8!}{5!3!} = \frac{8\times7\times6\times5\times4\times3\times2\times1}{(5\times4\times3\times2\times1)(3\times2\times1)} = \frac{8\times7\times6}{3\times2\times1} = 56$$

Using $(0.7)^5(0.3)^3 \approx 0.00454$, the final probability is about $56 \times 0.00454 \approx 0.254$.

COMPUTING BINOMIAL PROBABILITIES

The first step in using the binomial model is to check that the model is appropriate. The second step is to identify n, p, and k. As the last stage use software or the formulas to determine the probability, then interpret the results.

If you must do calculations by hand, it's often useful to cancel out as many terms as possible in the top and bottom of the binomial coefficient.

GUIDED PRACTICE 4.31

G

If we randomly sampled 40 case files from the insurance agency discussed earlier, how many of the cases would you expect to not have exceeded the deductible in a given year? What is the standard deviation of the number that would not have exceeded the deductible?[22]

GUIDED PRACTICE 4.32

G

The probability that a random smoker will develop a severe lung condition in his or her lifetime is about 0.3. If you have 4 friends who smoke, are the conditions for the binomial model satisfied?[23]

---

[22]We are asked to determine the expected number (the mean) and the standard deviation, both of which can be directly computed from the formulas: $\mu = np = 40 \times 0.7 = 28$ and $\sigma = \sqrt{np(1-p)} = \sqrt{40 \times 0.7 \times 0.3} = 2.9$. Because very roughly 95% of observations fall within 2 standard deviations of the mean (see Section 2.1.4), we would probably observe at least 22 but fewer than 34 individuals in our sample who would not exceed the deductible.

[23]One possible answer: if the friends know each other, then the independence assumption is probably not satisfied. For example, acquaintances may have similar smoking habits, or those friends might make a pact to quit together.

GUIDED PRACTICE 4.33

Suppose these four friends do not know each other and we can treat them as if they were a random sample from the population. Is the binomial model appropriate? What is the probability that[24]

(a) None of them will develop a severe lung condition?

(b) One will develop a severe lung condition?

(c) That no more than one will develop a severe lung condition?

GUIDED PRACTICE 4.34

What is the probability that at least 2 of your 4 smoking friends will develop a severe lung condition in their lifetimes?[25]

GUIDED PRACTICE 4.35

Suppose you have 7 friends who are smokers and they can be treated as a random sample of smokers.[26]

(a) How many would you expect to develop a severe lung condition, i.e. what is the mean?

(b) What is the probability that at most 2 of your 7 friends will develop a severe lung condition.

Next we consider the first term in the binomial probability, n choose k under some special scenarios.

GUIDED PRACTICE 4.36

Why is it true that $\binom{n}{0} = 1$ and $\binom{n}{n} = 1$ for any number n?[27]

GUIDED PRACTICE 4.37

How many ways can you arrange one success and $n-1$ failures in n trials? How many ways can you arrange $n-1$ successes and one failure in n trials?[28]

---

[24]To check if the binomial model is appropriate, we must verify the conditions. (i) Since we are supposing we can treat the friends as a random sample, they are independent. (ii) We have a fixed number of trials ($n = 4$). (iii) Each outcome is a success or failure. (iv) The probability of a success is the same for each trials since the individuals are like a random sample ($p = 0.3$ if we say a "success" is someone getting a lung condition, a morbid choice). Compute parts (a) and (b) using the binomial formula: $P(0) = \binom{4}{0}(0.3)^0(0.7)^4 = 1 \times 1 \times 0.7^4 = 0.2401$, $P(1) = \binom{4}{1}(0.3)^1(0.7)^3 = 0.4116$. Note: $0! = 1$. Part (c) can be computed as the sum of parts (a) and (b): $P(0) + P(1) = 0.2401 + 0.4116 = 0.6517$. That is, there is about a 65% chance that no more than one of your four smoking friends will develop a severe lung condition.

[25]The complement (no more than one will develop a severe lung condition) as computed in Guided Practice 4.33 as 0.6517, so we compute one minus this value: 0.3483.

[26](a) $\mu = 0.3 \times 7 = 2.1$. (b) $P(0, 1, \text{ or } 2 \text{ develop severe lung condition}) = P(k = 0) + P(k = 1) + P(k = 2) = 0.6471$.

[27]Frame these expressions into words. How many different ways are there to arrange 0 successes and n failures in n trials? (1 way.) How many different ways are there to arrange n successes and 0 failures in n trials? (1 way.)

[28]One success and $n-1$ failures: there are exactly n unique places we can put the success, so there are n ways to arrange one success and $n-1$ failures. A similar argument is used for the second question. Mathematically, we show these results by verifying the following two equations:

$$\binom{n}{1} = n, \qquad \binom{n}{n-1} = n$$

4.3.2 Normal approximation to the binomial distribution

The binomial formula is cumbersome when the sample size (n) is large, particularly when we consider a range of observations. In some cases we may use the normal distribution as an easier and faster way to estimate binomial probabilities.

E

EXAMPLE 4.38

Approximately 15% of the US population smokes cigarettes. A local government believed their community had a lower smoker rate and commissioned a survey of 400 randomly selected individuals. The survey found that only 42 of the 400 participants smoke cigarettes. If the true proportion of smokers in the community was really 15%, what is the probability of observing 42 or fewer smokers in a sample of 400 people?

We leave the usual verification that the four conditions for the binomial model are valid as an exercise.

The question posed is equivalent to asking, what is the probability of observing $k = 0$, 1, 2, ..., or 42 smokers in a sample of $n = 400$ when $p = 0.15$? We can compute these 43 different probabilities and add them together to find the answer:

$$\begin{aligned} &P(k = 0 \text{ or } k = 1 \text{ or } \cdots \text{ or } k = 42) \\ &\quad = P(k = 0) + P(k = 1) + \cdots + P(k = 42) \\ &\quad = 0.0054 \end{aligned}$$

If the true proportion of smokers in the community is $p = 0.15$, then the probability of observing 42 or fewer smokers in a sample of $n = 400$ is 0.0054.

The computations in Example 4.38 are tedious and long. In general, we should avoid such work if an alternative method exists that is faster, easier, and still accurate. Recall that calculating probabilities of a range of values is much easier in the normal model. We might wonder, is it reasonable to use the normal model in place of the binomial distribution? Surprisingly, yes, if certain conditions are met.

G

GUIDED PRACTICE 4.39

Here we consider the binomial model when the probability of a success is $p = 0.10$. Figure 4.9 shows four hollow histograms for simulated samples from the binomial distribution using four different sample sizes: $n = 10, 30, 100, 300$. What happens to the shape of the distributions as the sample size increases? What distribution does the last hollow histogram resemble?[29]

NORMAL APPROXIMATION OF THE BINOMIAL DISTRIBUTION

The binomial distribution with probability of success p is nearly normal when the sample size n is sufficiently large that np and $n(1 - p)$ are both at least 10. The approximate normal distribution has parameters corresponding to the mean and standard deviation of the binomial distribution:

$$\mu = np \qquad\qquad \sigma = \sqrt{np(1-p)}$$

The normal approximation may be used when computing the range of many possible successes. For instance, we may apply the normal distribution to the setting of Example 4.38.

[29] The distribution is transformed from a blocky and skewed distribution into one that rather resembles the normal distribution in last hollow histogram.

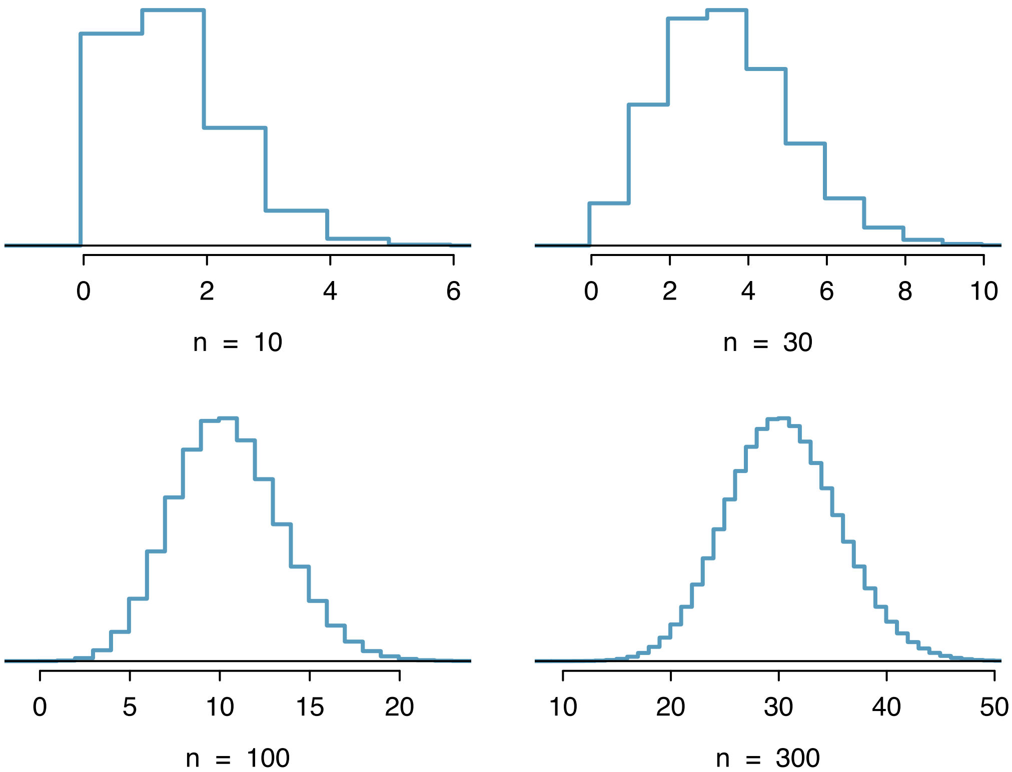


Figure 4.9: Hollow histograms of samples from the binomial model when $p = 0.10$. The sample sizes for the four plots are $n = 10$, 30, 100, and 300, respectively.

EXAMPLE 4.40

How can we use the normal approximation to estimate the probability of observing 42 or fewer smokers in a sample of 400, if the true proportion of smokers is $p = 0.15$?

Showing that the binomial model is reasonable was a suggested exercise in Example 4.38. We also verify that both np and $n(1 - p)$ are at least 10:

$$np = 400 \times 0.15 = 60 \qquad n(1 - p) = 400 \times 0.85 = 340$$

With these conditions checked, we may use the normal approximation in place of the binomial distribution using the mean and standard deviation from the binomial model:

$$\mu = np = 60 \qquad \sigma = \sqrt{np(1 - p)} = 7.14$$

We want to find the probability of observing 42 or fewer smokers using this model.

GUIDED PRACTICE 4.41

Use the normal model $N(\mu = 60, \sigma = 7.14)$ to estimate the probability of observing 42 or fewer smokers. Your answer should be approximately equal to the solution of Example 4.38: 0.0054. [30]

[30] Compute the Z-score first: $Z = \frac{42-60}{7.14} = -2.52$. The corresponding left tail area is 0.0059.

4.3.3 The normal approximation breaks down on small intervals

The normal approximation to the binomial distribution tends to perform poorly when estimating the probability of a small range of counts, even when the conditions are met.

Suppose we wanted to compute the probability of observing 49, 50, or 51 smokers in 400 when $p = 0.15$. With such a large sample, we might be tempted to apply the normal approximation and use the range 49 to 51. However, we would find that the binomial solution and the normal approximation notably differ:

Binomial: 0.0649 Normal: 0.0421

We can identify the cause of this discrepancy using Figure 4.10, which shows the areas representing the binomial probability (outlined) and normal approximation (shaded). Notice that the width of the area under the normal distribution is 0.5 units too slim on both sides of the interval.

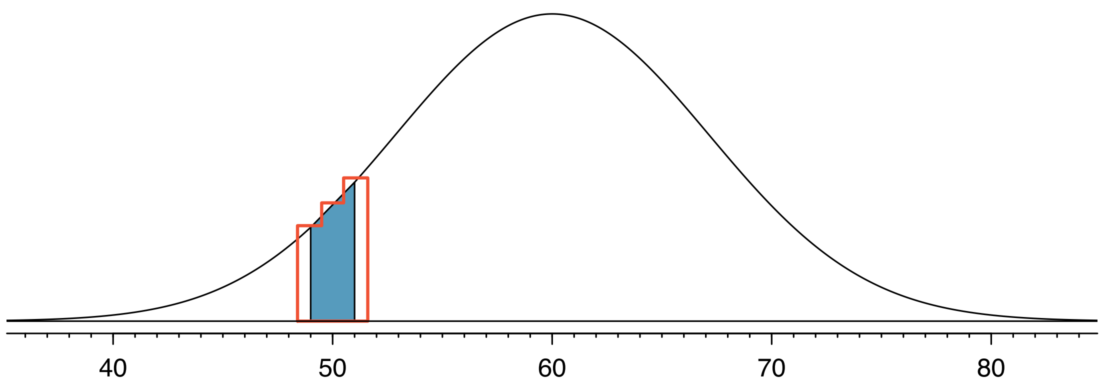


Figure 4.10: A normal curve with the area between 49 and 51 shaded. The outlined area represents the exact binomial probability.

IMPROVING THE NORMAL APPROXIMATION FOR THE BINOMIAL DISTRIBUTION

The normal approximation to the binomial distribution for intervals of values is usually improved if cutoff values are modified slightly. The cutoff values for the lower end of a shaded region should be reduced by 0.5, and the cutoff value for the upper end should be increased by 0.5.

The tip to add extra area when applying the normal approximation is most often useful when examining a range of observations. In the example above, the revised normal distribution estimate is 0.0633, much closer to the exact value of 0.0649. While it is possible to also apply this correction when computing a tail area, the benefit of the modification usually disappears since the total interval is typically quite wide.

Exercises

4.17 Underage drinking, Part I. Data collected by the Substance Abuse and Mental Health Services Administration (SAMSHA) suggests that 69.7% of 18-20 year olds consumed alcoholic beverages in any given year.[31]

(a) Suppose a random sample of ten 18-20 year olds is taken. Is the use of the binomial distribution appropriate for calculating the probability that exactly six consumed alcoholic beverages? Explain.

(b) Calculate the probability that exactly 6 out of 10 randomly sampled 18- 20 year olds consumed an alcoholic drink.

(c) What is the probability that exactly four out of ten 18-20 year olds have *not* consumed an alcoholic beverage?

(d) What is the probability that at most 2 out of 5 randomly sampled 18-20 year olds have consumed alcoholic beverages?

(e) What is the probability that at least 1 out of 5 randomly sampled 18-20 year olds have consumed alcoholic beverages?

4.18 Chicken pox, Part I. The National Vaccine Information Center estimates that 90% of Americans have had chickenpox by the time they reach adulthood.[32]

(a) Suppose we take a random sample of 100 American adults. Is the use of the binomial distribution appropriate for calculating the probability that exactly 97 out of 100 randomly sampled American adults had chickenpox during childhood? Explain.

(b) Calculate the probability that exactly 97 out of 100 randomly sampled American adults had chickenpox during childhood.

(c) What is the probability that exactly 3 out of a new sample of 100 American adults have *not* had chickenpox in their childhood?

(d) What is the probability that at least 1 out of 10 randomly sampled American adults have had chickenpox?

(e) What is the probability that at most 3 out of 10 randomly sampled American adults have *not* had chickenpox?

4.19 Underage drinking, Part II. We learned in Exercise 4.17 that about 70% of 18-20 year olds consumed alcoholic beverages in any given year. We now consider a random sample of fifty 18-20 year olds.

(a) How many people would you expect to have consumed alcoholic beverages? And with what standard deviation?

(b) Would you be surprised if there were 45 or more people who have consumed alcoholic beverages?

(c) What is the probability that 45 or more people in this sample have consumed alcoholic beverages? How does this probability relate to your answer to part (b)?

4.20 Chickenpox, Part II. We learned in Exercise 4.18 that about 90% of American adults had chickenpox before adulthood. We now consider a random sample of 120 American adults.

(a) How many people in this sample would you expect to have had chickenpox in their childhood? And with what standard deviation?

(b) Would you be surprised if there were 105 people who have had chickenpox in their childhood?

(c) What is the probability that 105 or fewer people in this sample have had chickenpox in their childhood? How does this probability relate to your answer to part (b)?

4.21 Game of dreidel. A dreidel is a four-sided spinning top with the Hebrew letters *nun*, *gimel*, *hei*, and *shin*, one on each side. Each side is equally likely to come up in a single spin of the dreidel. Suppose you spin a dreidel three times. Calculate the probability of getting

(a) at least one *nun*?

(b) exactly 2 *nuns*?

(c) exactly 1 *hei*?

(d) at most 2 *gimels*?

Photo by Staccabees, cropped (http://flic.kr/p/7gLZTf) CC BY 2.0 license

[31]SAMHSA, Office of Applied Studies, National Survey on Drug Use and Health, 2007 and 2008.
[32]National Vaccine Information Center, Chickenpox, The Disease & The Vaccine Fact Sheet.

4.22 Arachnophobia. A Gallup Poll found that 7% of teenagers (ages 13 to 17) suffer from arachnophobia and are extremely afraid of spiders. At a summer camp there are 10 teenagers sleeping in each tent. Assume that these 10 teenagers are independent of each other.[33]

(a) Calculate the probability that at least one of them suffers from arachnophobia.

(b) Calculate the probability that exactly 2 of them suffer from arachnophobia.

(c) Calculate the probability that at most 1 of them suffers from arachnophobia.

(d) If the camp counselor wants to make sure no more than 1 teenager in each tent is afraid of spiders, does it seem reasonable for him to randomly assign teenagers to tents?

4.23 Eye color, Part II. Exercise 4.13 introduces a husband and wife with brown eyes who have 0.75 probability of having children with brown eyes, 0.125 probability of having children with blue eyes, and 0.125 probability of having children with green eyes.

(a) What is the probability that their first child will have green eyes and the second will not?

(b) What is the probability that exactly one of their two children will have green eyes?

(c) If they have six children, what is the probability that exactly two will have green eyes?

(d) If they have six children, what is the probability that at least one will have green eyes?

(e) What is the probability that the first green eyed child will be the 4^{th} child?

(f) Would it be considered unusual if only 2 out of their 6 children had brown eyes?

4.24 Sickle cell anemia. Sickle cell anemia is a genetic blood disorder where red blood cells lose their flexibility and assume an abnormal, rigid, “sickle” shape, which results in a risk of various complications. If both parents are carriers of the disease, then a child has a 25% chance of having the disease, 50% chance of being a carrier, and 25% chance of neither having the disease nor being a carrier. If two parents who are carriers of the disease have 3 children, what is the probability that

(a) two will have the disease?

(b) none will have the disease?

(c) at least one will neither have the disease nor be a carrier?

(d) the first child with the disease will the be 3^{rd} child?

4.25 Exploring permutations. The formula for the number of ways to arrange n objects is $n! = n \times (n-1) \times \cdots \times 2 \times 1$. This exercise walks you through the derivation of this formula for a couple of special cases.

A small company has five employees: Anna, Ben, Carl, Damian, and Eddy. There are five parking spots in a row at the company, none of which are assigned, and each day the employees pull into a random parking spot. That is, all possible orderings of the cars in the row of spots are equally likely.

(a) On a given day, what is the probability that the employees park in alphabetical order?

(b) If the alphabetical order has an equal chance of occurring relative to all other possible orderings, how many ways must there be to arrange the five cars?

(c) Now consider a sample of 8 employees instead. How many possible ways are there to order these 8 employees' cars?

4.26 Male children. While it is often assumed that the probabilities of having a boy or a girl are the same, the actual probability of having a boy is slightly higher at 0.51. Suppose a couple plans to have 3 kids.

(a) Use the binomial model to calculate the probability that two of them will be boys.

(b) Write out all possible orderings of 3 children, 2 of whom are boys. Use these scenarios to calculate the same probability from part (a) but using the addition rule for disjoint outcomes. Confirm that your answers from parts (a) and (b) match.

(c) If we wanted to calculate the probability that a couple who plans to have 8 kids will have 3 boys, briefly describe why the approach from part (b) would be more tedious than the approach from part (a).

[33] Gallup Poll, What Frightens America's Youth?, March 29, 2005.

4.4 Negative binomial distribution

The geometric distribution describes the probability of observing the first success on the n^{th} trial. The **negative binomial distribution** is more general: it describes the probability of observing the k^{th} success on the n^{th} trial.

E

EXAMPLE 4.42

Each day a high school football coach tells his star kicker, Brian, that he can go home after he successfully kicks four 35 yard field goals. Suppose we say each kick has a probability p of being successful. If p is small – e.g. close to 0.1 – would we expect Brian to need many attempts before he successfully kicks his fourth field goal?

We are waiting for the fourth success ($k = 4$). If the probability of a success (p) is small, then the number of attempts (n) will probably be large. This means that Brian is more likely to need many attempts before he gets $k = 4$ successes. To put this another way, the probability of n being small is low.

To identify a negative binomial case, we check 4 conditions. The first three are common to the binomial distribution.

IS IT NEGATIVE BINOMIAL? FOUR CONDITIONS TO CHECK

(1) The trials are independent.
(2) Each trial outcome can be classified as a success or failure.
(3) The probability of a success (p) is the same for each trial.
(4) The last trial must be a success.

G

GUIDED PRACTICE 4.43

Suppose Brian is very diligent in his attempts and he makes each 35 yard field goal with probability $p = 0.8$. Take a guess at how many attempts he would need before making his fourth kick.[34]

E

EXAMPLE 4.44

In yesterday's practice, it took Brian only 6 tries to get his fourth field goal. Write out each of the possible sequence of kicks.

Because it took Brian six tries to get the fourth success, we know the last kick must have been a success. That leaves three successful kicks and two unsuccessful kicks (we label these as failures) that make up the first five attempts. There are ten possible sequences of these first five kicks, which are shown in Figure 4.11. If Brian achieved his fourth success ($k = 4$) on his sixth attempt ($n = 6$), then his order of successes and failures must be one of these ten possible sequences.

G

GUIDED PRACTICE 4.45

Each sequence in Figure 4.11 has exactly two failures and four successes with the last attempt always being a success. If the probability of a success is $p = 0.8$, find the probability of the first sequence.[35]

[34]One possible answer: since he is likely to make each field goal attempt, it will take him at least 4 attempts but probably not more than 6 or 7.

[35]The first sequence: $0.2 \times 0.2 \times 0.8 \times 0.8 \times 0.8 \times 0.8 = 0.0164$.

Kick Attempt

	1	2	3	4	5	6
1	F	F	$\overset{1}{S}$	$\overset{2}{S}$	$\overset{3}{S}$	$\overset{4}{S}$
2	F	$\overset{1}{S}$	F	$\overset{2}{S}$	$\overset{3}{S}$	$\overset{4}{S}$
3	F	$\overset{1}{S}$	$\overset{2}{S}$	F	$\overset{3}{S}$	$\overset{4}{S}$
4	F	$\overset{1}{S}$	$\overset{2}{S}$	$\overset{3}{S}$	F	$\overset{4}{S}$
5	$\overset{1}{S}$	F	F	$\overset{2}{S}$	$\overset{3}{S}$	$\overset{4}{S}$
6	$\overset{1}{S}$	F	$\overset{2}{S}$	F	$\overset{3}{S}$	$\overset{4}{S}$
7	$\overset{1}{S}$	F	$\overset{2}{S}$	$\overset{3}{S}$	F	$\overset{4}{S}$
8	$\overset{1}{S}$	$\overset{2}{S}$	F	F	$\overset{3}{S}$	$\overset{4}{S}$
9	$\overset{1}{S}$	$\overset{2}{S}$	F	$\overset{3}{S}$	F	$\overset{4}{S}$
10	$\overset{1}{S}$	$\overset{2}{S}$	$\overset{3}{S}$	F	F	$\overset{4}{S}$

Figure 4.11: The ten possible sequences when the fourth successful kick is on the sixth attempt.

If the probability Brian kicks a 35 yard field goal is $p = 0.8$, what is the probability it takes Brian exactly six tries to get his fourth successful kick? We can write this as

$$\begin{aligned}
&P(\text{it takes Brian six tries to make four field goals})\\
&\quad = P(\text{Brian makes three of his first five field goals, and he makes the sixth one})\\
&\quad = P(1^{st} \text{ sequence OR } 2^{nd} \text{ sequence OR ... OR } 10^{th} \text{ sequence})
\end{aligned}$$

where the sequences are from Figure 4.11. We can break down this last probability into the sum of ten disjoint possibilities:

$$\begin{aligned}
&P(1^{st} \text{ sequence OR } 2^{nd} \text{ sequence OR ... OR } 10^{th} \text{ sequence})\\
&\quad = P(1^{st} \text{ sequence}) + P(2^{nd} \text{ sequence}) + \cdots + P(10^{th} \text{ sequence})
\end{aligned}$$

The probability of the first sequence was identified in Guided Practice 4.45 as 0.0164, and each of the other sequences have the same probability. Since each of the ten sequence has the same probability, the total probability is ten times that of any individual sequence.

The way to compute this negative binomial probability is similar to how the binomial problems were solved in Section 4.3. The probability is broken into two pieces:

$$\begin{aligned}
&P(\text{it takes Brian six tries to make four field goals})\\
&\quad = [\text{Number of possible sequences}] \times P(\text{Single sequence})
\end{aligned}$$

Each part is examined separately, then we multiply to get the final result.

We first identify the probability of a single sequence. One particular case is to first observe all the failures ($n - k$ of them) followed by the k successes:

$$\begin{aligned}
&P(\text{Single sequence})\\
&\quad = P(n - k \text{ failures and then } k \text{ successes})\\
&\quad = (1 - p)^{n-k} p^k
\end{aligned}$$

We must also identify the number of sequences for the general case. Above, ten sequences were identified where the fourth success came on the sixth attempt. These sequences were identified by fixing the last observation as a success and looking for all the ways to arrange the other observations. In other words, how many ways could we arrange $k-1$ successes in $n-1$ trials? This can be found using the n choose k coefficient but for $n-1$ and $k-1$ instead:

$$\binom{n-1}{k-1} = \frac{(n-1)!}{(k-1)!\,((n-1)-(k-1))!} = \frac{(n-1)!}{(k-1)!\,(n-k)!}$$

This is the number of different ways we can order $k-1$ successes and $n-k$ failures in $n-1$ trials. If the factorial notation (the exclamation point) is unfamiliar, see page 150.

NEGATIVE BINOMIAL DISTRIBUTION

The negative binomial distribution describes the probability of observing the k^{th} success on the n^{th} trial, where all trials are independent:

$$P(\text{the } k^{th} \text{ success on the } n^{th} \text{ trial}) = \binom{n-1}{k-1} p^k (1-p)^{n-k}$$

The value p represents the probability that an individual trial is a success.

E

EXAMPLE 4.46

Show using the formula for the negative binomial distribution that the probability Brian kicks his fourth successful field goal on the sixth attempt is 0.164.

The probability of a single success is $p = 0.8$, the number of successes is $k = 4$, and the number of necessary attempts under this scenario is $n = 6$.

$$\binom{n-1}{k-1} p^k (1-p)^{n-k} \;=\; \frac{5!}{3!2!}(0.8)^4(0.2)^2 \;=\; 10 \times 0.0164 \;=\; 0.164$$

G

GUIDED PRACTICE 4.47

The negative binomial distribution requires that each kick attempt by Brian is independent. Do you think it is reasonable to suggest that each of Brian's kick attempts are independent?[36]

G

GUIDED PRACTICE 4.48

Assume Brian's kick attempts are independent. What is the probability that Brian will kick his fourth field goal within 5 attempts?[37]

[36]Answers may vary. We cannot conclusively say they are or are not independent. However, many statistical reviews of athletic performance suggests such attempts are very nearly independent.

[37]If his fourth field goal ($k = 4$) is within five attempts, it either took him four or five tries ($n = 4$ or $n = 5$). We have $p = 0.8$ from earlier. Use the negative binomial distribution to compute the probability of $n = 4$ tries and $n = 5$ tries, then add those probabilities together:

$$\begin{aligned} &P(n = 4 \text{ OR } n = 5) = P(n = 4) + P(n = 5) \\ &\quad = \binom{4-1}{4-1} 0.8^4 + \binom{5-1}{4-1}(0.8)^4(1-0.8) = 1 \times 0.41 + 4 \times 0.082 = 0.41 + 0.33 = 0.74 \end{aligned}$$

BINOMIAL VERSUS NEGATIVE BINOMIAL

In the binomial case, we typically have a fixed number of trials and instead consider the number of successes. In the negative binomial case, we examine how many trials it takes to observe a fixed number of successes and require that the last observation be a success.

GUIDED PRACTICE 4.49

On 70% of days, a hospital admits at least one heart attack patient. On 30% of the days, no heart attack patients are admitted. Identify each case below as a binomial or negative binomial case, and compute the probability.[38]

G

(a) What is the probability the hospital will admit a heart attack patient on exactly three days this week?

(b) What is the probability the second day with a heart attack patient will be the fourth day of the week?

(c) What is the probability the fifth day of next month will be the first day with a heart attack patient?

[38] In each part, $p = 0.7$. (a) The number of days is fixed, so this is binomial. The parameters are $k = 3$ and $n = 7$: 0.097. (b) The last "success" (admitting a heart attack patient) is fixed to the last day, so we should apply the negative binomial distribution. The parameters are $k = 2$, $n = 4$: 0.132. (c) This problem is negative binomial with $k = 1$ and $n = 5$: 0.006. Note that the negative binomial case when $k = 1$ is the same as using the geometric distribution.

Exercises

4.27 Rolling a die. Calculate the following probabilities and indicate which probability distribution model is appropriate in each case. You roll a fair die 5 times. What is the probability of rolling

(a) the first 6 on the fifth roll?
(b) exactly three 6s?
(c) the third 6 on the fifth roll?

4.28 Playing darts. Calculate the following probabilities and indicate which probability distribution model is appropriate in each case. A very good darts player can hit the bull's eye (red circle in the center of the dart board) 65% of the time. What is the probability that he

(a) hits the bullseye for the 10^{th} time on the 15^{th} try?
(b) hits the bullseye 10 times in 15 tries?
(c) hits the first bullseye on the third try?

4.29 Sampling at school. For a sociology class project you are asked to conduct a survey on 20 students at your school. You decide to stand outside of your dorm's cafeteria and conduct the survey on a random sample of 20 students leaving the cafeteria after dinner one evening. Your dorm is comprised of 45% males and 55% females.

(a) Which probability model is most appropriate for calculating the probability that the 4^{th} person you survey is the 2^{nd} female? Explain.
(b) Compute the probability from part (a).
(c) The three possible scenarios that lead to 4^{th} person you survey being the 2^{nd} female are

$$\{M, M, F, F\}, \{M, F, M, F\}, \{F, M, M, F\}$$

One common feature among these scenarios is that the last trial is always female. In the first three trials there are 2 males and 1 female. Use the binomial coefficient to confirm that there are 3 ways of ordering 2 males and 1 female.

(d) Use the findings presented in part (c) to explain why the formula for the coefficient for the negative binomial is $\binom{n-1}{k-1}$ while the formula for the binomial coefficient is $\binom{n}{k}$.

4.30 Serving in volleyball. A not-so-skilled volleyball player has a 15% chance of making the serve, which involves hitting the ball so it passes over the net on a trajectory such that it will land in the opposing team's court. Suppose that her serves are independent of each other.

(a) What is the probability that on the 10^{th} try she will make her 3^{rd} successful serve?
(b) Suppose she has made two successful serves in nine attempts. What is the probability that her 10^{th} serve will be successful?
(c) Even though parts (a) and (b) discuss the same scenario, the probabilities you calculated should be different. Can you explain the reason for this discrepancy?

4.5 Poisson distribution

EXAMPLE 4.50

There are about 8 million individuals in New York City. How many individuals might we expect to be hospitalized for acute myocardial infarction (AMI), i.e. a heart attack, each day? According to historical records, the average number is about 4.4 individuals. However, we would also like to know the approximate distribution of counts. What would a histogram of the number of AMI occurrences each day look like if we recorded the daily counts over an entire year?

A histogram of the number of occurrences of AMI on 365 days for NYC is shown in Figure 4.12.[39] The sample mean (4.38) is similar to the historical average of 4.4. The sample standard deviation is about 2, and the histogram indicates that about 70% of the data fall between 2.4 and 6.4. The distribution's shape is unimodal and skewed to the right.

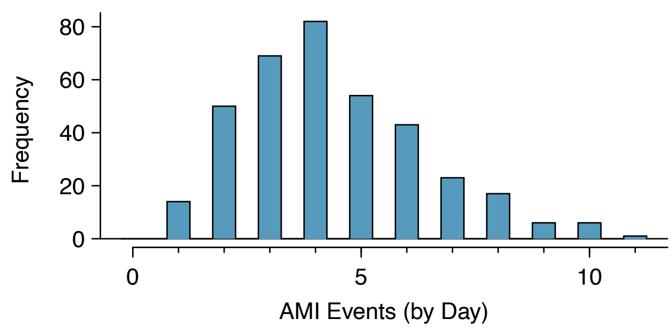


Figure 4.12: A histogram of the number of occurrences of AMI on 365 separate days in NYC.

The **Poisson distribution** is often useful for estimating the number of events in a large population over a unit of time. For instance, consider each of the following events:

- having a heart attack,
- getting married, and
- getting struck by lightning.

The Poisson distribution helps us describe the number of such events that will occur in a day for a fixed population if the individuals within the population are independent. The Poisson distribution could also be used over another unit of time, such as an hour or a week.

The histogram in Figure 4.12 approximates a Poisson distribution with rate equal to 4.4. The **rate** for a Poisson distribution is the average number of occurrences in a mostly-fixed population per unit of time. In Example 4.50, the time unit is a day, the population is all New York City residents, and the historical rate is 4.4. The parameter in the Poisson distribution is the rate – or how many events we expect to observe – and it is typically denoted by λ (the Greek letter *lambda*) or μ. Using the rate, we can describe the probability of observing exactly k events in a single unit of time.

---

[39]These data are simulated. In practice, we should check for an association between successive days.

POISSON DISTRIBUTION

Suppose we are watching for events and the number of observed events follows a Poisson distribution with rate λ. Then

$$P(\text{observe } k \text{ events}) = \frac{\lambda^k e^{-\lambda}}{k!}$$

where k may take a value 0, 1, 2, and so on, and $k!$ represents k-factorial, as described on page 150. The letter $e \approx 2.718$ is the base of the natural logarithm. The mean and standard deviation of this distribution are λ and $\sqrt{\lambda}$, respectively.

We will leave a rigorous set of conditions for the Poisson distribution to a later course. However, we offer a few simple guidelines that can be used for an initial evaluation of whether the Poisson model would be appropriate.

A random variable may follow a Poisson distribution if we are looking for the number of events, the population that generates such events is large, and the events occur independently of each other.

Even when events are not really independent – for instance, Saturdays and Sundays are especially popular for weddings – a Poisson model may sometimes still be reasonable if we allow it to have a different rate for different times. In the wedding example, the rate would be modeled as higher on weekends than on weekdays. The idea of modeling rates for a Poisson distribution against a second variable such as the day of week forms the foundation of some more advanced methods that fall in the realm of **generalized linear models**. In Chapters 8 and 9, we will discuss a foundation of linear models.

Exercises

4.31 Customers at a coffee shop. A coffee shop serves an average of 75 customers per hour during the morning rush.

(a) Which distribution have we studied that is most appropriate for calculating the probability of a given number of customers arriving within one hour during this time of day?
(b) What are the mean and the standard deviation of the number of customers this coffee shop serves in one hour during this time of day?
(c) Would it be considered unusually low if only 60 customers showed up to this coffee shop in one hour during this time of day?
(d) Calculate the probability that this coffee shop serves 70 customers in one hour during this time of day.

4.32 Stenographer's typos. A very skilled court stenographer makes one typographical error (typo) per hour on average.

(a) What probability distribution is most appropriate for calculating the probability of a given number of typos this stenographer makes in an hour?
(b) What are the mean and the standard deviation of the number of typos this stenographer makes?
(c) Would it be considered unusual if this stenographer made 4 typos in a given hour?
(d) Calculate the probability that this stenographer makes at most 2 typos in a given hour.

4.33 How many cars show up? For Monday through Thursday when there isn't a holiday, the average number of vehicles that visit a particular retailer between 2pm and 3pm each afternoon is 6.5, and the number of cars that show up on any given day follows a Poisson distribution.

(a) What is the probability that exactly 5 cars will show up next Monday?
(b) What is the probability that 0, 1, or 2 cars will show up next Monday between 2pm and 3pm?
(c) There is an average of 11.7 people who visit during those same hours from vehicles. Is it likely that the number of people visiting by car during this hour is also Poisson? Explain.

4.34 Lost baggage. Occasionally an airline will lose a bag. Suppose a small airline has found it can reasonably model the number of bags lost each weekday using a Poisson model with a mean of 2.2 bags.

(a) What is the probability that the airline will lose no bags next Monday?
(b) What is the probability that the airline will lose 0, 1, or 2 bags on next Monday?
(c) Suppose the airline expands over the course of the next 3 years, doubling the number of flights it makes, and the CEO asks you if it's reasonable for them to continue using the Poisson model with a mean of 2.2. What is an appropriate recommendation? Explain.

Chapter exercises

4.35 Roulette winnings. In the game of roulette, a wheel is spun and you place bets on where it will stop. One popular bet is that it will stop on a red slot; such a bet has an 18/38 chance of winning. If it stops on red, you double the money you bet. If not, you lose the money you bet. Suppose you play 3 times, each time with a $1 bet. Let Y represent the total amount won or lost. Write a probability model for Y.

4.36 Speeding on the I-5, Part I. The distribution of passenger vehicle speeds traveling on the Interstate 5 Freeway (I-5) in California is nearly normal with a mean of 72.6 miles/hour and a standard deviation of 4.78 miles/hour.[40]

(a) What percent of passenger vehicles travel slower than 80 miles/hour?
(b) What percent of passenger vehicles travel between 60 and 80 miles/hour?
(c) How fast do the fastest 5% of passenger vehicles travel?
(d) The speed limit on this stretch of the I-5 is 70 miles/hour. Approximate what percentage of the passenger vehicles travel above the speed limit on this stretch of the I-5.

4.37 University admissions. Suppose a university announced that it admitted 2,500 students for the following year's freshman class. However, the university has dorm room spots for only 1,786 freshman students. If there is a 70% chance that an admitted student will decide to accept the offer and attend this university, what is the approximate probability that the university will not have enough dormitory room spots for the freshman class?

4.38 Speeding on the I-5, Part II. Exercise 4.36 states that the distribution of speeds of cars traveling on the Interstate 5 Freeway (I-5) in California is nearly normal with a mean of 72.6 miles/hour and a standard deviation of 4.78 miles/hour. The speed limit on this stretch of the I-5 is 70 miles/hour.

(a) A highway patrol officer is hidden on the side of the freeway. What is the probability that 5 cars pass and none are speeding? Assume that the speeds of the cars are independent of each other.
(b) On average, how many cars would the highway patrol officer expect to watch until the first car that is speeding? What is the standard deviation of the number of cars he would expect to watch?

4.39 Auto insurance premiums. Suppose a newspaper article states that the distribution of auto insurance premiums for residents of California is approximately normal with a mean of $1,650. The article also states that 25% of California residents pay more than $1,800.

(a) What is the Z-score that corresponds to the top 25% (or the 75^{th} percentile) of the standard normal distribution?
(b) What is the mean insurance cost? What is the cutoff for the 75th percentile?
(c) Identify the standard deviation of insurance premiums in California.

4.40 SAT scores. SAT scores (out of 1600) are distributed normally with a mean of 1100 and a standard deviation of 200. Suppose a school council awards a certificate of excellence to all students who score at least 1350 on the SAT, and suppose we pick one of the recognized students at random. What is the probability this student's score will be at least 1500? (The material covered in Section 3.2 on conditional probability would be useful for this question.)

4.41 Married women. The American Community Survey estimates that 47.1% of women ages 15 years and over are married.[41]

(a) We randomly select three women between these ages. What is the probability that the third woman selected is the only one who is married?
(b) What is the probability that all three randomly selected women are married?
(c) On average, how many women would you expect to sample before selecting a married woman? What is the standard deviation?
(d) If the proportion of married women was actually 30%, how many women would you expect to sample before selecting a married woman? What is the standard deviation?
(e) Based on your answers to parts (c) and (d), how does decreasing the probability of an event affect the mean and standard deviation of the wait time until success?

[40] S. Johnson and D. Murray. "Empirical Analysis of Truck and Automobile Speeds on Rural Interstates: Impact of Posted Speed Limits". In: *Transportation Research Board 89th Annual Meeting*. 2010.
[41] U.S. Census Bureau, 2010 American Community Survey, Marital Status.

4.42 Survey response rate. Pew Research reported that the typical response rate to their surveys is only 9%. If for a particular survey 15,000 households are contacted, what is the probability that at least 1,500 will agree to respond?[42]

4.43 Overweight baggage. Suppose weights of the checked baggage of airline passengers follow a nearly normal distribution with mean 45 pounds and standard deviation 3.2 pounds. Most airlines charge a fee for baggage that weigh in excess of 50 pounds. Determine what percent of airline passengers incur this fee.

4.44 Heights of 10 year olds, Part I. Heights of 10 year olds, regardless of gender, closely follow a normal distribution with mean 55 inches and standard deviation 6 inches.

(a) What is the probability that a randomly chosen 10 year old is shorter than 48 inches?
(b) What is the probability that a randomly chosen 10 year old is between 60 and 65 inches?
(c) If the tallest 10% of the class is considered "very tall", what is the height cutoff for "very tall"?

4.45 Buying books on Ebay. Suppose you're considering buying your expensive chemistry textbook on Ebay. Looking at past auctions suggests that the prices of this textbook follow an approximately normal distribution with mean \$89 and standard deviation \$15.

(a) What is the probability that a randomly selected auction for this book closes at more than \$100?
(b) Ebay allows you to set your maximum bid price so that if someone outbids you on an auction you can automatically outbid them, up to the maximum bid price you set. If you are only bidding on one auction, what are the advantages and disadvantages of setting a bid price too high or too low? What if you are bidding on multiple auctions?
(c) If you watched 10 auctions, roughly what percentile might you use for a maximum bid cutoff to be somewhat sure that you will win one of these ten auctions? Is it possible to find a cutoff point that will ensure that you win an auction?
(d) If you are willing to track up to ten auctions closely, about what price might you use as your maximum bid price if you want to be somewhat sure that you will buy one of these ten books?

4.46 Heights of 10 year olds, Part II. Heights of 10 year olds, regardless of gender, closely follow a normal distribution with mean 55 inches and standard deviation 6 inches.

(a) The height requirement for *Batman the Ride* at Six Flags Magic Mountain is 54 inches. What percent of 10 year olds cannot go on this ride?
(b) Suppose there are four 10 year olds. What is the chance that at least two of them will be able to ride *Batman the Ride*?
(c) Suppose you work at the park to help them better understand their customers' demographics, and you are counting people as they enter the park. What is the chance that the first 10 year old you see who can ride *Batman the Ride* is the 3rd 10 year old who enters the park?
(d) What is the chance that the fifth 10 year old you see who can ride *Batman the Ride* is the 12th 10 year old who enters the park?

4.47 Heights of 10 year olds, Part III. Heights of 10 year olds, regardless of gender, closely follow a normal distribution with mean 55 inches and standard deviation 6 inches.

(a) What fraction of 10 year olds are taller than 76 inches?
(b) If there are 2,000 10 year olds entering Six Flags Magic Mountain in a single day, then compute the expected number of 10 year olds who are at least 76 inches tall. (You may assume the heights of the 10-year olds are independent.)
(c) Using the binomial distribution, compute the probability that 0 of the 2,000 10 year olds will be at least 76 inches tall.
(d) The number of 10 year olds who enter Six Flags Magic Mountain and are at least 76 inches tall in a given day follows a Poisson distribution with mean equal to the value found in part (b). Use the Poisson distribution to identify the probability no 10 year old will enter the park who is 76 inches or taller.

4.48 Multiple choice quiz. In a multiple choice quiz there are 5 questions and 4 choices for each question (a, b, c, d). Robin has not studied for the quiz at all, and decides to randomly guess the answers. What is the probability that

(a) the first question she gets right is the 3^{rd} question?
(b) she gets exactly 3 or exactly 4 questions right?
(c) she gets the majority of the questions right?

---

[42]Pew Research Center, Assessing the Representativeness of Public Opinion Surveys, May 15, 2012.