Section 3.2: Conditional Probability

Diez, D. M., Çetinkaya-Rundel, M., Barr, C. D. (2019). OpenIntro Statistics (4th ed.). OpenIntro. https://www.openintro.org/book/os/ CC BY-SA 3.0

STAT 1201
Introduction to Probability and Statistics

ONLINE AND DISTANCE EDUCATION

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

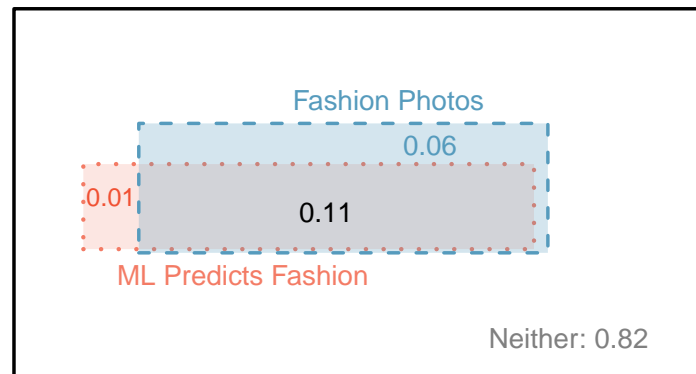Figure 3.11: Contingency table summarizing the `photo_classify` data set.

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?

Ⓔ

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:

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

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:

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

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

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

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

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

means the same thing as

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

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`:

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

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:

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

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:

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

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:

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

$$= \frac{\# \text{ cases where } \texttt{truth} \text{ is } \texttt{fashion} \text{ and } \texttt{mach\_learn} \text{ is } \texttt{pred\_fashion}}{\# \text{ cases where } \texttt{mach\_learn} \text{ is } \texttt{pred\_fashion}}$$

$$= \frac{197}{219} = 0.900$$

We considered only those cases that met the condition, $\texttt{mach\_learn}$ is $\texttt{pred\_fashion}$, 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 ($\texttt{mach\_learn}$ is $\texttt{pred\_fashion}$). 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

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

$$= \frac{\# \, (\texttt{truth} \text{ is } \texttt{fashion} \text{ and } \texttt{mach\_learn} \text{ is } \texttt{pred\_fashion})}{\# \, (\texttt{mach\_learn} \text{ is } \texttt{pred\_fashion})}$$

$$= \frac{108}{120} = \frac{0.108}{0.120} = 0.90$$

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

(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 $\texttt{truth}$ 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 $\texttt{pred\_fashion}$:

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

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

[29]$P(\texttt{result = died} \mid \texttt{inoculated = yes}) = \frac{P(\texttt{result = died and inoculated = yes})}{P(\texttt{inoculated = yes})} = \frac{0.0010}{0.0392} = 0.0255$ (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.

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.

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).

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]

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

Ⓖ

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]

Ⓖ

(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

Ⓖ

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>

Inoculated   Result

                                    lived, 0.9754
                                    ------------- 0.0392*0.9754 = 0.03824
            yes, 0.0392

                                    died, 0.0246
                                    ------------- 0.0392*0.0246 = 0.00096

                                    lived, 0.8589
                                    ------------- 0.9608*0.8589 = 0.82523
            no, 0.9608

                                    died, 0.1411
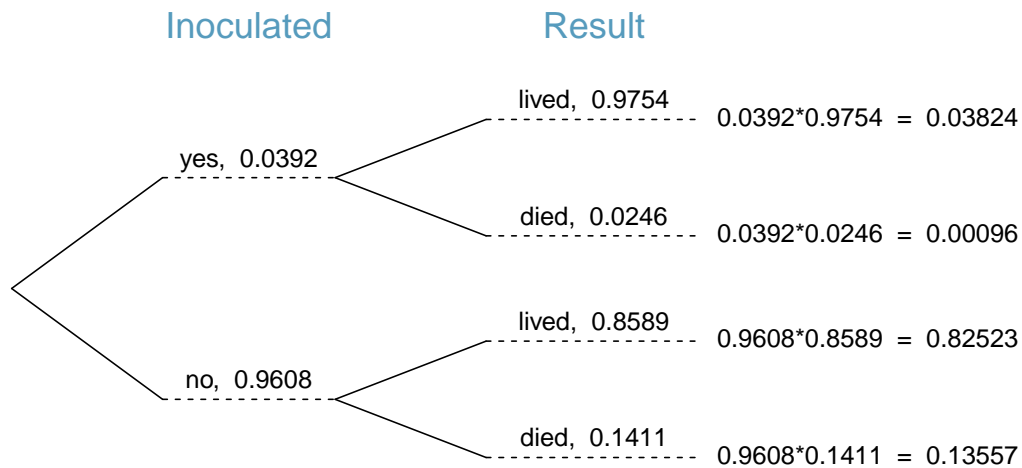   ------------- 0.9608*0.1411 = 0.13557

</figure>

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:

$$P(\texttt{inoculated = yes and result = lived})$$
$$= P(\texttt{inoculated = yes}) \times P(\texttt{result = lived|inoculated = yes})$$
$$= 0.0392 \times 0.9754 = 0.0382$$

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:

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

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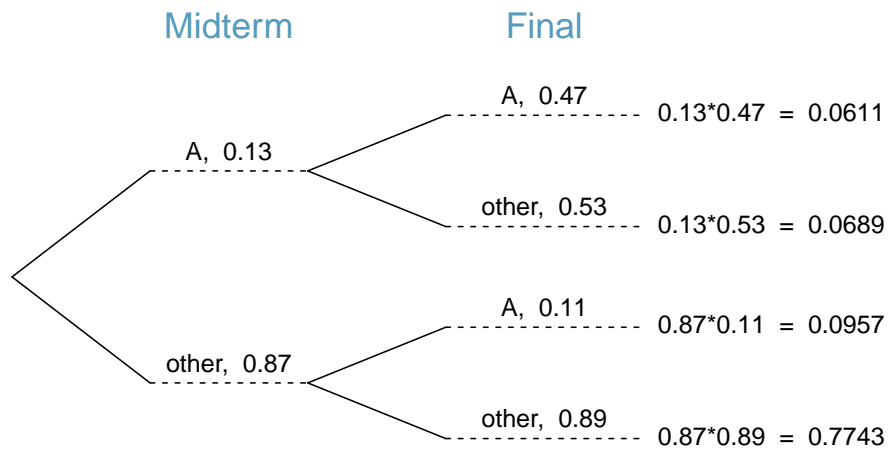
When constructing a tree diagram, variables provided with marginal probabilities are often used to create the tree's primary branches; in this case, the marginal probabilities are provided for midterm grades. The final grades, which correspond to the conditional probabilities provided, will be shown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

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

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

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:

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

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) $P(\text{construct tree diagram} \mid \text{passed}) = \frac{0.7566}{0.8820} = 0.8578$.
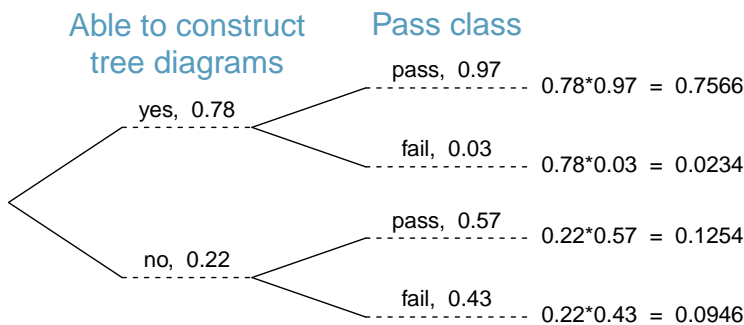
EXAMPLE 3.42

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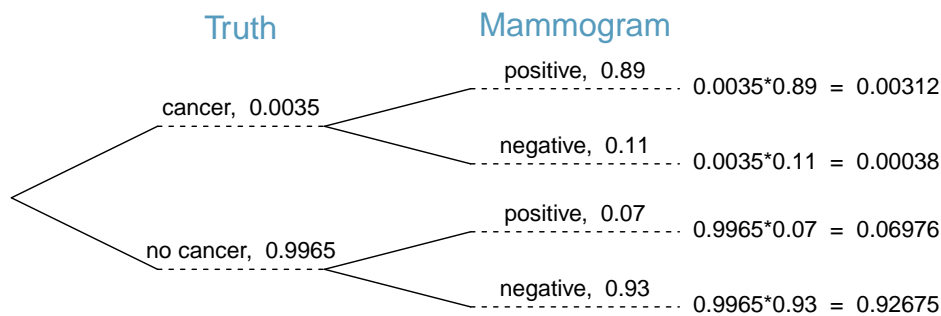
The probability the patient has breast cancer and the mammogram is positive is

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

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

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

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

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

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.

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

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

$$P(\text{has BC} \mid \text{mammogram}^+)$$
$$= \frac{P(\text{mammogram}^+ \mid \text{has BC})P(\text{has BC})}{P(\text{mammogram}^+ \mid \text{no BC})P(\text{no BC}) + P(\text{mammogram}^+ \mid \text{has BC})P(\text{has BC})}$$

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.
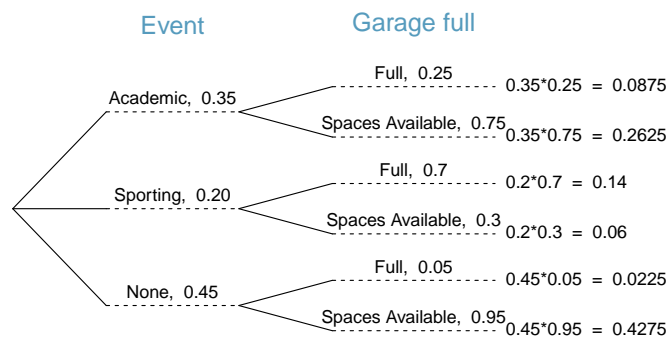
GUIDED PRACTICE 3.45

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: $1.00 - 0.56 - 0.35 = 0.09$.