⊙ **Guided Practice A.32** Three people are selected at random.[22]

(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 Equation (A.29).

Example A.33 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 Equation A.29 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.

A.2 Conditional probability

Are students more likely to use marijuana when their parents used drugs? The `drug_use` data set contains a sample of 445 cases with two variables, `student` and `parents`, and is summarized in Table A.11.[23] The `student` variable is either `uses` or `not`, where a student is labeled as `uses` if she has recently used marijuana. The `parents` variable takes the value `used` if at least one of the parents used drugs, including alcohol.

		parents		
		used	not	Total
student	uses	125	94	219
	not	85	141	226
	Total	210	235	445

Table A.11: Contingency table summarizing the `drug_use` data set.

Example A.34 If at least one parent used drugs, what is the chance their child (`student`) uses?

We will estimate this probability using the data. Of the 210 cases in this data set where `parents` = `used`, 125 represent cases where `student` = `uses`:

$$P(\texttt{student} = \texttt{uses} \text{ given } \texttt{parents} = \texttt{used}) = \frac{125}{210} = 0.60$$

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

[23] Ellis GJ and Stone LH. 1979. Marijuana Use in College: An Evaluation of a Modeling Explanation. Youth and Society 10:323-334.

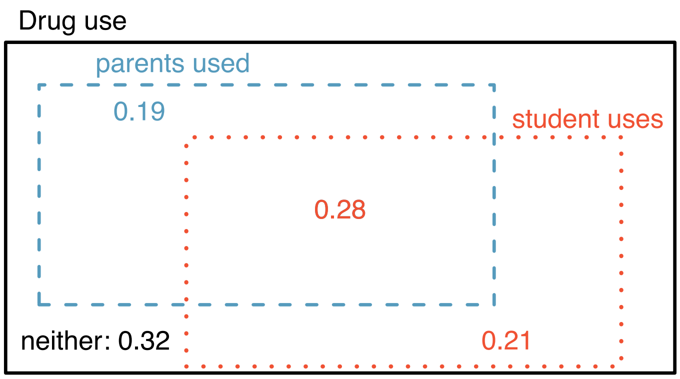

Figure A.12: A Venn diagram using boxes for the `drug_use` data set.

	`parents: used`	`parents: not`	Total
`student: uses`	0.28	0.21	0.49
`student: not`	0.19	0.32	0.51
Total	0.47	0.53	1.00

Table A.13: Probability table summarizing parental and student drug use.

Example A.35 A student is randomly selected from the study and she does not use drugs. What is the probability that at least one of her parents used?

If the student does not use drugs, then she is one of the 226 students in the second row. Of these 226 students, 85 had at least one parent who used drugs:

$$P(\texttt{parents} = \texttt{used} \text{ given } \texttt{student} = \texttt{not}) = \frac{85}{226} = 0.376$$

A.2.1 Marginal and joint probabilities

Table A.13 includes row and column totals for each variable separately in the `drug_use` data set. These totals represent **marginal probabilities** for the sample, which are the probabilities based on a single variable without conditioning on any other variables. For instance, a probability based solely on the `student` variable is a marginal probability:

$$P(\texttt{student} = \texttt{uses}) = \frac{219}{445} = 0.492$$

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

$$P(\texttt{student} = \texttt{uses} \text{ and } \texttt{parents} = \texttt{not}) = \frac{94}{445} = 0.21$$

It is common to substitute a comma for "and" in a joint probability, although either is acceptable.

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 `drug_use` sample. These proportions are computed by dividing each count in Table A.11 by 445 to obtain the proportions in Table A.13. The joint probability distribution of the `parents` and `student` variables is shown in Table A.14.

Joint outcome	Probability
`parents` = `used`, `student` = `uses`	0.28
`parents` = `used`, `student` = `not`	0.19
`parents` = `not`, `student` = `uses`	0.21
`parents` = `not`, `student` = `not`	0.32
Total	1.00

Table A.14: A joint probability distribution for the `drug_use` data set.

⊙ **Guided Practice A.36** Verify Table A.14 represents a probability distribution: events are disjoint, all probabilities are non-negative, and the probabilities sum to 1.[24]

We can compute marginal probabilities using joint probabilities in simple cases. For example, the probability a random student from the study uses drugs is found by summing the outcomes from Table A.14 where `student` = `uses`:

$$\begin{aligned} &P(\underline{\texttt{student} = \texttt{uses}}) \\ &\quad = P(\texttt{parents} = \texttt{used}, \underline{\texttt{student} = \texttt{uses}}) + \\ &\qquad\qquad P(\texttt{parents} = \texttt{not}, \underline{\texttt{student} = \texttt{uses}}) \\ &\quad = 0.28 + 0.21 = 0.49 \end{aligned}$$

A.2.2 Defining conditional probability

There is some connection between drug use of parents and of the student: drug use of one is associated with drug use of the other.[25] In this section, we discuss how to use information about associations between two variables to improve probability estimation.

The probability that a random student from the study uses drugs is 0.49. Could we update this probability if we knew that this student's parents used drugs? Absolutely. To do so, we limit our view to only those 210 cases where parents used drugs and look at the fraction where the student uses drugs:

$$P(\texttt{student} = \texttt{uses} \text{ given } \texttt{parents} = \texttt{used}) = \frac{125}{210} = 0.60$$

We call this a **conditional probability** because we computed the probability under a condition: `parents` = `used`. 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 separate the text inside our probability notation into the outcome of interest and the condition:

$$\begin{aligned} &P(\texttt{student} = \texttt{uses} \text{ given } \texttt{parents} = \texttt{used}) \\ &= P(\texttt{student} = \texttt{uses} \mid \texttt{parents} = \texttt{used}) = \frac{125}{210} = 0.60 \end{aligned} \tag{A.37}$$

$P(A|B)$
Probability of outcome A given B

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

In Equation (A.37), we computed the probability a student uses based on the condition that

[24] Each of the four outcome combination are disjoint, all probabilities are indeed non-negative, and the sum of the probabilities is $0.28 + 0.19 + 0.21 + 0.32 = 1.00$.

[25] This is an observational study and no causal conclusions may be reached.

at least one parent used as a fraction:

$$\begin{aligned} &P(\texttt{student} = \texttt{uses} \mid \texttt{parents} = \texttt{used}) \\ &\quad = \frac{\#\text{ times } \texttt{student} = \texttt{uses} \text{ and } \texttt{parents} = \texttt{used}}{\#\text{ times } \texttt{parents} = \texttt{used}} \\ &\quad = \frac{125}{210} = 0.60 \end{aligned} \tag{A.38}$$

We considered only those cases that met the condition, $\texttt{parents} = \texttt{used}$, and then we computed the ratio of those cases that satisfied our outcome of interest, the student uses.

Counts are not always available for data, and instead only marginal and joint probabilities may be provided. 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 Equation (A.38) as an example demonstrating this technique.

We considered only those cases that satisfied the condition, $\texttt{parents} = \texttt{used}$. Of these cases, the conditional probability was the fraction who represented the outcome of interest, $\texttt{student} = \texttt{uses}$. Suppose we were provided only the information in Table A.13 on page 307, i.e. only probability data. Then if we took a sample of 1000 people, we would anticipate about 47% or $0.47 \times 1000 = 470$ would meet our information criterion. Similarly, we would expect about 28% or $0.28 \times 1000 = 280$ to meet both the information criterion and represent our outcome of interest. Thus, the conditional probability could be computed:

$$\begin{aligned} P(\texttt{student} = \texttt{uses} \mid \texttt{parents} = \texttt{used}) &= \frac{\#\ (\texttt{student} = \texttt{uses} \text{ and } \texttt{parents} = \texttt{used})}{\#\ (\texttt{parents} = \texttt{used})} \\ &= \frac{280}{470} = \frac{0.28}{0.47} = 0.60 \end{aligned} \tag{A.39}$$

In Equation (A.39), we examine exactly the fraction of two probabilities, 0.28 and 0.47, which we can write as

$$P(\texttt{student} = \texttt{uses} \text{ and } \texttt{parents} = \texttt{used}) \quad \text{and} \quad P(\texttt{parents} = \texttt{used}).$$

The fraction of these probabilities represents our general formula for conditional probability.

Conditional Probability

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

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)} \tag{A.40}$$

Guided Practice A.41 (a) Write out the following statement in conditional probability notation: "*The probability a random case has* $\texttt{parents} = \texttt{not}$ *if it is known that* $\texttt{student} = \texttt{not}$". Notice that the condition is now based on the student, not the parent. (b) Determine the probability from part (a). Table A.13 on page 307 may be helpful.[26]

[26](a) $P(\texttt{parent} = \texttt{not}|\texttt{student} = \texttt{not})$. (b) Equation (A.40) for conditional probability indicates we should first find $P(\texttt{parents} = \texttt{not} \text{ and } \texttt{student} = \texttt{not}) = 0.32$ and $P(\texttt{student} = \texttt{not}) = 0.51$. Then the ratio represents the conditional probability: $0.32/0.51 = 0.63$.

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

Table A.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

Table A.16: Table proportions for the `smallpox` data, computed by dividing each count by the table total, 6224.

⊙ **Guided Practice A.42** (a) Determine the probability that one of the parents had used drugs if it is known the student does not use drugs. (b) Using the answers from part (a) and Guided Practice A.41(b), compute

$$P(\texttt{parents} = \texttt{used}|\texttt{student} = \texttt{not}) + P(\texttt{parents} = \texttt{not}|\texttt{student} = \texttt{not})$$

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

⊙ **Guided Practice A.43** The data indicate that drug use of parents and children are associated. Does this mean the drug use of parents causes the drug use of the students?[28]

A.2.3 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.[29] 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 A.15 and A.16.

⊙ **Guided Practice A.44** Write out, in formal notation, the probability a randomly selected person who was not inoculated died from smallpox, and find this probability.[30]

[27](a) This probability is $\frac{P(\texttt{parents} = \texttt{used and student} = \texttt{not})}{P(\texttt{student} = \texttt{not})} = \frac{0.19}{0.51} = 0.37$. (b) The total equals 1. (c) Under the condition the student does not use drugs, the parents must either use drugs or not. The complement still appears to work *when conditioning on the same information*.

[28]No. This was an observational study. Two potential confounding variables include `income` and `region`. Can you think of others?

[29]Fenner F. 1988. *Smallpox and Its Eradication (History of International Public Health, No. 6)*. Geneva: World Health Organization. ISBN 92-4-156110-6.

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

⊙ **Guided Practice A.45** Determine the probability that an inoculated person died from smallpox. How does this result compare with the result of Guided Practice A.44?[31]

⊙ **Guided Practice A.46** 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?[32]

A.2.4 General multiplication rule

Section A.1.6 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 definition for conditional probability in Equation (A.40) on page 309.

● **Example A.47** 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 Table A.16. We want to determine

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

and we are given that

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

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 Table A.16 at the intersection of `no` and `lived` (with a small rounding error).

⊙ **Guided Practice A.48** 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.[33]

---

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

[32] 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).

[33] The answer is 0.0382, which can be verified using Table A.16.

Guided Practice A.49 If 97.45% of the people who were inoculated lived, what proportion of inoculated people must have died?[34]

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 A.50 Based on the probabilities computed above, does it appear that inoculation is effective at reducing the risk of death from smallpox?[35]

A.2.5 Independence considerations in conditional probability

If two processes 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 A.51 Let X and Y represent the outcomes of rolling two dice. (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 $P(Y = \texttt{1} \,|X = \texttt{1})$. (d) What is $P(Y = 1)$? Is this different from the answer from part (c)? Explain.[36]

We can show in Guided Practice A.51(c) that the conditioning information has no influence by using the Multiplication Rule for independence processes:

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

Guided Practice A.52 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?[37]

[34] There were only two possible outcomes: `lived` or `died`. This means that 100% - 97.45% = 2.55% of the people who were inoculated died.

[35] 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 A.46, 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.)

[36] Brief solutions: (a) 1/6. (b) 1/36. (c) $\frac{P(Y=\texttt{1} \text{ and } X=\texttt{1})}{P(X=\texttt{1})} = \frac{1/36}{1/6} = 1/6$. (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.

[37] He has forgotten that the next roulette spin is independent of the previous spins. Casinos do employ this practice; they post 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**.

A.2.6 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 A.17. The first branch for `inoculation` is said to be the **primary** branch while the other branches are **secondary**.

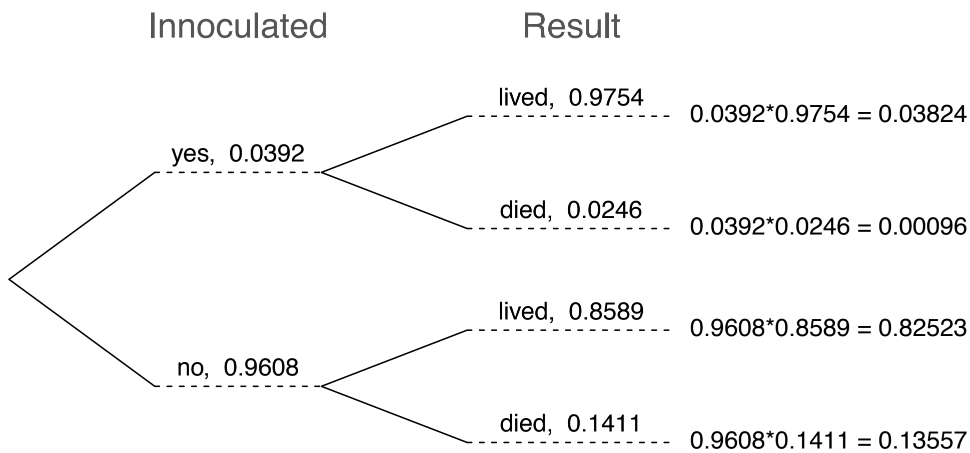

Figure A.17: A tree diagram of the `smallpox` data set.

Tree diagrams are annotated with marginal and conditional probabilities, as shown in Figure A.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 A.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 A.53 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, as shown in Figure A.18. 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.

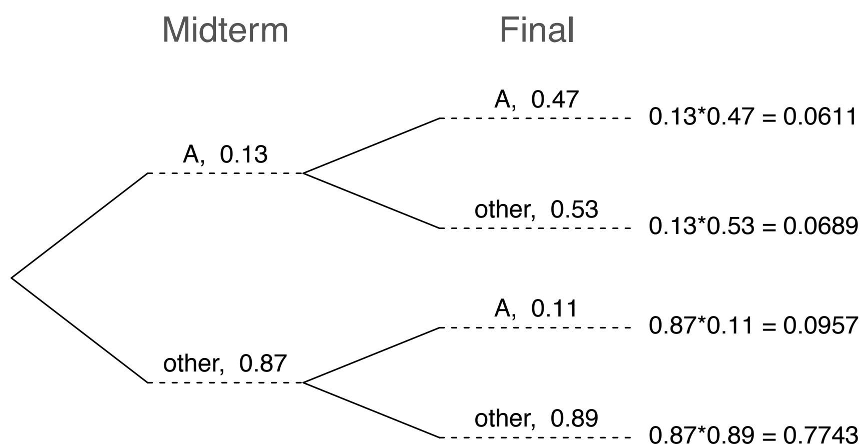

Figure A.18: A tree diagram describing the `midterm` and `final` variables.

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

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

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.