Example 2.47 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 2.16. 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 2.16 at the intersection of `no` and `lived` (with a small rounding error).

Exercise 2.48 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]

Exercise 2.49 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:

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

Exercise 2.50 Based on the probabilities computed above, does it appear that inoculation is effective at reducing the risk of death from smallpox?[35]

---

[33] The answer is 0.0382, which can be verified using Table 2.16.

[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 Exercise 2.46, 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.)

2.2.5 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.

Exercise 2.51 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 $P(Y = 1 | X = 1)$. (d) What is $P(Y = 1)$? Is this different from the answer from part (c)? Explain.[36]

We can show in Exercise 2.51(c) that the conditioning information has no influence by using the Multiplication Rule for independence processes:

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

Exercise 2.52 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]

2.2.6 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 2.17. The first branch for `inoculation` is said to be the **primary** branch while the other branches are **secondary**.

Tree diagrams are annotated with marginal and conditional probabilities, as shown in Figure 2.17. 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 2.17 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

[36] 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.

[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**.

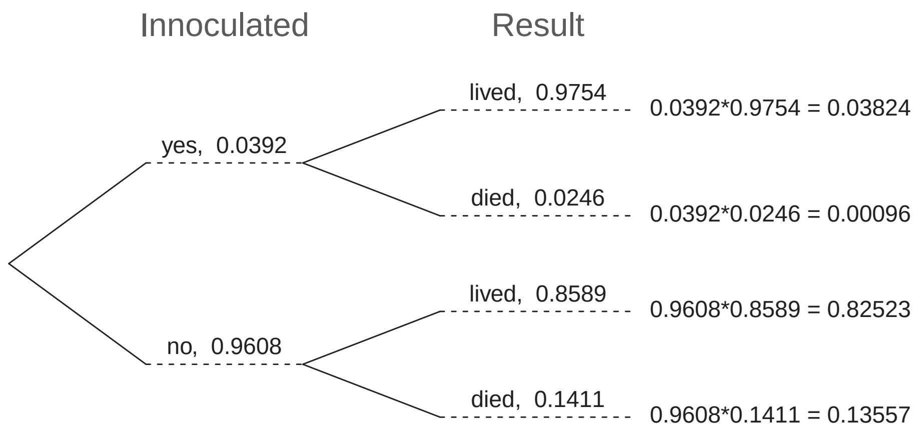


Figure 2.17: A tree diagram of the `smallpox` data set.

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 2.53 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 2.18. 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.

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

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

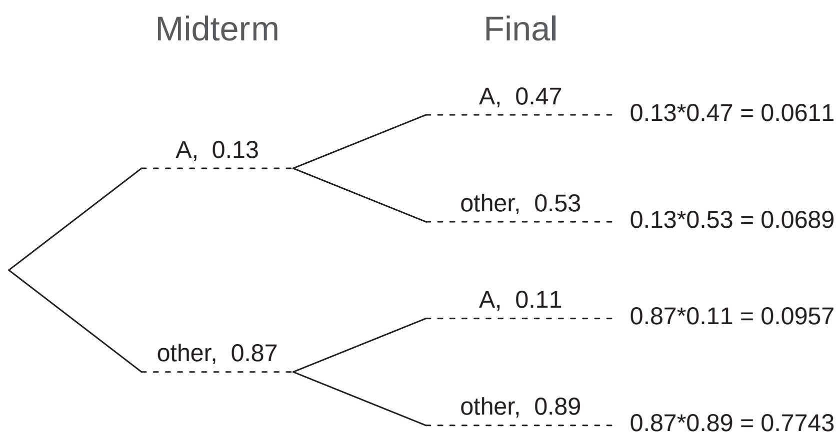


Figure 2.18: A tree diagram describing the `midterm` and `final` variables.

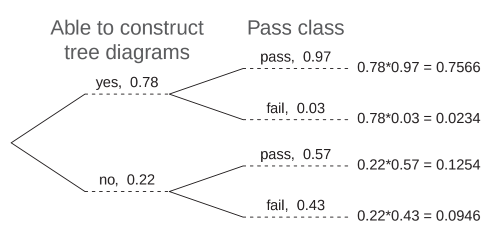
The marginal probability, $P(\texttt{final} = \texttt{A})$, was calculated by adding up all the joint probabilities on the right side of the tree that correspond to $\texttt{final} = \texttt{A}$. We may now finally take the ratio of the two probabilities:

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

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

Exercise 2.54 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.[38]

[38](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) $P(\text{construct tree diagram} \mid \text{passed}) = \frac{0.7566}{0.8820} = 0.8578$.



2.2.7 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.

Example 2.55 In Canada, about 0.35% of women over 40 will be diagnosed with 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.[39] 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?

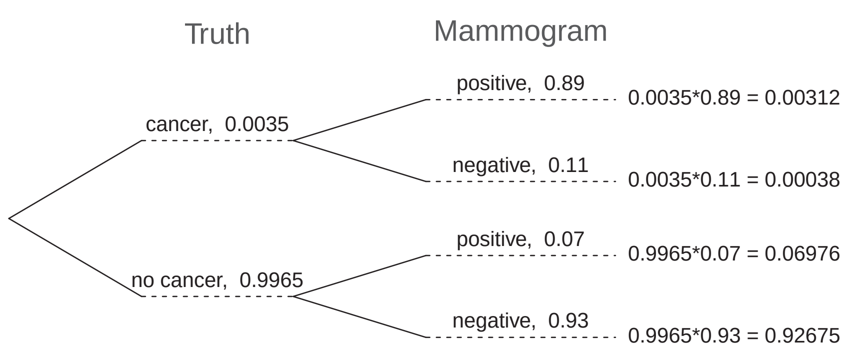


Figure 2.19: Tree diagram for Example 2.55, computing the probability a random patient who tests positive on a mammogram actually has breast cancer.

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

[39]The probabilities reported here were obtained using studies reported at www.breastcancer.org and www.ncbi.nlm.nih.gov/pmc/articles/PMC1173421.

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 actually having breast cancer and "mammogram$^+$" means the mammogram screening was positive. A tree diagram is useful for identifying each probability and is shown in Figure 2.19. 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:

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

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 2.55 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 2.55. 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.

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

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:

$$\frac{P(B|A_1)P(A_1)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots + P(B|A_k)P(A_k)} \tag{2.56}$$

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

Bayes' Theorem is just 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.

TIP: Only use Bayes' Theorem when tree diagrams are difficult

Drawing a tree diagram makes it easier to understand how two variables are connected. Use Bayes' Theorem only when there are so many scenarios that drawing a tree diagram would be complex.

Exercise 2.57 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.[40]

Example 2.58 Here we solve the same problem presented in Exercise 2.57, 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

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

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):

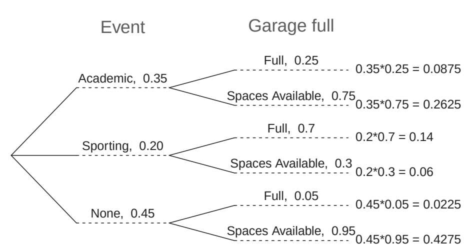
$$\begin{aligned} P(A_1|B) &= \frac{P(B|A_1)P(A_1)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3)} \\ &= \frac{(0.7)(0.2)}{(0.7)(0.2) + (0.25)(0.35) + (0.05)(0.45)} \\ &= 0.56 \end{aligned}$$

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.

Exercise 2.59 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.[41]

---

[40]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.



[41]Short answer:

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

Exercise 2.60 In Exercise 2.57 and 2.59, you found that if the parking lot is full, the probability a sporting event is 0.56 and the probability there is an academic event is 0.35. Using this information, compute $P(\text{no event} \mid \text{the lot is full})$.[42]

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.

2.3 Sampling from a small population (special topic)

Example 2.61 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.

Example 2.62 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

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

Exercise 2.63 What rule permitted us to multiply the probabilities in Example 2.62?[43]

---

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

[43]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 2.64 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.

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

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.

Exercise 2.65 Under the setup of Example 2.64, what is the probability of being picked to answer all three questions?[44]

If we sample from a small population **without replacement**, we no longer have independence between our observations. In Example 2.62, 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 2.64, the professor sampled her students **with replacement**: she repeatedly sampled the entire class without regard to who she already picked.

Exercise 2.66 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?[45]

Exercise 2.67 Compare your answers in Exercise 2.66. How much influence does the sampling method have on your chances of winning a prize?[46]

Had we repeated Exercise 2.66 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.

---

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

[45] (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.

[46] 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.