Conditional Probability Example Problems

1. Conditional Proportions

In a lecture class of 150 students, 110 students are freshmen, 50 own a dog, and 25 are freshmen who own a dog. Suppose a student is selected at random.

(a) What proportion of dog owners in the class are freshmen?

(b) What proportion of freshmen in the class are dog owners?

(c) Do you get the same answer for both parts? Why or why not?

2. Conditional Probabilities and Diagnostic Testing

You think you might have a disease that affects 0.1% of the population. You take a test that has 99% sensitivity and 95% specificity. The test comes back positive. What is the probability that you have the disease?

3. Conditional Probabilities and Bayes's Rule

In a large national population of college students, 61% attend four-year institutions and the rest attend two-year institutions. Males make up 44% of the students in the four-year institutions and 41% of the students in the two-year institutions (assume the rest are female). Consider randomly selecting a female student from the population. What is the probability that she attends a four-year institution?

Conditional Probability Example Solutions

1. Conditional Proportions

In a lecture class of 150 students, 110 students are freshmen, 50 own a dog, and 25 are freshmen who own a dog. Suppose a student is selected at random.

(a) What proportion of dog owners in the class are freshmen?

The information that we are given is that the student owns a dog. There are 50 students who own dogs. Of those 50 students, 25 are freshman. Therefore the proportion of dog owners in the class that are freshmen is:

$$\frac{number\ of\ students\ who\ are\ freshmen\ \textbf{and}\ own\ dogs}{total\ number\ of\ students\ who\ own\ dogs} = \frac{25}{50} = \frac{1}{2} = 0.5$$

(b) What proportion of freshmen in the class are dog owners?

The information that we are given is that the student is a freshman. There are 110 freshman students in the class. Of those 110 freshmen, 25 have dogs. Therefore the proportion of freshmen in the class who are dog owners is:

$$\frac{number\ of\ students\ who\ are\ freshmen\ \textbf{and}\ own\ dogs}{total\ number\ of\ students\ who\ are\ freshmen} = \frac{25}{110} = \frac{5}{22} = 0.227$$

(c) Do you get the same answer for both parts? Why or why not?

No. For both parts, the numerator of the fraction (number of students who are freshmen and own dogs) stays the same, but since we are given different information in the two parts, the denominator of the fraction changes. When we are given the information that the student owns a dog, the denominator is the number of students who own dogs; when we are given the information that the students is a freshman, the denominator is the number of students who are freshmen.

2. Conditional Probabilities and Diagnostic Testing

You think you might have a disease that affects 0.1% of the population. You take a test that has 99% sensitivity and 95% specificity. The test comes back positive. What is the probability that you have the disease?

For those of us still getting comfortable with statistics word problems, it helps to explicitly write out the goal of the problem and organize the information we are given.

The goal of this problem is to find a conditional probability; specifically, that probability is the probability of having the disease, given that your test comes back positive.

The information we know consists of one marginal proportion (unconditional probability) and two conditional proportions/probability. In order presented in the problem, they are:

1) 0.1% of the population has the disease (marginal proportion/unconditional probability)

2) 99% of people who have the disease will test positive (conditional proportion - definition of sensitivity)
3) 95% of people who don't have the disease will test negative (conditional proportion - definition of specificity)

Using this information, we can use a two-way table, tree diagram, or Bayes's Rule directly to compute the probability of having the disease given the positive test result.

(a) Using a two-way table:

Let us claim that there are 10,000 people in our population (this works better with even larger numbers if you don't want to use decimals). Since 0.1% of the population has the disease, we expect that 10 out of those 10,000 people actually have the disease and the rest (9990) do not. Thus, we start by filling out our table:

	Test +	Test -	Total
Disease			10
No Disease			9990
Total			10,000

Next, by the definition of sensitivity we know that there is a 99% of testing positive if we have the disease. Similarly, by the definition of specificity, we have a 95% chance of testing negative if we do not have the disease. Thus, we fill in those cells in the table (note again that we do not round to whole numbers if possible):

	Test +	Test -	Total
Disease	(10)(99%) = 9.9		10
No Disease		(9990)(95%) = 9490.5	9990
Total			10,000

Finally, we do our usual addition and subtraction to fill in the rest of the numbers in the two-way table:

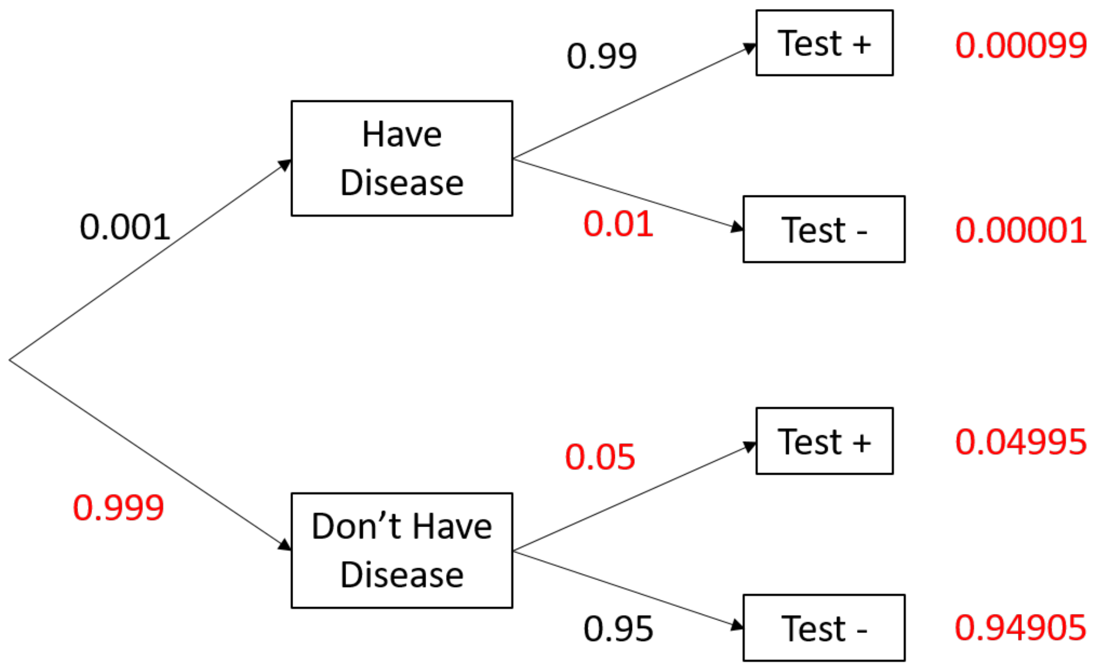
	Test +	Test -	Total
Disease	9.9	0.1	10
No Disease	499.5	9490.5	9990
Total	509.4	9490.6	10,000

Finally, we write out our conditional probability statement and fill in the appropriate numbers:

$$P(disease \mid test\ positive) = \frac{number\ who\ have\ disease\ and\ test\ positive}{number\ who\ test\ positive} = \frac{9.9}{50.94} = 0.019$$

(b) Using a tree diagram:

First, we have to figure out which variable (disease or test result) we have *unconditional* probabilities for and which variable we have *conditional* probabilities for. In this case, we have unconditional probabilities (marginal proportions) for having vs. not having the disease. Sensitivity and specificity are conditional probabilities for test results given disease status. Therefore, the first branch in our tree needs to be for disease, and the second for test result given disease. We fill in the known proportions and then use them to fill in the proportions for the rest of the branches, using the fact that the probability of "exiting" a node is 1 (we must go somewhere until we hit the end of the path). In the tree diagram below, black numbers are given probabilities and red numbers are calculated probabilities. The red numbers at the right depict the probability of ending up at each node:



Then, we write the formula for conditional probability:

$$P(disease \mid test\ positive) = \frac{P(disease \cap test\ positive)}{P(test\ positive)}$$

We can read the value of the top probability directly off the tree diagram. For the bottom probability, we have to add the probabilities of all paths through the diagram that end at "Test +":

$$\frac{P(disease \cap test\ positive)}{P(test\ positive)} = \frac{0.00099}{0.00099 + 0.04995} = \frac{0.00099}{0.05094} = 0.019$$

Note that in this example, I converted all the proportions to decimals. In the example in lab, I kept everything in percent. It doesn't really matter.

(c) Using Bayes's Theorem/Rule:
Let A be the event "we have the disease" and B be the event "we test positive for the disease." Thus, we want to find

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

However, we do not know $P(B)$ unconditionally (nor are we given $P(A \cap B)$. Therefore, we must use Bayes's Rule to expand the numerator and denominator of the fraction:

$$P(A \mid B) = \frac{P(A)P(B \mid A)}{P(A)P(B \mid A) + P(A^C)P(B \mid A^C)}$$

Note that A^C means "we do not have the disease." From the information given in the problem, our test has 99% sensitivity, so $P(B \mid A) = 0.99$. Similarly, since our test has 95% specificity, $P(B^C \mid A^C) = 0.95$, so $P(B \mid A^C) = 1 - P(B^C \mid A^C) = 0.05$. Finally, because 0.1% of people have the disease, $P(A) = 0.001$ and $P(A^C) = 1 - P(A) = 0.999$. Therefore:

$$P(A \mid B) = \frac{P(A)P(B \mid A)}{P(A)P(B \mid A) + P(A^C)P(B \mid A^C)} = \frac{(0.1\%)(99\%)}{(0.1\%)(99\%) + (99.9\%)(5\%)}$$

$$= \frac{9.9}{9.9 + 499.5} = \frac{9.9}{509.4} = 0.019$$

or to be technically correct,

$$= \frac{0.00099}{0.00099 + 0.04995} = \frac{0.00099}{0.05094} = 0.019$$

Using any method, we find that the probability of having the disease, given a positive result, is 0.019. This test has a positive predictive value of 1.9%. (Or, there is still only a 1.9% chance of having the disease, even given the positive test result.)

Incidentally, using any of those three methods, we can also find that the negative predictive value of the test (the probability of not having the disease, given a negative test result) is about 0.99999 (99.999%). Essentially, with rare diseases, if the test comes back negative, you almost assuredly don't have the disease, but if the test comes back positive, you still are much more likely to be a false positive (this is why testing for rare diseases usually doesn't stop at the initial screen).

3. Conditional Probabilities and Bayes's Rule

In a large national population of college students, 61% attend four-year institutions and the rest attend two-year institutions. Males make up 44% of the students in the four-year institutions and 41% of the students in the two-year institutions (assume the rest are female). Consider randomly selecting a female student from the population. What is the probability that she attends a four-year institution?

For those of us still getting comfortable with statistics word problems, it helps to explicitly write out the goal of the problem and organize the information we are given.

The goal of this problem is to find a conditional probability; specifically, that probability is the probability of picking a student at a four-year institution, conditioned upon the fact that the student we randomly selected is female.

The information we know consists of one marginal proportion (unconditional probability) and two conditional proportions/probability. In order presented in the problem, they are:

1) 61% of the population attend four-year institutions (marginal proportion/unconditional probability)
2) 44% of four-year students are male (conditional proportion/probability)
3) 41% of two-year students are male (conditional proportion/probability)

Using this information, we can use a two-way table, tree diagram, or Bayes's Rule directly to compute the (conditional) proportion of female students who attend four-year institutions. Recall that when we randomly select 1 individual from a population, we can consider the probability of some event to be the relative frequency of that event in the population. Thefore, if we find this conditional proportion, it will also be the conditional probability we are interested in.

(a) Using a two-way table:

Let us claim that there are 10,000 people in our sample. We know that 61% of those 10,000 people attend four-year institutions and the rest (39%) attend two-year institutions. Thus, we start by filling out our table:

	Male	Female	Total
4-Year			6100
2-Year			3900
Total			10,000

Next, we know that 44% of the students at four-year institutions are male, and 41% of the students at two-year institutions are male, so we use those numbers to find the number of male students at each type of institution.

	Male	Female	Total
4-Year	$(6100)(44\%) = 2684$		6100
2-Year	$(39)(41\%) = 1599$		3900
Total			10,000

Finally, we do our usual addition and subtraction to fill in the rest of the numbers in the two-way table:

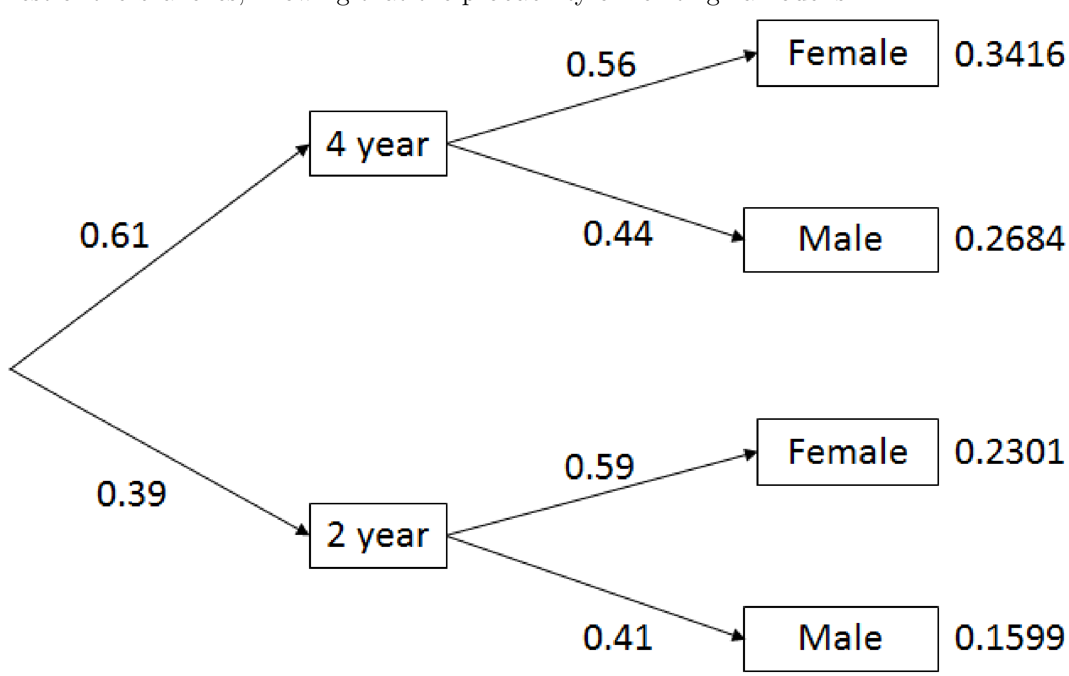
	Male	Female	Total
4-Year	2684	3416	6100
2-Year	1599	2301	3900
Total	4283	5717	10,000

Finally, we write out our conditional probability statement and fill in the appropriate numbers:

$$P(four\ year\ institution \mid female) = \frac{number\ of\ female\ four\ year\ students}{total\ number\ of\ female\ students} = \frac{3416}{5717} = 0.5975$$

(b) Using a tree diagram:

First, we have to figure out which variable (college type or gender) we have *unconditional* probabilities for and which variable we have *conditional* probabilities for. In this case, we have unconditional probabilities (marginal proportions) for college type, and conditional probabilities for gender given college type. Therefore, the first branch in our tree needs to be for college type, and the second for gender. We fill in the given proportions and then use them to fill in the proportions for the rest of the branches, knowing that the probability of "exiting" a node is 1:



Then, we write the formula for conditional probability:

$$P(four\ year\ institution \mid female) = \frac{P(four\ year\ institution \cap female)}{P(female)}$$

We can read the value of the top probability directly off the tree diagram. For the bottom probability, we have to add the probabilities of all paths through the diagram that end at "Female":

$$\frac{P(four\ year\ institution \cap female)}{P(female)} = \frac{0.3416}{0.3416 + 0.2301} = \frac{0.3416}{0.5717} = 0.5975$$

(Again, you can use percentages instead of decimals and the only thing that will change is that the numbers in the numerator and denominator will be bigger - the decimal the fraction evaluates to will be the same)

(c) Using Bayes's Theorem/Rule:

Let A be the event "the person we select attends a four year institution" and B be the event "the person we select is female." Thus, we want to find

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

However, we do not know $P(B)$ unconditionally (nor are we given $P(A \cap B)$. Therefore, we must use Bayes's Rule to expand the numerator and denominator of the fraction:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B \mid A)P(A) + P(B \mid A^C)P(A^C)}$$

Note that A^C means "the person we select attends a two-year institution." From the information given in the problem, 44% of people at four-year institutions are male, so 56% are female. Therefore $P(B \mid A) = 0.56$. Similarly, 41% of people at two-year institutions are male, so 59% are female and $P(B \mid A^C) = 0.59$. We note that 61% of students attend four-year institutions, so $P(A) = 0.61$ and $P(A^C) = 1 - P(A) = 0.39$. Therefore:

$$P(A \mid B) = \frac{P(A)P(B \mid A)}{P(A)P(B \mid A) + P(A^C)P(B \mid A^C)} = \frac{(61\%)(56\%)}{(61\%)(56\%) + (39\%)(59\%)}$$

$$= \frac{3416}{3416 + 2301} = \frac{3416}{5717} = 0.5975$$

or to be technically correct,

$$= \frac{0.3416}{0.3416 + 0.2301} = \frac{0.3416}{0.5717} = 0.5975$$

Using any method, we find that the probability of selecting a student at a four-year institution, given that the student is female, is 0.5975.