Independence and Conditional Probability

August 5, 2019

Midterm

The Midterm is next week Tuesday, August 13.

- Approximately 50 multiple choice questions.
- You do not need a scantron.
- Questions will be mostly conceptual.
- You may bring any basic or graphing calculator.
- I will bring extra scratch paper.

Extra Credit Opportunity

- Write an exam question that would be appropriate for your midterm.
- The midterm will cover material from Chapters 1, 2, and 3.
- Your exam question must come from material covered in class, your homeworks, or your labs.
- Questions may be either multiple choice or short answer.
- To receive any credit, you must write an original question and provide both the question and the correct answer.

These can be submitted on iLearn (Assignments tab). It opens today at 9:30am and will close on Thursday at 11:59pm.

Independence

- Independence of random processes is similar to independence of variables and observations.
- We say that two random processes are **independent** if knowing the outcome of one provides no useful information about the outcome of the other.

Independence

For example, consider our discussion on rolling 2 six-sided dice.

- The roll of the first die has no effect on the roll of the second die.
- Thus our two dice rolls are independent of one another.

Independence

We've already calculated the probability of the two rolls both being a `1`

- $1/6$ of the time the first roll is a `1`
- A further $1/6$ of *those* times the second is also a `1`.
- So we decided that the probability was $(1/6) \times (1/6) = 1/36$.

Multiplying these probabilities together works because the two events are independent.

Multiplication Rule for Independent Processes

Let A and B be events from two different and independent processes. Then the probability that both A and B occur can be calculated as the product of their separate probabilities:

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

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

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

Example

About 9% of people are left-handed. Suppose 2 people are selected at random from the U.S. population. Because the sample size of 2 is very small relative to the population, it is reasonable to assume these two people are independent.

1. What is the probability that both are left-handed?
2. What is the probability that both are right-handed?

Example: Both Left-Handed

What is the probability that both are left-handed?

- Let L_1 be the event that the first person is left-handed and L_2 the event that the second person is left-handed.
- We are told that 9% of people are left-handed, so $P(L_1) = P(L_2) = 0.09$.

Example: Both Left-Handed

What is the probability that both are left-handed?

- We are assuming that these people are independent, so we can use the multiplication rule:

$$\begin{aligned} P(L_1 \text{ and } L_2) &= P(L_1) \times P(L_2) \\ &= (0.09) \times (0.09) \\ &= 0.0081 \end{aligned}$$

or 0.81% (this is highly unlikely!)

Example: Both Right-Handed

What is the probability that both are right-handed?

- First, assume that everyone is either right- or left-handed.
- Then L_1^c is the event that the first person is right-handed and L_2^c is the event that the second person is right-handed.
- From the previous slide, we decided that $P(L_1) = P(L_2) = 0.09$
- So $P(L_1^c) = 1 - P(L_1) = 1 - 0.09 = 0.91$ and $P(L_2^c) = 0.91$

Example: Both Right-Handed

What is the probability that both are right-handed?

- We are still assuming that these people are independent, so we can again use the multiplication rule:

$$\begin{aligned} P(L_1^c \text{ and } L_2^c) &= P(L_1^c) \times P(L_2^c) \\ &= (0.91) \times (0.91) \\ &= 0.8281 \end{aligned}$$

or 82.81%.

Disjoint Events - Independent?

If two events are disjoint, are they independent?

Disjoint Events- Independent?

If two events are disjoint, are they independent?

- Recall that independent events have no relationship with one another.
- This means that if we know something about event A, we don't get any information about event B.
- For disjoint events, if event A occurs, we can be totally certain that event B did not occur.
- Therefore they are *dependent*.

Example

Consider two disjoint events for rolling a six-sided die. Let $A = \{1\}$ be the event that I roll a `1` and $B = \{2\}$ the event that I roll a `2`.

- If I know that A occurred, then I can be 100% sure that B did not occur.
- If I know that A did not occur, then I know that the roll must be a `2`, `3`, `4`, `5`, or `6`.
  - Now there are five possible options instead of six!
  - We've narrowed down our options, so knowing that I did not roll a `1` has given us some useful information.

Therefore A and B can't be independent.

Conditional Probability

We can get far more information out of the relationships between multiple variables than we can from a single variable.
For example

- Recall our case study on the malaria vaccine.
- We can look at P(infection), but that doesn't tell us anything about the efficacy of the vaccine.
- Instead, we want to look at the probability that a person develops infection *if they were vaccinated.*
- We compare this to the probability that a person develops infection if they were not vaccinated.

Contingency Table Probabilities

Let's consider a data set on a machine learning classifier.

- The classifier is designed to take images and determine whether each one is about fashion.
- The classifier groups 1822 photos into either "fashion" or "not fashion".
- Separately, these photos are grouped into "fashion" and "not fashion" by a group of people.
  - We take these groupings as the truth that the classifier is trying to get at.

Contingency Table Probabilities

We can take these groupings and build them into a contingency table.

		truth		
		Fashion	Not	Total
classifier	Fashion	197	22	219
	Not	112	1491	1603
	Total	309	1513	1822

Contingency Table Probabilities

We think about this a lot with classification problems!

		truth		
		fashion	not fashion	Total
classifier	pred fashion	197	22	219
	pred not	112	1491	1603
	Total	309	1513	1822

- When we build our classifier, we want to know the rate at which it correctly and incorrectly identifies `fashion` and `not fashion`.
- This will give us an idea of how successful our classifier is.
  - Is it a good classifier?
  - Should we try a different machine learning algorithm?

Example: Contingency Table Probabilities

1. If the photo is actually about fashion, what is the probability that the classifier correctly identified it as being about fashion?
2. If the classifier predicted that a photo was not about fashion, what is the probability that it was incorrect?

Example: Contingency Table Probabilities

If the photo is actually about fashion, what is the probability that the classifier correctly identified it as being about fashion?

		`truth`		
		`fashion`	`not fashion`	Total
`classifier`	`pred fashion`	197	22	219
	`pred not`	112	1491	1603
	Total	309	1513	1822

- We know that the photo is actually about fashion, so we focus our attention to the column where `truth` is `fashion`.
- Then within this column, we look for the number of times the classifier `pred fashion` out of the total number of `fashion` photos.

Example: Contingency Table Probabilities

If the photo is actually about fashion, what is the probability that the classifier correctly identified it as being about fashion?

		truth		
		fashion	not fashion	Total
classifier	pred fashion	197	22	219
	pred not	112	1491	1603
	Total	309	1513	1822

$$P(\texttt{classifier} \text{ is } \texttt{pred fashion} \; \textit{given} \; \texttt{truth} \text{ is } \texttt{fashion}) = \frac{197}{309}$$

or 0.638, a reasonable correct identification rate for fashion.

Example: Contingency Table Probabilities

If the classifier predicted that a photo was not about fashion, what is the probability that it was incorrect?

		`truth`		
		`fashion`	`not fashion`	Total
`classifier`	`pred fashion`	197	22	219
	`pred not`	112	1491	1603
	Total	309	1513	1822

- We know that `classifier` is `pred not` fashion, so we focus our attention to this row.
- We want to know the probability that it was incorrect, or in `truth` is `fashion`.

Example: Contingency Table Probabilities

If the classifier predicted that a photo was not about fashion, what is the probability that it was incorrect?

		truth		
		fashion	not fashion	Total
classifier	pred fashion	197	22	219
	pred not	112	1491	1603
	Total	309	1513	1822

$$P(\texttt{truth}\text{ is }\texttt{fashion}\ \textit{given}\ \texttt{classifier}\text{ is }\texttt{pred not}) = \frac{112}{1603}$$

or 0.070, a low misidentification rate for fashion photos.

Marginal and Joint Probabilities

		truth		
		fashion	not fashion	Total
classifier	pred fashion	197	22	219
	pred not	112	1491	1603
	Total	309	1513	1822

- We've now used our contingency table to think about two types of probabilities.
    - The probability for a single event (from the row and column of totals).
    - The probability for multiple events together (from the numbers in the middle).

Marginal Probabilities

- A **marginal probability** is a probability based on a single variable.
- Think of the *margins* as the edges of a contingency table where we have the information for each variable individually.

Marginal Probabilities

		truth		
		fashion	not fashion	Total
classifier	pred fashion	197	22	219
	pred not	112	1491	1603
	Total	309	1513	1822

A probability based solely on our `classifier` is a marginal probability. It is based on a single variable without regard to any other variables.

$$P(\texttt{classifier}\text{ is }\texttt{pred fashion}) = 219/1822$$

Joint Probabilities

- A **joint probability** is a probability for two or more variables together.
- Think of this as a probability that two or more variables occur *jointly* (together).

Joint Probabilities

		truth		
		fashion	not fashion	Total
classifier	pred fashion	197	22	219
	pred not	112	1491	1603
	Total	309	1513	1822

The probability that our `classifier` is `pred fashion` and the truth is `fashion` is a joint probability. It is based on two variables together.

$$P(\texttt{classifier}\text{ is }\texttt{pred fashion}\text{ and }\texttt{truth}\text{ is }\texttt{fashion}) = 197/1822$$

Table Proportions

We can examine marginal and joint probabilities using table proportions. **Table proportions** are computed by dividing each count in a contingency table by the table's grand total.

		truth		
		fashion	not fashion	Total
classifier	pred fashion	0.108	0.012	0.120
	pred not	0.062	0.818	0.880
	Total	0.170	0.830	1.000

Joint Probability Distributions

A joint probability distribution is just a probability distribution for multiple variables together.

Joint Outcome	Probability
`classifier` is `pred fashion` and `truth` is `fashion`	0.108
`classifier` is `pred fashion` and `truth` is `not fashion`	0.012
`classifier` is `pred not` and `truth` is `fashion`	0.062
`classifier` is `pred not` and `truth` is `not fashion`	0.818
Total	1.000

Note: A marginal probability distribution is the type of probability distribution we introduced last week!

Marginal and Joint Probabilities

We can compute marginal probabilities using joint probabilities.

Joint Outcome	Probability
`classifier` is `pred fashion` and `truth` is `fashion`	0.108
`classifier` is `pred fashion` and `truth` is `not fashion`	0.012
`classifier` is `pred not` and `truth` is `fashion`	0.062
`classifier` is `pred not` and `truth` is `not fashion`	0.818
Total	1.000

For example,

$$
\begin{aligned}
&P(\texttt{truth}\text{ is }\texttt{fashion})\\
&\qquad = P(\texttt{classifier}\text{ is }\texttt{pred fashion}\text{ and }\texttt{truth}\text{ is }\texttt{fashion})\\
&\qquad\quad + P(\texttt{classifier}\text{ is }\texttt{pred not}\text{ and }\texttt{truth}\text{ is }\texttt{fashion})\\
&\qquad = 0.108 + 0.062\\
&\qquad = 0.170
\end{aligned}
$$

Marginal and Joint Probabilities

This makes sense based on our table proportions!

		truth		
		fashion	not fashion	Total
classifier	pred fashion	0.108	0.012	0.120
	pred not	0.062	0.818	0.880
	Total	0.170	0.830	1.000

- All of these numbers are directly proportional to our original contingency table.
- The row and column of totals represent the marginal probabilities.
- These totals are the actual sums of their respective rows/columns.

Defining Conditional Probability

- The `classifier` predicts whether a photo is about `fashion`, but it is not perfect.
- We'd like to know how we can use these predictions to improve our understanding of the second variable, the `truth`.
- We might want to know, for example, the probability that the `truth` is `fashion` *given* that the `classifier predicts fashion`.

Defining Conditional Probability

The probability that a random photo from the data set is actually about fashion is 0.17. Suppose we know that `classifier` is `pred fashion`.

- Now we can get a better estimate of the probability that the `truth` is `fashion`.
- We do this by restricting our attention to the 219 cases where the `classifier` is `pred fashion`.
- Then we look at the fraction of *these* photos where the `truth` is `fashion` (197 cases).

$$P(\texttt{truth}\text{ is }\texttt{fashion}\text{ given }\texttt{classifier}\text{ is }\texttt{pred fashion}) = \frac{197}{219}$$

Defining Conditional Probability

- When we are given some useful information that allows us to restrict our attention, we call these probabilities **conditional probabilities**.
- We can say that we condition based on some *given* information, or that we computed the probability under the *condition* that the `classifier` is `pred fashion`.

Defining Conditional Probability

There are two important aspects to a conditional probability:

1. The **outcome of interest** is whatever we want to know about.
2. The **condition** is information we know to be true, a known outcome or event.

Conditional Probability Notation

We separate our outcome of interest from our condition in our probability notation with a vertical bar:

$$P(\texttt{truth}\text{ is }\texttt{fashion}\text{ given }\texttt{classifier}\text{ is }\texttt{pred fashion})$$

becomes

$$P(\texttt{truth}\text{ is }\texttt{fashion} \mid \texttt{classifier}\text{ is }\texttt{pred fashion}) = \frac{197}{219}$$

We read the vertical bar as the word *given*.

Defining Conditional Probability

Earlier, we computed

$$P(\texttt{truth} \text{ is } \texttt{fashion} \text{ given } \texttt{classifier} \text{ is } \texttt{pred fashion}) = 0.900$$

by restricting our attention to the data where `classifier` is `pred fashion`.

From this row where `classifier` is `pred fashion`, we took the number of cases where `truth` is `fashion` and divided by the row total to get our answer.

Defining Conditional Probability

However, we don't always have access to the count data. Instead we are given only the probabilities.

		truth		
		fashion	not fashion	Total
classifier	pred fashion	0.108	0.012	0.120
	pred not	0.062	0.818	0.880
	Total	0.170	0.830	1.000

Defining Conditional Probability

- Suppose we took a sample of 1000 photos.
- We could multiply each probability by 1000 to get an estimate of how many would fall into each place in our contingency table.
- We would anticipate $0.120 \times 1000 = 120$ to be the number of cases where `classifier` is `pred fashion`.
- We would expect to see $0.108 \times 1000 = 108$ cases where `truth` is `fashion` and `classifier` is `pred fashion`

Defining Conditional Probability

We can use these numbers to compute our conditional probability. (Using our count data, we found $197/219 = 0.90$.)

$$
\begin{aligned}
&P(\texttt{truth} \text{ is } \texttt{fashion} \text{ given } \texttt{classifier} \text{ is } \texttt{pred fashion}) \\
&\quad= \frac{\#\text{ cases } (\texttt{truth} \text{ is } \texttt{fashion} \text{ and } \texttt{classifier} \text{ is } \texttt{pred fashion})}{\#\text{ cases } (\texttt{classifier} \text{ is } \texttt{pred fashion})} \\
&\quad= \frac{108}{120} = \frac{0.108 \times 1000}{0.120 \times 1000} = \frac{0.108}{0.120} = 0.90
\end{aligned}
$$

Defining Conditional Probability

This is the ratio, or fraction, or two probabilities. We can rewrite this as

$$\begin{aligned}
&P(\texttt{truth}\text{ is }\texttt{fashion}\text{ given }\texttt{classifier}\text{ is }\texttt{pred fashion})\\
&\quad= \frac{P(\texttt{truth}\text{ is }\texttt{fashion}\text{ and }\texttt{classifier}\text{ is }\texttt{pred fashion})}{P(\texttt{classifier}\text{ is }\texttt{pred fashion})}\\
&\quad= \frac{0.108}{0.120} = 0.90
\end{aligned}$$

Defining Conditional Probability

This leads us to the general **conditional probability formula**:

Let A and B be outcomes. The conditional probability of outcome A occurring given the condition that B has occurred is

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

Example

Find the probability that the classifier is incorrect when classifying a photo about fashion.

Example

Find the probability that the classifier is incorrect when classifying a photo about fashion.

- We *know* that the photo is about fashion.
  - We can write that `truth` is `fashion`.
  - This information is given, or our condition.
- From that, we want to know the probability that the classifier is wrong.
  - We want to know the probability that the `classifier` results in `not fashion`.

Example

Find the probability that the classifier is incorrect when classifying a photo about fashion.

Putting this all together, we want

$$P(\texttt{classifier}\text{ is }\texttt{not fashion} \mid \texttt{truth}\text{ is }\texttt{fashion})$$

Example

Using our formula

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

we let A be the event that `classifier` is `not fashion` and B the event that `truth` is `fashion`. Then

$$\begin{aligned} &P(\texttt{classifier} \text{ is } \texttt{not fashion} \mid \texttt{truth} \text{ is } \texttt{fashion}) \\ &\quad = \frac{P(\texttt{classifier} \text{ is } \texttt{not fashion} \text{ and } \texttt{truth} \text{ is } \texttt{fashion})}{P(\texttt{truth} \text{ is } \texttt{fashion})} \end{aligned}$$

Example

		truth		
		fashion	not fashion	Total
classifier	pred fashion	0.108	0.012	0.120
	pred not	0.062	0.818	0.880
	Total	0.170	0.830	1.000

$$
\begin{aligned}
&P(\texttt{classifier}\text{ is }\texttt{not fashion} \mid \texttt{truth}\text{ is }\texttt{fashion})\\
&\quad= \frac{P(\texttt{classifier}\text{ is }\texttt{not fashion}\ \text{ and }\ \texttt{truth}\text{ is }\texttt{fashion})}{P(\texttt{truth}\text{ is }\texttt{fashion})}\\
&\quad= \frac{0.062}{0.170} = 0.363
\end{aligned}
$$

Example: Smallpox

The smallpox data set is a sample of 6224 individuals from the year 1721.

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

Example: Smallpox

The smallpox data set has the following table proportions:

		inoculated		
		yes	no	Total
result	lived	0.038	0.825	0.863
	died	0.001	0.136	0.137
	Total	0.039	0.961	1.000

Let's find the probability that an inoculated person died from smallpox.

Example: Smallpox

Find the probability that an inoculated person died from smallpox.

- We are told that the person is inoculated. This is our condition.
- We want to know the probability that this person died.
- This is the probability that a person died given that they were inoculated

$$P(\texttt{died} \mid \texttt{inoculated})$$

Example: Smallpox

Find the probability that an inoculated person died from smallpox.

		inoculated		
		yes	no	Total
result	lived	0.038	0.825	0.863
	died	0.001	0.136	0.137
	Total	0.039	0.961	1.000

$$\begin{aligned} P(\texttt{died} \mid \texttt{inoculated}) &= \frac{P(\texttt{died}\text{ and }\texttt{inoculated})}{P(\texttt{inoculated})} \\ &= \frac{0.001}{0.039} = 0.026 \end{aligned}$$

General Multiplication Rule

In the previous section, we talked about the multiplication rule for independent events. The **general multiplication rule** is for all events, whether or not they are independent.

Let A and B be any two outcomes or events. Then

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

Notice that this is not new information! This is just a rearrangement of the formula for conditional probability.

Example

Let's return to the `smallpox` data set, but suppose we only have two pieces of information:

1. 96.08% of people were not inoculated.
2. 85.88% of people who were not inoculated ended up surviving.

Can we compute the probability that a resident was not inoculated and lived?

Example

Compute the probability that a resident was not inoculated and lived.

First, let's rewrite the information we were given in probability notation.

- 96.08% of people were not inoculated
  $\rightarrow P(\texttt{inoculated} = \texttt{no}) = 0.9608$
- 85.88% of people who were not inoculated ended up surviving
  $\rightarrow P(\texttt{result} = \texttt{lived} \mid \texttt{inoculated} = \texttt{no}) = 0.8588$

Example

Compute the probability that a resident was not inoculated and lived.

Then we use this information with the general multiplication rule.

$$\begin{aligned} &P(\texttt{result} = \texttt{lived} \text{ and } \texttt{inoculated} = \texttt{no}) \\ &\quad = P(\texttt{result} = \texttt{lived} \mid \texttt{inoculated} = \texttt{no}) \times P(\texttt{inoculated} = \texttt{no}) \\ &\quad = 0.9608 \times 08588 \\ &\quad = 0.8251. \end{aligned}$$

Sum of Conditional Probabilities

Let $A_1, \ldots, A_k$ represent all the disjoint outcomes for a variable or process. Then if B is some event,

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

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

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

Why are these true? Let's look at a Venn diagram.

Independence Considerations

For two independent events, knowing the outcome of one should give us no information about the probability of the other. Consider X and Y, the outcomes for rolling two six-sided dice.

1. Find $P(X = 1)$.
2. Find $P(X = 1 \text{ and } Y = 1)$.
3. Find $P(Y = 1 | X = 1)$.

Knowing the outcome of X doesn't give us any additional information about Y.

Independence Considerations

We can use the Multiplication Rule to show that the conditioning information has no influence for independent processes:

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

Example: The Gambler's Fallacy

A roulette wheel has 18 black slots, 18 red slots, and 2 green slots (38 total slots).

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?

Example: The Gambler's Fallacy

What is wrong with Ron's reasoning?

- It's true that there is close to a $1/64 = 0.016$ chance that we get `black` six times in a row.
  - $P(\texttt{black}_1) \times \cdots \times P(\texttt{black}_5) \times P(\texttt{black}_6) = (9/19)^6 = 0.011$
- But there's also a $1/64$ chance that we get `black` five times in a row followed by `red`.
  - $P(\texttt{black}_1) \times \cdots \times P(\texttt{black}_5) \times P(\texttt{red}_6) = (9/19)^6 = 0.011$

Example: The Gambler's Fallacy

What is wrong with Ron's reasoning?

- Each spin is independent of the previous spins!
- This means that each spin has a 18/38 chance of being `black`!
- Ron has a $1 - \frac{18}{38} = 0.538$ chance of losing his entire paycheck.

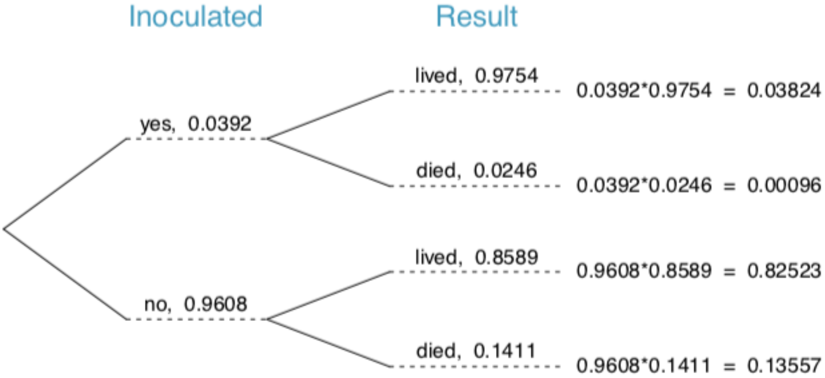
Tree Diagrams

Tree diagrams help organize outcomes and probabilities based on the structure of the data. They are especially useful when the data can be put into some kind of sequential structure.

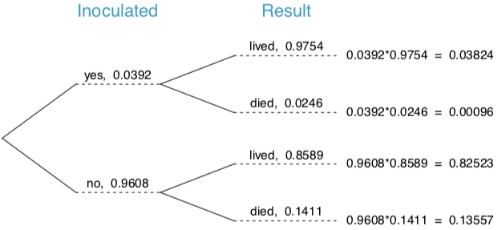
Tree Diagrams

- The `smallpox` data can be structured this way.
- We split the data by `inoculation` (`yes` or `no`).
- Then we split by `result` (`lived` or `died`).

Tree Diagrams

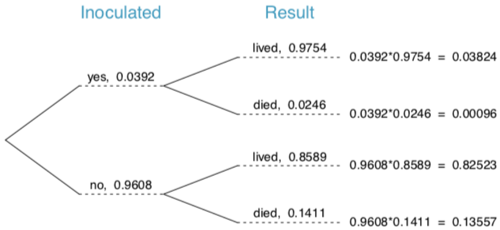


Tree Diagrams



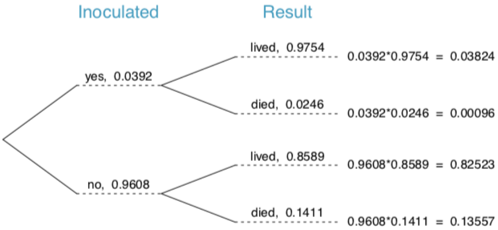
- The first branch, for `inoculation`, is called the **primary branch**.
- All other branches, in this case for `result` are **secondary branches**.

Tree Diagrams



- The probabilities for the primary branch are marginal.
  - For `inoculation` is `yes`, the marginal probability is $P(\texttt{inoculation} \text{ is } \texttt{yes}) = 0.0392$.
- The probabilities for the secondary branches are conditional.
  - For `result` is `lived` on the `inoculation` is `yes` branch, we have $P(\texttt{result} \text{ is } \texttt{lived} \mid \texttt{inoculation} \text{ is } \texttt{yes}) = 0.9754$

Tree Diagrams



- Joint probabilities are shown to the right of each secondary branch.
- These are computed using the General Multiplication Rule

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

where the primary branch represents event B and the secondary branch event A.

Example: Exam Scores

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.
- 11% of the students who earned lower than an A on the midterm received an A on the final.
- You pick up a final exam at random and notice the student received an A.

What is the probability that this student earned an A on the midterm?

Example: Exam Scores

Let's start by writing the given information in probability notation.

- $P(\texttt{midterm} = \texttt{A}) = 0.13$
- $P(\texttt{final} = \texttt{A} \mid \texttt{midterm} = \texttt{A}) = 0.47$
- $P(\texttt{final} = \texttt{A} \mid \texttt{midterm} = \texttt{not A}) = 0.11$

We want to know the probability that a student who earned an A on the final also earned an A on the midterm:

$$P(\texttt{midterm} = \texttt{A} \mid \texttt{final} = \texttt{A})$$

Example: Exam Scores

Now that we've formalized the information from the problem statement, we can consider our next steps.

It's not yet clear how to calculate

$$P(\texttt{midterm} = \texttt{A} \mid \texttt{final} = \texttt{A}),$$

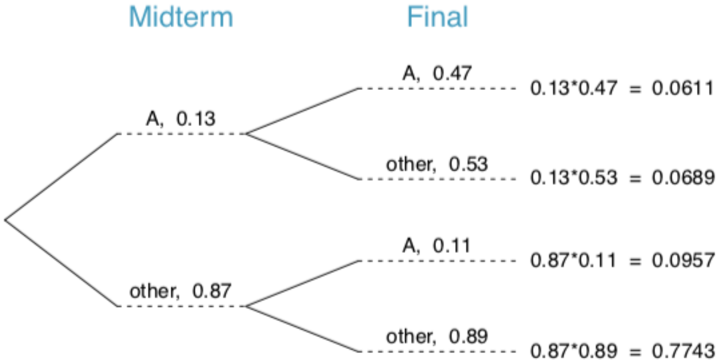
so let's use what we know to draw a tree diagram.

Example: Exam Scores

We will use this information to draw our tree diagram.

- $P(\texttt{midterm} = \texttt{A}) = 0.13$
- $P(\texttt{final} = \texttt{A} \mid \texttt{midterm} = \texttt{A}) = 0.47$
- $P(\texttt{final} = \texttt{A} \mid \texttt{midterm} = \texttt{not A}) = 0.11$

Example: Exam Scores



Can we use this to calculate $P(\texttt{midterm} = \texttt{A} \mid \texttt{final} = \texttt{A})$?

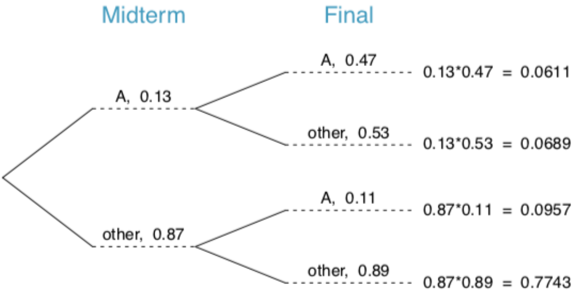
Example: Exam Scores

First, consider our conditional probability formula.

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

We can get all of the probabilities on the right hand side of the formula by using our tree diagram!

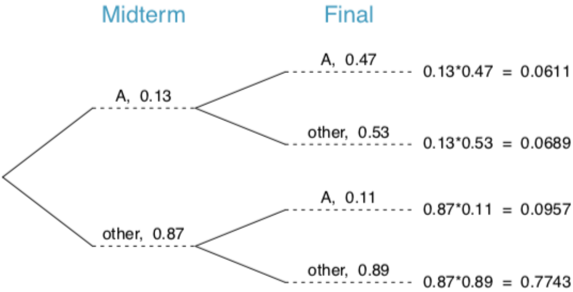
Example: Exam Scores



First,

$$P(\texttt{midterm} = \texttt{A} \text{ and } \texttt{final} = \texttt{A}) = 0.0611.$$

Example: Exam Scores



Then

$$
\begin{aligned}
&P(\texttt{final} = \texttt{A}) \\
&\quad = P(\texttt{midterm} = \texttt{not A} \text{ and } \texttt{final} = \texttt{A}) + P(\texttt{midterm} = \texttt{A} \text{ and } \texttt{final} = \texttt{A}) \\
&\quad = 0.0957 + 0.0611 = 0.1568
\end{aligned}
$$

Example: Exam Scores

Plugging these in,

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

So the probability that a student earned an A on the midterm, given that their final exam score was an A, is about 39%.

Bayes' Theorem

That was a lot of work!

Bayes' Theorem will help minimize this work so that we can more easily calculate

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

when we have information about

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