Ling/CSC 439/539
Introduction to Probability Theory

Fall 2017

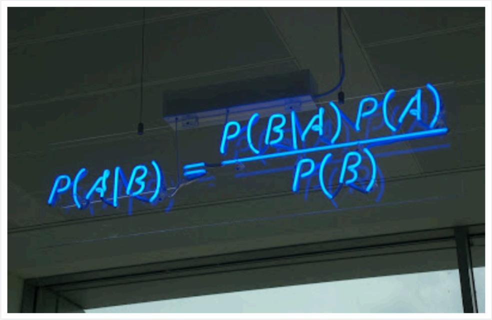
Take-away

Probability Theory Summary

- Representing conditional probabilities using contingency tables and probability trees
- The chain rule
- Bayes rule
- The law of total probability (aka the partition rule)
- Independent events

Motivation

Eleven Equations True Computer Science Geeks Should (at Least Pretend to) Know



http://www.elegantcoding.com/2011/11/eleven-equations-true-computer-science.html

How juries are fooled by statistics

http://www.ted.com/talks/lang/eng/peter_donnelly_shows_how_stats_fool_juries.html?

Outline

Definition

- How would you define "probability"?

Example: Blood Types

- Human blood is classified by the presence or absence of two antigens, called A and B. This gives rise to four types: O, A, B, and AB.

		Antigen B Absent	Antigen B Present	Marginal
Antigen A	Absent	0.44	0.10	0.54
	Present	0.42	0.04	0.46
	Marginal	0.86	0.14	1.00

Table: Probability Estimates for U.S. Blood Types

- Let A be the event "presence of antigen A" and B be the event "presence of antigen B"

Example: Blood Types

		Antigen B Absent	Antigen B Present	Marginal
Antigen A	Absent	0.44	0.10	0.54
	Present	0.42	0.04	0.46
	Marginal	0.86	0.14	1.00

Table: Probability Estimates for U.S. Blood Types

- Exercise!

Outline

Examples of Conditional Probabilities

Conditional Probability

Bayes' Rule

Independence

Conditioning Changes the Sample Space

- Before we knew anything, anything in sample space S could occur.

Conditioning Changes the Sample Space

- Before we knew anything, anything in sample space S could occur.
- After we know B happened, we are only choosing from within B.

Conditioning Changes the Sample Space

- Before we knew anything, anything in sample space S could occur.
- After we know B happened, we are only choosing from within B.
- **The set B becomes our new sample space, with an updated probability of 1!**

Conditioning Changes the Sample Space

- Before we knew anything, anything in sample space S could occur.
- After we know B happened, we are only choosing from within B.
- **The set B becomes our new sample space, with an updated probability of 1!**
- Instead of asking "In what proportion of S is D true?", we now ask "In what proportion of B is D true?"

Conditioning Changes the Sample Space

- We calculate conditional *proportions* from frequencies by restricting attention to ("conditioning on") a particular category, and dividing the joint frequency by the restricted total.
- Conditional probability is defined in exactly the same way, replacing proportions with probabilities.

Conditional Probability

To find the probability of *A given B*, consider the ways A can occur *in the context of B* (i.e., $A \cap B$), out of all the ways B can occur.

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

Joint Probability from Conditional Probability

- We can rearrange the formula for conditional probability.

Joint Probability from Conditional Probability

- We can rearrange the formula for conditional probability.
- Knowing that

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

Joint Probability from Conditional Probability

- We can rearrange the formula for conditional probability.
- Knowing that

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

- We can easily derive this:

Joint Probability from Conditional Probability

- We can rearrange the formula for conditional probability.
- Knowing that

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

- We can easily derive this:

The "Chain Rule" of Probability

For any events, A and B, the joint probability $P(A \cap B)$ can be computed as

$$P(A \cap B) = P(A|B) \times P(B)$$

Or, since $P(A \cap B) = P(B \cap A)$

$$P(A \cap B) = P(B|A) \times P(A)$$

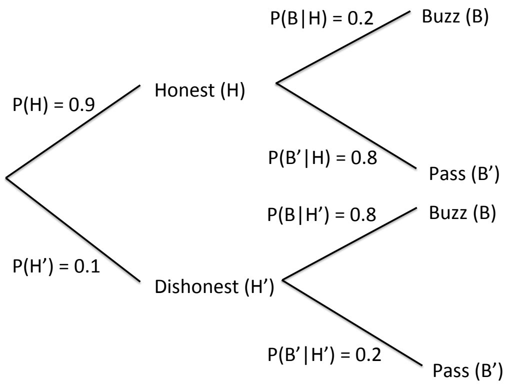
Example: Lie Detector

- A store owner discovers that some of her employees have taken cash. She decides to use a lie detector to discover who they are.
- Suppose that 10% of employees stole, but 100% say they didn't.
- The lie detector buzzes 80% of the time that someone lies, and 20% of the time that someone is telling the truth.
- If the detector buzzes, what's the probability that the person was lying?

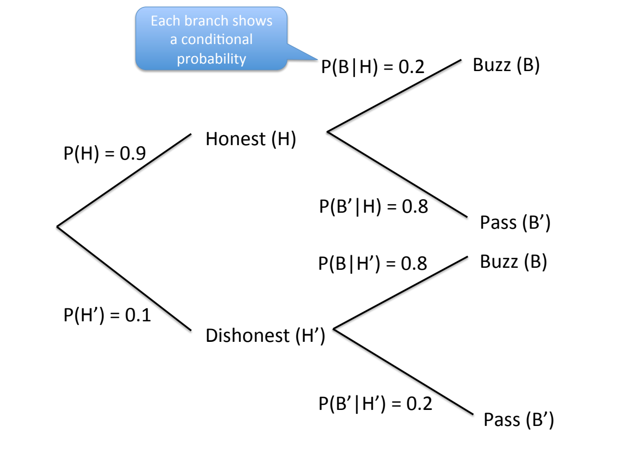
Let's View These Rules Using Probability Trees

Probability tree == decision tree with probabilities attached to each decision

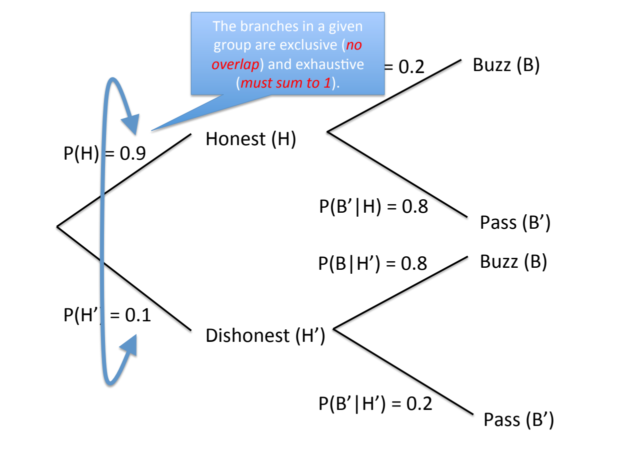
Let's View These Rules Using Probability Trees



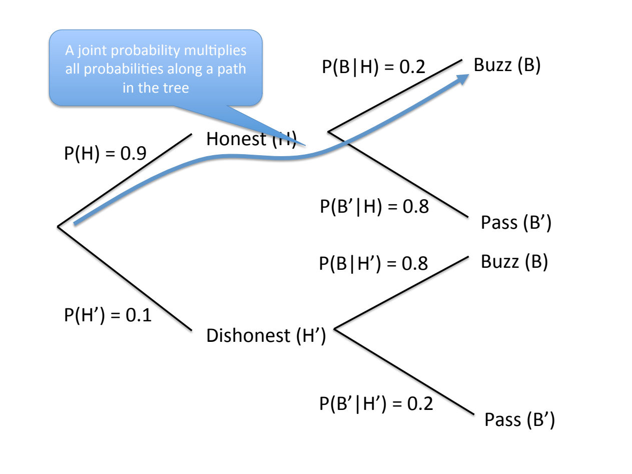
Let's View These Rules Using Probability Trees



Let's View These Rules Using Probability Trees



Let's View These Rules Using Probability Trees



Let's View These Rules Using Probability Trees

- What is $P(\text{Pass} \cap \text{Dishonest})$?

Let's View These Rules Using Probability Trees

- What is $P(\text{Pass} \cap \text{Dishonest})$?
- What is $P(\text{Buzz})$? Hint: which branches end up with the Buzz event?

Let's View These Rules Using Probability Trees

- What is $P(\text{Pass} \cap \text{Dishonest})$?
- What is $P(\text{Buzz})$? Hint: which branches end up with the Buzz event?
- What is $P(\text{Dishonest}|\text{Buzz})$?

Let's View These Rules Using Probability Trees

- What is $P(\text{Pass} \cap \text{Dishonest})$?
- What is $P(\text{Buzz})$? Hint: which branches end up with the Buzz event?
- What is $P(\text{Dishonest}|\text{Buzz})$?

- Computing conditional probabilities going against the flow of the tree is a bit tricky. We will see a better way next!

Outline

Reversing Conditional Probabilities

- 16% of the students are Nutrition Science majors and 55% are female. Of the Nutrition Science majors, 54% are female. What proportion of female students in the class are Nutrition Science majors?

Reversing Conditional Probabilities

- 16% of the students are Nutrition Science majors and 55% are female. Of the Nutrition Science majors, 54% are female. What proportion of female students in the class are Nutrition Science majors?
- I'm looking for $P()$

Reversing Conditional Probabilities

- 16% of the students are Nutrition Science majors and 55% are female. Of the Nutrition Science majors, 54% are female. What proportion of female students in the class are Nutrition Science majors?
- I'm looking for $P(N|F)$

Reversing Conditional Probabilities

- 16% of the students are Nutrition Science majors and 55% are female. Of the Nutrition Science majors, 54% are female. What proportion of female students in the class are Nutrition Science majors?
- I'm looking for $P(N|F)$
- I can find it by using the fact that

$$P(N|F) = \frac{P()}{P()}$$

Reversing Conditional Probabilities

- 16% of the students are Nutrition Science majors and 55% are female. Of the Nutrition Science majors, 54% are female. What proportion of female students in the class are Nutrition Science majors?
- I'm looking for $P(N|F)$
- I can find it by using the fact that

$$P(N|F) = \frac{P(F \cap N)}{P(F)}$$

Reversing Conditional Probabilities

- 16% of the students are Nutrition Science majors and 55% are female. Of the Nutrition Science majors, 54% are female. What proportion of female students in the class are Nutrition Science majors?
- I'm looking for $P(N|F)$
- I can find it by using the fact that

$$P(N|F) = \frac{P(F \cap N)}{P(F)}$$

- I have $P(F) =$

Reversing Conditional Probabilities

- 16% of the students are Nutrition Science majors and 55% are female. Of the Nutrition Science majors, 54% are female. What proportion of female students in the class are Nutrition Science majors?
- I'm looking for $P(N|F)$
- I can find it by using the fact that

$$P(N|F) = \frac{P(F \cap N)}{P(F)}$$

- I have $P(F) = 0.55$

Reversing Conditional Probabilities

- 16% of the students are Nutrition Science majors and 55% are female. Of the Nutrition Science majors, 54% are female. What proportion of female students in the class are Nutrition Science majors?
- I'm looking for $P(N|F)$
- I can find it by using the fact that

$$P(N|F) = \frac{P(F \cap N)}{P(F)}$$

- I have $P(F) = 0.55$
- To find $P(F \cap N)$ I can use the fact that

$$P(F \cap N) = P() \times P()$$

Reversing Conditional Probabilities

- 16% of the students are Nutrition Science majors and 55% are female. Of the Nutrition Science majors, 54% are female. What proportion of female students in the class are Nutrition Science majors?
- I'm looking for $P(N|F)$
- I can find it by using the fact that

$$P(N|F) = \frac{P(F \cap N)}{P(F)}$$

- I have $P(F) = 0.55$
- To find $P(F \cap N)$ I can use the fact that

$$P(F \cap N) = P(N) \times P(F|N)$$

Reversing Conditional Probabilities

- 16% of the students are Nutrition Science majors and 55% are female. Of the Nutrition Science majors, 54% are female. What proportion of female students in the class are Nutrition Science majors?
- I'm looking for $P(N|F)$
- I can find it by using the fact that

$$P(N|F) = \frac{P(F \cap N)}{P(F)}$$

- I have $P(F) = 0.55$
- To find $P(F \cap N)$ I can use the fact that

$$P(F \cap N) = P(N) \times P(F|N)$$

- So altogether, I have

$$P(N|F) = \frac{P()P()}{P()}$$

Reversing Conditional Probabilities

- 16% of the students are Nutrition Science majors and 55% are female. Of the Nutrition Science majors, 54% are female. What proportion of female students in the class are Nutrition Science majors?
- I'm looking for $P(N|F)$
- I can find it by using the fact that

$$P(N|F) = \frac{P(F \cap N)}{P(F)}$$

- I have $P(F) = 0.55$
- To find $P(F \cap N)$ I can use the fact that

$$P(F \cap N) = P(N) \times P(F|N)$$

- So altogether, I have

$$P(N|F) = \frac{P(N)P(F|N)}{P(F)}$$

Reversing Conditional Probabilities

- 16% of the students are Nutrition Science majors and 55% are female. Of the Nutrition Science majors, 54% are female. What proportion of female students in the class are Nutrition Science majors?
- I'm looking for $P(N|F)$
- I can find it by using the fact that

$$P(N|F) = \frac{P(F \cap N)}{P(F)}$$

- I have $P(F) = 0.55$
- To find $P(F \cap N)$ I can use the fact that

$$P(F \cap N) = P(N) \times P(F|N)$$

- So altogether, I have

$$P(N|F) = \frac{P(N)P(F|N)}{P(F)} = \frac{0.54 \times 0.16}{0.55}$$

Bayes' Rule

We can reverse conditional probabilities using **Bayes' Rule**

Bayes' Rule

For any two events, A and B, we have

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

Bayes' Rule

We can reverse conditional probabilities using **Bayes' Rule**

Bayes' Rule

For any two events, A and B, we have

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

- Remember: use Bayes Rule to reverse conditional probabilities.
- Useful to infer causes from effects (inferential statistics)!
- Very easy to derive from the chain rule, so remember that first.

Bayes' Rule

Bayes' Rule

For any two events, A and B, we have

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

- The **prior** (baseline) probability of having Lycanthropy is 1 in 1000.
- A test has been developed which gives a positive result for 9 in 10 werewolves and 1 in 20 non-werewolves.
- What's the conditional probability of having Lycanthropy, given a positive test result?

Bayes' Rule

Bayes' Rule

For any two events, A and B, we have

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

- The **prior** (baseline) probability of having Lycanthropy is 1 in 1000: $P() = 1/1000$

Bayes' Rule

Bayes' Rule

For any two events, A and B, we have

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

- The **prior** (baseline) probability of having Lycanthropy is 1 in 1000: $P(L) = 1/1000$

Bayes' Rule

Bayes' Rule

For any two events, A and B, we have

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

- The **prior** (baseline) probability of having Lycanthropy is 1 in 1000: $P(L) = 1/1000$
- A test has been developed which gives a positive result for 9 in 10 werewolves: $P() = 9/10$

Bayes' Rule

Bayes' Rule

For any two events, A and B, we have

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

- The **prior** (baseline) probability of having Lycanthropy is 1 in 1000: $P(L) = 1/1000$
- A test has been developed which gives a positive result for 9 in 10 werewolves: $P(T|L) = 9/10$

Bayes' Rule

Bayes' Rule

For any two events, A and B, we have

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

- The **prior** (baseline) probability of having Lycanthropy is 1 in 1000: $P(L) = 1/1000$
- A test has been developed which gives a positive result for 9 in 10 werewolves: $P(T|L) = 9/10$
- and 1 in 20 non-werewolves: $P() = 1/20$

Bayes' Rule

Bayes' Rule

For any two events, A and B, we have

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

- The **prior** (baseline) probability of having Lycanthropy is 1 in 1000: $P(L) = 1/1000$
- A test has been developed which gives a positive result for 9 in 10 werewolves: $P(T|L) = 9/10$
- and 1 in 20 non-werewolves: $P(T|L') = 1/20$

Bayes' Rule

Bayes' Rule

For any two events, A and B, we have

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

- The **prior** (baseline) probability of having Lycanthropy is 1 in 1000: $P(L) = 1/1000$
- A test has been developed which gives a positive result for 9 in 10 werewolves: $P(T|L) = 9/10$
- and 1 in 20 non-werewolves: $P(T|L') = 1/20$
- What's the conditional probability of having Lycanthropy, given a positive test result: $P() =?$

Bayes' Rule

Bayes' Rule

For any two events, A and B, we have

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

- The **prior** (baseline) probability of having Lycanthropy is 1 in 1000: $P(L) = 1/1000$
- A test has been developed which gives a positive result for 9 in 10 werewolves: $P(T|L) = 9/10$
- and 1 in 20 non-werewolves: $P(T|L') = 1/20$
- What's the conditional probability of having Lycanthropy, given a positive test result: $P(L|T) =?$

Bayes' Rule

Bayes' Rule for Werewolves

We have

$$P(L|T) = \frac{P(L)P(T|L)}{P(T)}$$

- Do we have everything we need?

Bayes' Rule

Bayes' Rule for Werewolves

We have

$$P(L|T) = \frac{P(L)P(T|L)}{P(T)}$$

- Do we have everything we need?
- We know $P(L) = 1/1000$, $P(T|L) = 9/10$. We're missing $P(T)$.

Bayes' Rule

Bayes' Rule for Werewolves

We have

$$P(L|T) = \frac{P(L)P(T|L)}{P(T)}$$

- Do we have everything we need?
- We know $P(L) = 1/1000$, $P(T|L) = 9/10$. We're missing $P(T)$.
- We also have $P(T|L')$, which we haven't used yet...

Bayes' Rule

Bayes' Rule for Werewolves

We have

$$P(L|T) = \frac{P(L)P(T|L)}{P(T)}$$

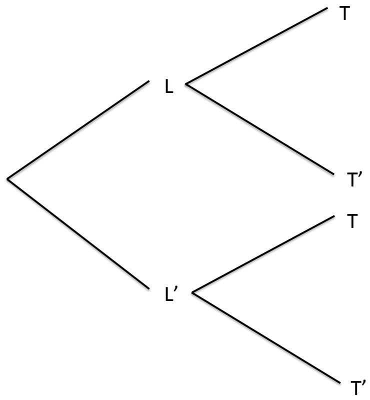
- Do we have everything we need?
- We know $P(L) = 1/1000$, $P(T|L) = 9/10$. We're missing $P(T)$.
- We also have $P(T|L')$, which we haven't used yet...
- Can we find $P(T)$ from what we have?

How to Compute $P(T)$?

- Two equivalent ways of doing it:
  - Contingency tables
  - Probability trees
- Use whichever is easiest for you.
  - In this course we will use probability trees

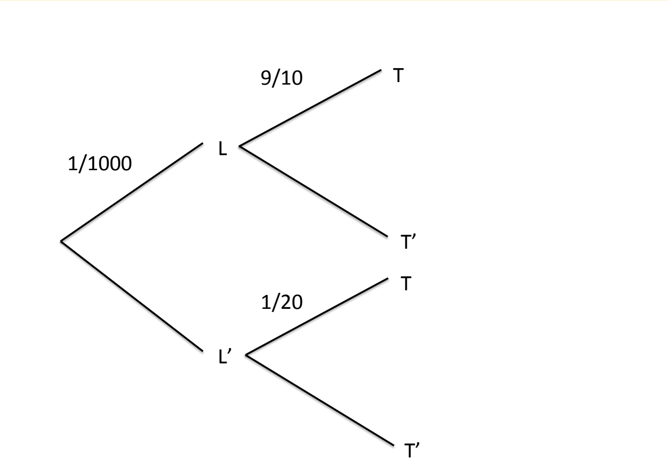
To the Probability Tree!

$$P(L) = 1/1000, P(T|L) = 9/10, P(T|L') = 1/20$$



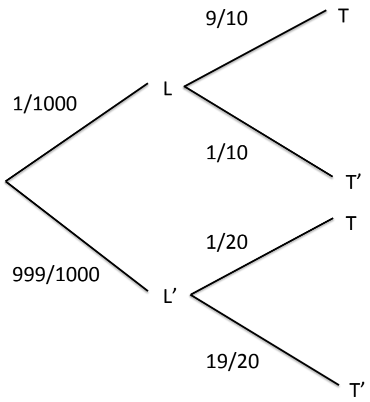
To the Probability Tree!

$$P(L) = 1/1000, P(T|L) = 9/10, P(T|L') = 1/20$$



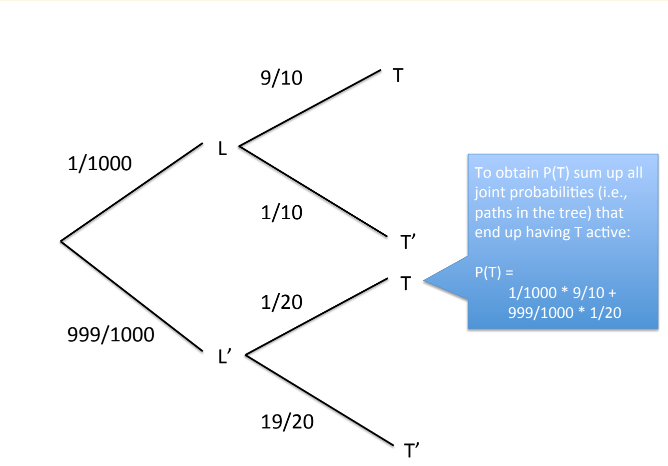
To the Probability Tree!

$$P(L) = 1/1000, P(T|L) = 9/10, P(T|L') = 1/20$$



To the Probability Tree!

$$P(L) = 1/1000, P(T|L) = 9/10, P(T|L') = 1/20$$



The Law of Total Probability

- What was involved in finding the marginal probability $P(T)$?
    1. $P(T)$ was the sum of two joint probabilities, $P(L \cap T)$ and $P(L' \cap T)$.

The Law of Total Probability

- What was involved in finding the marginal probability $P(T)$?
  1. $P(T)$ was the sum of two joint probabilities, $P(L \cap T)$ and $P(L' \cap T)$.
  2. To find the joint probabilities, we multiplied a marginal by a conditional, according to the "chain rule".

The Law of Total Probability

- What was involved in finding the marginal probability $P(T)$?
  1. $P(T)$ was the sum of two joint probabilities, $P(L \cap T)$ and $P(L' \cap T)$.
  2. To find the joint probabilities, we multiplied a marginal by a conditional, according to the "chain rule".

The Law of Total Probability

- What was involved in finding the marginal probability $P(T)$?
  1. $P(T)$ was the sum of two joint probabilities, $P(L \cap T)$ and $P(L' \cap T)$.
  2. To find the joint probabilities, we multiplied a marginal by a conditional, according to the "chain rule".

The Law of Total Probability

For any events A and B, we have $P(B) = P(A \cap B) + P(A' \cap B)$. Together with two applications of the chain rule, this gives us

$$P(B) = P(A)P(B|A) + P(A')P(B|A')$$

Bayes' Rule Revisited

Incorporating the Law of Total Probability into Bayes' Rule, we get

Bayes Rule (version 2)

For any events A and B, we have

$$\begin{aligned} P(A|B) &= \frac{P(A \cap B)}{P(A \cap B) + P(A' \cap B)} \\ &= \frac{P(A)P(B|A)}{P(A)P(B|A) + P(A')P(B|A')} \end{aligned}$$

Outline

Probabilistic Independence

- 10% of employees are dishonest.

Probabilistic Independence

- 10% of employees are dishonest.
- There's a 5% chance of rain tomorrow.

Probabilistic Independence

- 10% of employees are dishonest.
- There's a 5% chance of rain tomorrow.
- What's the probability an employee is dishonest if it rains tomorrow?

Probabilistic Independence

- 10% of employees are dishonest.
- There's a 5% chance of rain tomorrow.
- What's the probability an employee is dishonest if it rains tomorrow?
- Probably your intuition is that one conveys no information about the other. What does this mean about the relationship between the conditional probability of D given R, and the marginal probability of D?

Probabilistic Independence

Independent Events

We say that event A is **independent** of event B if conditioning on B does not change the probability of A, that is if

$$P(A|B) = P(A)$$

Probabilistic Independence

Independent Events

We say that event A is **independent** of event B if conditioning on B does not change the probability of A, that is if

$$P(A|B) = P(A)$$

- If the employee is dishonest, what's the probability that it will rain tomorrow?

Probabilistic Independence

Independent Events

We say that event A is **independent** of event B if conditioning on B does not change the probability of A, that is if

$$P(A|B) = P(A)$$

- If the employee is dishonest, what's the probability that it will rain tomorrow?
- It seems like independence should be symmetric. Is it?

Probabilistic Independence

- If A is independent of B, then $P(A|B) = P(A)$. Is $P(B|A)$ also equal to $P(B)$?

Probabilistic Independence

- If A is independent of B, then $P(A|B) = P(A)$. Is $P(B|A)$ also equal to $P(B)$?
- Using Bayes' rule, we have

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

Probabilistic Independence

- If A is independent of B, then $P(A|B) = P(A)$. Is $P(B|A)$ also equal to $P(B)$?
- Using Bayes' rule, we have

$$\begin{aligned} P(B|A) &= \frac{P(B)P(A|B)}{P(A)} \\ &= \frac{P(B)P(A)}{P(A)} \end{aligned}$$

Probabilistic Independence

- If A is independent of B, then $P(A|B) = P(A)$. Is $P(B|A)$ also equal to $P(B)$?
- Using Bayes' rule, we have

$$
\begin{aligned}
P(B|A) &= \frac{P(B)P(A|B)}{P(A)} \\
&= \frac{P(B)P(A)}{P(A)} \\
&= P(B)
\end{aligned}
$$

Probabilistic Independence

- If A is independent of B, then $P(A|B) = P(A)$. Is $P(B|A)$ also equal to $P(B)$?
- Using Bayes' rule, we have

$$\begin{aligned} P(B|A) &= \frac{P(B)P(A|B)}{P(A)} \\ &= \frac{P(B)P(A)}{P(A)} \\ &= P(B) \end{aligned}$$

- So independence is in fact symmetric.

Probabilistic Independence

- It turns out that if A and B are independent, then their joint probability has a particularly simple form

$$\begin{aligned} P(A \cap B) &= P(A)P(B|A) \\ &= P(A)P(B) \end{aligned}$$

Probabilistic Independence

- It turns out that if A and B are independent, then their joint probability has a particularly simple form

$$\begin{aligned} P(A \cap B) &= P(A)P(B|A) \\ &= P(A)P(B) \end{aligned}$$

Independence (version 2)

If A and B are independent events, then

$$P(A \cap B) = P(A)P(B)$$

Probabilistic Independence

- It turns out that if A and B are independent, then their joint probability has a particularly simple form

$$\begin{aligned} P(A \cap B) &= P(A)P(B|A) \\ &= P(A)P(B) \end{aligned}$$

Independence (version 2)

If A and B are independent events, then

$$P(A \cap B) = P(A)P(B)$$

- What is $P(A \cup B)$ for independent events?

Are these Independent Events?

1. E: First coin comes up heads, F: Second coin comes up heads

Are these Independent Events?

1. E: First coin comes up heads, F: Second coin comes up heads
2. E: First coin comes up heads, F: First coin comes up tails

Are these Independent Events?

1. E: First coin comes up heads, F: Second coin comes up heads
2. E: First coin comes up heads, F: First coin comes up tails
3. E: Sample a nutrition science major, F: Sample a female

Are these Independent Events?

1. E: First coin comes up heads, F: Second coin comes up heads
2. E: First coin comes up heads, F: First coin comes up tails
3. E: Sample a nutrition science major, F: Sample a female
4. E: Draw a yellow ball on first pick, F: Draw a yellow ball on second pick (with replacement)

Are these Independent Events?

1. E: First coin comes up heads, F: Second coin comes up heads
2. E: First coin comes up heads, F: First coin comes up tails
3. E: Sample a nutrition science major, F: Sample a female
4. E: Draw a yellow ball on first pick, F: Draw a yellow ball on second pick (with replacement)
5. E: Draw a yellow ball on first pick, F: Draw a yellow ball on second pick (without replacement)

Are these Independent Events?

1. We remove socks from a drawer with black and white, one after another, until we get a match. Is the color of the first sock independent of the color of the second?

Are these Independent Events?

1. We remove socks from a drawer with black and white, one after another, until we get a match. Is the color of the first sock independent of the color of the second?
2. E_1: it rains today. E_2: today is Thursday.

Exercise: the Case of the Two Classes

- The Health Club is wondering how best to market their new yoga class, and the Head of Marketing wonders if someone who goes swimming is more likely to go to a yoga class. "Maybe we could offer some sort of discount to the swimmers to get them to try our yoga."
- The CEO disagrees: "I think you're wrong. I think people who go swimming and people who go to yoga are independent. I don't think people who go swimming are more likely to do yoga than anybody else."
- They ask a group of 96 people whether they go to swimming or yoga classes. Out of this group, 32 go to yoga and 72 go swimming. 24 people go to both.
- Are the two classes dependent or independent?

Take-away

Probability Theory Summary

- Representing conditional probabilities using contingency tables and probability trees
- The chain rule
- Bayes rule
- The law of total probability (aka the partition rule)
- Independent events

For more see this course + textbook:

Math 363:
`http://math.arizona.edu/~jwatkins/math363f15.htm`