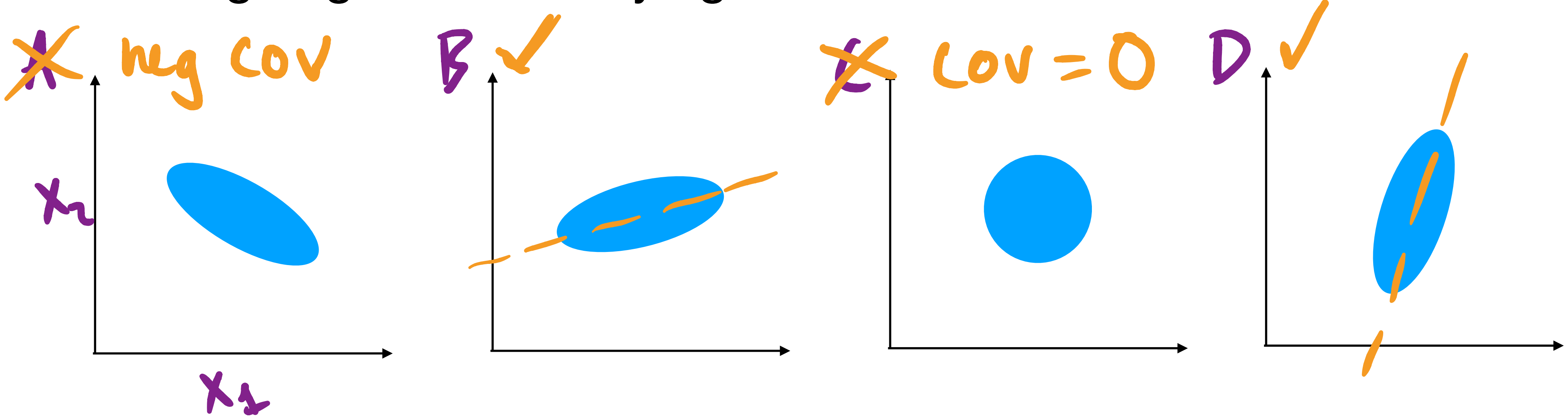
conditional probabilities, Bayes' rule

slope is pos

You are told that $cov(x_1, x_2) = 10$. What do you know about x_1 and x_2? Which of the following might be underlying distributions of the data?



Conditional Probabilities

- A **conditional probability** is a calculation of probabilities for **dependent** random variables.

- It translates to "if variable Y has value y_j, then what is the probability that variable X has value x_i?"

- dependent. r.v. tables

ICA Question 1: Conditional Probabilities

What is $P(X = 3)$?

$0.06 + 0.25 + 0.16 = 0.47$

What is $P(X = 3)$ if we already know that the value of Y is 7?

$$\frac{0.25}{0.1 + 0.15 + 0.25} = 0.5 \qquad \frac{0.25}{0.5} = 0.5$$

$$P(X = 2 \mid Y = 7) = \frac{0.15}{0.5} = 0.3$$

X = how I feel

	1	2	3
6	0.15	0.09	0.06
7	0.1	0.15	$\boxed{0.25}$
8	0.02	0.02	0.16

Y → # of hours I slept

Conditional Probability

- The calculation that we actually just did was:
- $$P(A \mid B) = \frac{P(A,B)}{P(B)} = \frac{P(A \cap B)}{P(B)} \qquad \frac{P(X=3, Y=7)}{P(Y=7)}$$

  given

- We can use this to evaluate many things!
  - What is the probability that school will close tomorrow based on if it's snowing today?
  - What is the probability that the next word in a phrase will be "turtle" given that the previous word was "a"?

Conditional Probability

- What is the probability that school will close tomorrow based on if it's snowing today?
  - What counts do we need for this?

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

P(snow)

~~P(school closed)~~

P(school closed + snow)

P(school closed | snow)

$\Downarrow$

$$\frac{P(\text{school closed} + \text{snow})}{P(\text{snow})}$$

$$P(\text{snow}) = \frac{\text{count}(\text{snow})}{\text{count}(\text{days})}$$

$$P(\text{school closed} + \text{snow}) = \frac{\text{count}(\text{sch. closed} + \text{snow})}{\text{count}(\text{days})}$$

$$P(A|B) = \frac{\dfrac{\text{count}(\text{sch. closed} + \text{snow})}{\cancel{\text{count}(\text{days})}}}{\dfrac{\text{count}(\text{snow})}{\cancel{\text{count}(\text{days})}}}$$

Conditional Probability

- What is the probability that the next word in a phrase will be "turtle" given that the previous word was "a"?

  - What counts do we need for this?

$P(\text{turtle} \mid a)$

$P(W_1 = \text{turtle} \mid W_0 = a)$

$$\frac{P(\text{a turtle})}{P(a)}$$

- count of "a turtle"
- count of "a ____"

ICA Question 2: Phrase Probabilities

Given the following data, calculate the probability that I will be late to school for each mode of transport.

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

Number of days: 50

Days that Felix was late: 20

Biked: 25

Rode the T: 20

Walked: 5

Late + bike : 5

Late + T : 13

Late + walk : 2

P(late | Bike) = 0.2 = $\dfrac{\frac{5}{\cancel{50}}}{\frac{25}{\cancel{50}}}$

P(late | T) = 0.65

P(late | walk) = 0.4

Conditional Probability

- Okay, but what if A and B are independent?

- What is the probability that I rolled a die and got a 6 given that I flipped a coin and got a tails?

- We can do the same calculation, but if $P(A \mid B) = P(A) = \dfrac{P(A, B)}{P(B)}$, then

  A and B are statistically independent

- $P(A = 6 \mid B = \text{heads})$

$$\frac{P(A = 6, B = \text{heads})}{P(B = \text{heads})} = \frac{1/6 \cdot 1/2}{1/2} = 1/6$$

ICA Question 3: P(A|B) = P(B|A)?

Given the following data, make an argument that P(A|B) = P(B|A) is or is not true. Let A be spam email and B be emails with "FREE".

$$P(\text{spam}|\text{FREE}) = \frac{10}{11}$$

$$P(\text{FREE}|\text{spam}) = \frac{10}{14}$$

$$P(A|B) \neq P(B|A)$$

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

	"FREE"	not "FREE"	
spam email	10	+ 4	14
not spam	+ 1	15	
	11		

ICA Question 4: P(A|B) = ______?

Given the following data, make an argument that P(A|B) = P(B|A) is or is not true.

P(FREE)

Calculate P(A|B) * P(B)

$$\frac{10}{\cancel{11}} * \frac{\cancel{11}}{30}$$

P(spam)

Calculate P(B|A) * P(A)

$$\frac{10}{\cancel{14}} * \frac{\cancel{14}}{30}$$

Do they have a relationship?

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

	"FREE"	not "FREE"	
spam email	10	4	14
not spam	1	15	
	11		

Bayes' rule

- Bayes' rule denotes the relationship between $P(A \mid B)$ and $P(B \mid A)$

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

- ... but, why might this be useful?

↳ what is prob. I have covid given test was negative?

↳ Naive Bayes classifier (ML)

↳ all over in NLP

ICA Question 5: Bayes rule denominator

We want to know the probability that an email is not *actually* spam given that our detection software claims that it *is* spam.

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

What should the calculation be here? (Just in terms of variables)

P(not spam | software says spam) =

(P(says spam | not spam) * P(not spam)) / P(says spam)

... and specifically for the denominator?

P(says spam) = P(says spam | spam) P(spam) + P(says spam | not spam) P(not spam)

Bayes' rule

- Bayes' rule denotes the relationship between $P(A \mid B)$ and $P(B \mid A)$
- $$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$
- When calculating $P(B)$ for the denominator, it's often useful to calculate this as the sum of $\sum_i P(B \mid A_i)P(A_i)$

all values A can take

ICA Question 5: Bayes rule denominator (cont'd)

We want to know the probability that an email is not *actually* spam given that our detection software claims that it *is* spam. P(not spam | says spam)

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

→ do not have P(says spam)

We have observed that P(not spam) = 16/30 and P(spam) = 14/30. Our software has a false positive rate of 20%. It also claims that it will flag 99.5% of all spam email as spam.

P(says spam | not spam)

P(says spam | spam)

$$\frac{.2 * \frac{16}{30}}{\left(.2 * \frac{16}{30}\right) + \left(.995 * \frac{14}{30}\right)} = 0.187$$

Naïve Bayes classifiers

- Bayes' rule denotes the relationship between $P(A \mid B)$ and $P(B \mid A)$

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

- Say that we want to know the probability that an email is spam given that it contains the word "FREE".

- Instead of calculating P(spam | FREE), we'll calculate

1) $$\frac{P(FREE \mid spam)P(spam)}{\cancel{P(FREE)} \to P(doc)}$$

2) $$\frac{P(FREE \mid not\ spam)P(not\ spam)}{\cancel{P(FREE)} \to P(doc)}$$

is #1 or #2 bigger?

ICA Question 6: Mini Naïve Bayes

yes, so this email is spam

$$\frac{10}{14} * \frac{14}{30} > \frac{1}{16} * \frac{16}{30}$$

Calculate all of the following:

P(FREE | spam) $= \frac{10}{14}$

P(FREE | not spam) $= 1/16$

P(spam) $= \frac{14}{30}$

P(not spam) $= \frac{16}{30}$

Now, make an argument for whether

P(FREE | spam)P(spam) / ~~P(document)~~

or

P(FREE | not spam)P(not spam) / ~~P(document)~~ is bigger

	"FREE"	not "FREE"	
spam email	10	4	
not spam	1	15	

Naïve Bayes classifiers

- Naïve Bayes classifiers are a little more complicated than this because we like to be able to have more than one feature
  - This is where "naïve" comes in...
- Gist is: calculate $\frac{P(class \mid features)P(class)}{P(features)}$ for each candidate class, then use the class with the **biggest** value as the overall label

Naïve Bayes classifiers

- Naïve Bayes classifiers are a little more complicated than this because we like to be able to have more than one feature
  - This is where "naïve" comes in...
- Gist is: calculate $\frac{P(class\,|\,features)P(class)}{P(features)}$ for each candidate class, then use the class with the **biggest** value as the overall label
- Neat!

Conditional Probability & natural language: wait, what?

- Say that I have the following sentences, what is P(turtle | ________) dependent on?
- "a"
- "I like my friend the"
- "I found a"

Admin

- **<u>All</u>** sections of 2810 will be dropping your lowest **<u>4</u>** ICAs.

  ↳ is in Canvas now

- Test 3 -> it's graded! Statistics look good at the moment. (Scores are higher than the first 2 tests)
  - We're working on double checking for consistency right now
  - Expect these grades before I see you next :)

Admin

- TRACE is available now!
- Please do fill these out. (in spite of survey fatigue)
  - I read them!
  - I use them to update and improve courses for the future.
- Specific feedback is helpful!
  - Something you liked? What was it?
  - Something that you'd like to see different? What was it?

Schedule

if you have HW 8 qs, I'm happy to answer them now (or mini-projects)

Turn in **ICA 21** on Canvas (make sure that this is submitted by 2pm!) - passcode is "dragon"

HW 8: due on Sunday @ 11:59pm

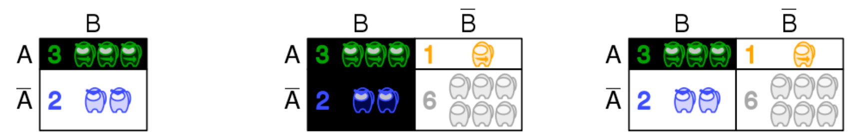
Test 4: May 4th, 1 - 3pm, Snell Engineering 108

Mon	Tue	Wed	Thu	Fri	Sat	Sun
April 11th Lecture 21 - conditional probabilities, bayes	**Felix OH Calendly**	**Felix OH Calendly**	**Felix OH Calendly** Lecture 22 - conditional independence, bayes nets			**HW 8 due @ 11:59pm**
April 18th **No lecture - Patriot's Day** (and no videos)	**Felix OH Calendly**	**Felix OH Calendly**	**Felix OH Calendly** Lecture 23 - Regression: R^2 & F			
April 25th Lecture 24 - presentations, wrap-up **Mini-project due @ 11:45am**		**HW 9 due @ 11:59pm**				

More recommended resources on these topics

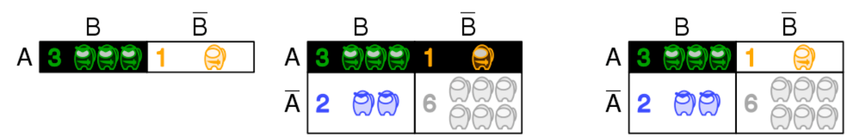
- Bayes theorem w/ Among Us characters: https://en.wikipedia.org/wiki/Conditional_probability#/media/File:Bayes_theorem_assassin.svg
    - (copied onto next slide)
- YouTube: 3Blue1Brown, Bayes theorem, the geometry of changing beliefs

Number of occurences		Being suspicious B	Not being suspicious $\overline{B}$	sum
An assassin	A	3	1	4
Not an assassin	$\overline{A}$	2	6	8
sum		5	7	12



$$P(A\text{, given }B) \cdot P(B) = P(A|B) \cdot P(B)$$

$$\frac{3}{\cancel{3+2}} \cdot \frac{\cancel{3+2}}{3+1+2+6} = \frac{3}{3+1+2+6}$$



$$P(B\text{, given }A) \cdot P(A) = P(B|A) \cdot P(A)$$

$$\frac{3}{\cancel{3+1}} \cdot \frac{\cancel{3+1}}{3+1+2+6} = \frac{3}{3+1+2+6}$$

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

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

https://en.wikipedia.org/wiki/Conditional_probability#/media/File:Bayes_theorem_assassin.svg