CSC380: Principles of Data Science

Probability 2

Xinchen Yu

- If we know that all outcomes are equally likely, we can use

We will use combinatorics to do counting

Number of elements in event set

$$P(E) = \frac{|E|}{|S|}$$

Number of possible outcomes (e.g. 36)

- If |E| is hard to calculate directly, we can try

  - the rules of probability

  - the Law of Total Probability, using an appropriate partition of sample space S

Rules of probability

- To recap and summarize:

Rules of Probability

1. **Non-negativity:** All probabilities are between 0 and 1 (inclusive)

2. **Unity of the sample space:** $P(S) = 1$

3. **Complement Rule:** $P(E^C) = 1 - P(E)$

4. **Probability of Unions:**
   - (a) *In general,* $P(E \cup F) = P(E) + P(F) - P(E \cap F)$
   - (b) *If E and F are disjoint, then* $P(E \cup F) = P(E) + P(F)$

- Conditional probability
- Probabilistic reasoning
  - contingency table
  - probability trees
- Bayes rule
- Independence of events
- Probability and combinatorics

Conditional Probability

Example: Seat Belts

		Child		
		Buck.	Unbuck.	Marginal
Parent	Buck.	0.48	0.12	0.60
	Unbuck.	0.10	0.30	0.40
	Marginal	0.58	0.42	1.00

Table: Probability Estimates for Seat Belt Status

Suppose we pick a family from US at random:

- What is the probability of the event "Child is Buckled"?
- What should our new estimate be if we know that "Parent is Buckled"?

Example: blood types

		Antigen B		
		Absent	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

- A: "presence of antigen A", and B: "presence of antigen B"
- Suppose someone of an unknown blood type gets a test that reveals the presence of antigen A.
    - What is the chance that event A happens to them now?
    - What is the chance that event B happens to them now?

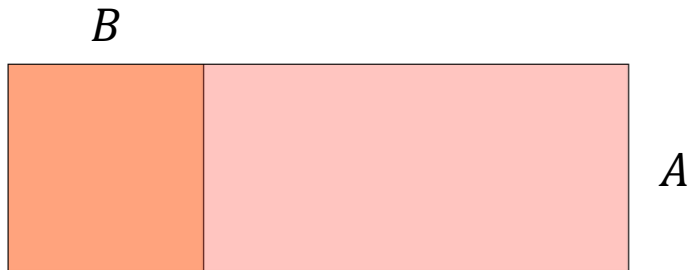# Relative area

- A: antigen A present B: antigen B present
- Given that A happens, what is the chance of B happening?
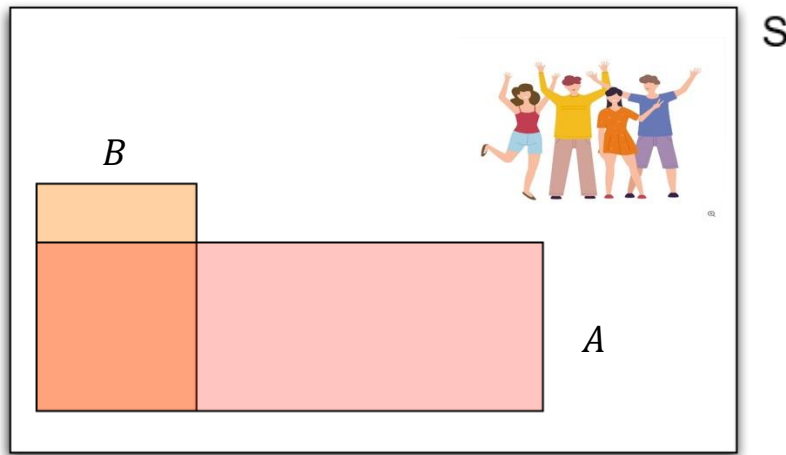
- Another way to think about this:
  - Restricted to people with antigen A present, what is the fraction of those people with antigen B?

Relative area

- Let's zoom into people with antigen A present.

B

A

- It's just as if the sample space had shrunk to include only A
- Now, probabilities correspond to proportions of A
- What does the orange square represent in the original sample space?
  - $A \cap B$
- How would we find the probability of B given A?

Conditioning changes the sample space

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

- After we know A happened, we are only choosing from within A. e.g., A: even numbers, B: get a 2

- The set A becomes our new sample space

- Instead of asking "In what proportion of S is B true?", we now ask "In what proportion of A is B true?" e.g., 1/6 vs 1/3

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

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)}$$
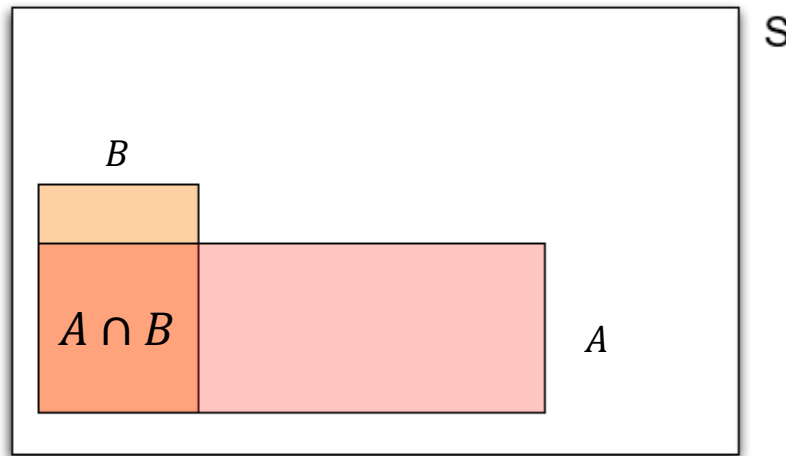
Every Probability is a Conditional Probability

- We can consider the original probabilities to be conditioned on the event S: at first what we know is that "something in S" occurs. E.g.

$$P(B) = P(B|S)$$
$$P(B \cap C) = P(B \cap C|S)$$

- $P(B|S)$ in words: what proportion of S does B happen?

- If we then learn that A occurs, A becomes our restricted sample space. $P(B|A)$ in words: what proportion of A does B happen?

Joint Probability and Conditional Probability

- We can rearrange $P(B \mid A) = \frac{P(A \cap B)}{P(A)}$ and derive:

The "Chain Rule" of Probability

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

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

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

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

When we have two events A and B…

- Conditional probability: $P(A|B), P(A^c|B), P(B|A)$ etc.

- Joint probability: $P(A, B)$ or $P(A^c, B)$ or …

- Marginal probability: $P(A)$ or $P(A^c)$

Example revisited: blood types

		Antigen B		
		Absent	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

- Suppose someone of an unknown blood type gets a test that reveals the presence of antigen A.
  - What is $P(A \mid A)$?

$$P(A \mid A) = \frac{P(A \cap A)}{P(A)} = 1$$

  - What is $P(B \mid A)$?

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{0.04}{0.46} = 0.087$$

Example revisited: Seat Belts

A: parent is buckled
C: child is buckled

		Child		
		Buck.	Unbuck.	Marginal
Parent	Buck.	0.48	0.12	0.60
	Unbuck.	0.10	0.30	0.40
	Marginal	0.58	0.42	1.00

Table: Probability Estimates for Seat Belt Status

Suppose we pick a family from US at random:

- What is the probability of the event "Child is Buckled"? $P(C)$

- What should our new estimate be if we know that ("given that") Parent is Buckled? $P(C \mid A)$

Example revisited: Seat Belts

A: parent is buckled
C: child is buckled

		Child		
		Buck.	Unbuck.	Marginal
Parent	Buck.	0.48	0.12	0.60
	Unbuck.	0.10	0.30	0.40
	Marginal	0.58	0.42	1.00

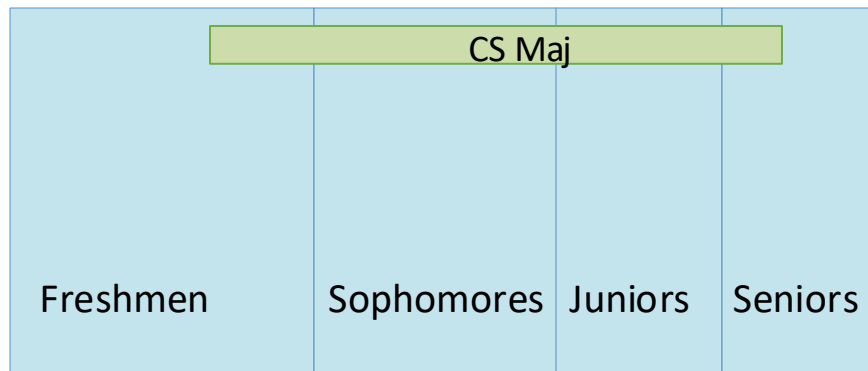Table: Probability Estimates for Seat Belt Status

Suppose we pick a family from the US at random:

- $P(C) = 0.58$

- $P(C \mid A) = \dfrac{P(C \cap A)}{P(A)} = \dfrac{0.48}{0.60} = 0.8$ Larger than $P(C)$

- Suppose we see a buckled parent, it is much more likely that we see their child buckled

Law of Total Probability Suppose $B_1, \ldots, B_n$ *form a partition* of the sample space S. Then,

$$P(A) = P(A, B_1) + \cdots + P(A, B_n)$$

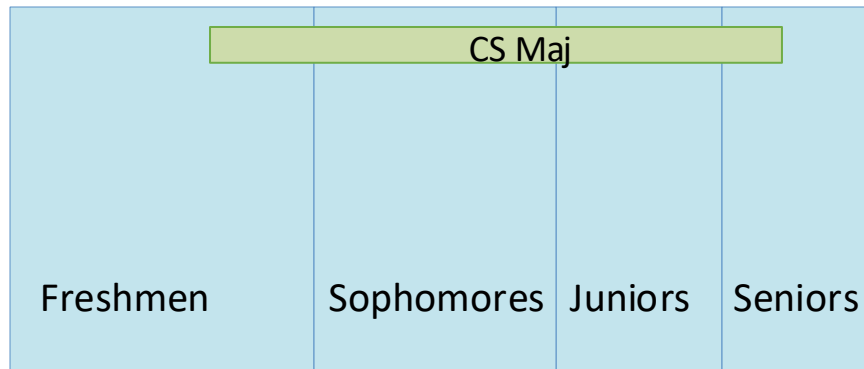|			
	CS Maj		
Freshmen	Sophomores	Juniors	Seniors

Law of Total Probability, revisited

Expanding each $P(A, B_i) = P(A \mid B_i)P(B_i)$, we have:

$$P(A) = \sum_{i=1}^{n} P(A \mid B_i)P(B_i)$$

A: student in CS major

B_i : student in class year i

$P(A \mid B_i)$ The fraction of CS major in class year i

Law of Total Probability, revisited

Example Suppose UA has an equal number of students in the 4 class years, and the fraction of CS major in these 4 class years are 10%, 10%, 20%, 80% respectively. What is fraction of CS majors?

- $P(B_1) = P(B_2) = P(B_3) = P(B_4) = 0.25$
- $P(\,C \mid B_1\,) = 0.1, \ldots, P(\,C \mid B_4\,) = 0.8$
- Calculate $P(C)$ by:

$$P(C) = \sum_{i=1}^{4} P(\,C \mid B_i\,)P(B_i) = 30\%$$

Probabilistic reasoning

Probabilistic reasoning

- We have some prior belief of an event A happening
  - $P(A)$, prior probability
  - e.g. me infected by COVID
- We see some new evidence B
  - e.g. I test COVID positive

- How does seeing B affect our belief about A?
  - $P(A \mid B)$, posterior probability

Another example: detector

A store owner discovers that some of her employees have taken cash. She decides to use a detector to discover who they are.

- Suppose that 10% of employees stole.
- The detector buzzes 80% of the time that someone stole, and 20% of the time that someone not stole
- Is the detector reliable? In other words, if the detector buzzes, what's the probability that the person did stole?

H: employee not stole

B: lie detector buzzes

Another example: detector

- Suppose that 10% of employees stole.

    H: employee did not stole $\quad P(H) = 0.9$

- The detector buzzes 80% of the time that someone stoles, and 20% of the time that someone not stole.

    $$P(B \mid H^C) = 0.8$$

    B: lie detector buzzes

    $$P(B \mid H) = 0.2$$

- If the detector buzzes, what's the probability that the person stole?

    $$P(H^C \mid B)$$

Detector analysis: Probability table

		Detector result		
		Pass	Buzz (B)	Marginal
Employee	Not stole (H)			
	Stole			
	Marginal			

$P(H) = 0.9$

$P(B \mid H^C) = 0.8$

$P(B \mid H) = 0.2$

Detector analysis: Probability table

$$P(H, B) = P(H) \cdot P(B \mid H) = 0.9 \times 0.2 = 0.18$$

		Detector result		
		Pass (B^C)	Buzz (B)	Marginal
Employee	Not stole (H)		0.18	0.9
	Stole (H^C)			0.1
	Marginal			

$$P(H) = 0.9$$

$$P(B \mid H^C) = 0.8$$

$$P(B \mid H) = 0.2$$

Detector analysis: Probability table

$$P(H) = P(H, B) + P(H, B^c) = 0.9$$

		Detector result		
		Pass (B^c)	Buzz (B)	Marginal
Employee	Not stole (H)	0.72	0.18	0.9
	Stole (H^c)			0.1
	Marginal			

$$P(H) = 0.9$$

$$P(B \mid H^c) = 0.8$$

$$P(B \mid H) = 0.2$$

Detector analysis: Probability table

		Detector result		
		Pass (B^C)	Buzz (B)	Marginal
Employee	Not stole (H)	0.72	0.18	0.9
	Stole (H^C)	0.02	0.08	0.1
	Marginal	0.74	0.26	1

$P(H) = 0.9$

$P(B \mid H^C) = 0.8$

$P(B \mid H) = 0.2$

Detector analysis: Probability table

		Detector result		
		Pass (B^C)	Buzz (B)	Marginal
Employee	Not stole (H)	0.72	0.18	0.9
	Stole (H^C)	0.02	0.08	0.1
	Marginal	0.74	0.26	1

- We have the full probability table. Can we calculate $P(H^C \mid B)$? Yes!

$$P(H^C \mid B) = \frac{P(H^C, B)}{P(B)} \qquad \frac{0.08}{0.26} = 0.307$$

It seems like the detector is not very reliable…

Recap

- Conditional probability: $P(B \mid A) = \frac{P(A \cap B)}{P(A)}$

- Law of total probability: $P(A) = \sum_{i=1}^{n} P(A \mid B_i) P(B_i)$

- If we know $\mathrm{P}(H), P(B|H^C), P(B|H)$:
  - $\mathrm{P}(H) \rightarrow P(H^C)$ Complement rule
  - $\mathrm{P}(H), P(B|H) \rightarrow P(B, H)$ Conditional probability
  - $\mathrm{P}(H^C), P(B|H^C) \rightarrow P(B, H^C)$ Conditional probability
  - $\mathrm{P(B)} \rightarrow P(B, H) + P(B, H^C)$ Law of total probability
  - $P(B), P(B, H) \rightarrow P(H|B)$ Conditional probability
  - $P(B), P(B, H^C) \rightarrow P(H^C|B)$ Conditional probability

- We can get $P(B), P(H|B), P(H^C|B)$

Today's plan

- Another tool: probability trees
- Bayes rule
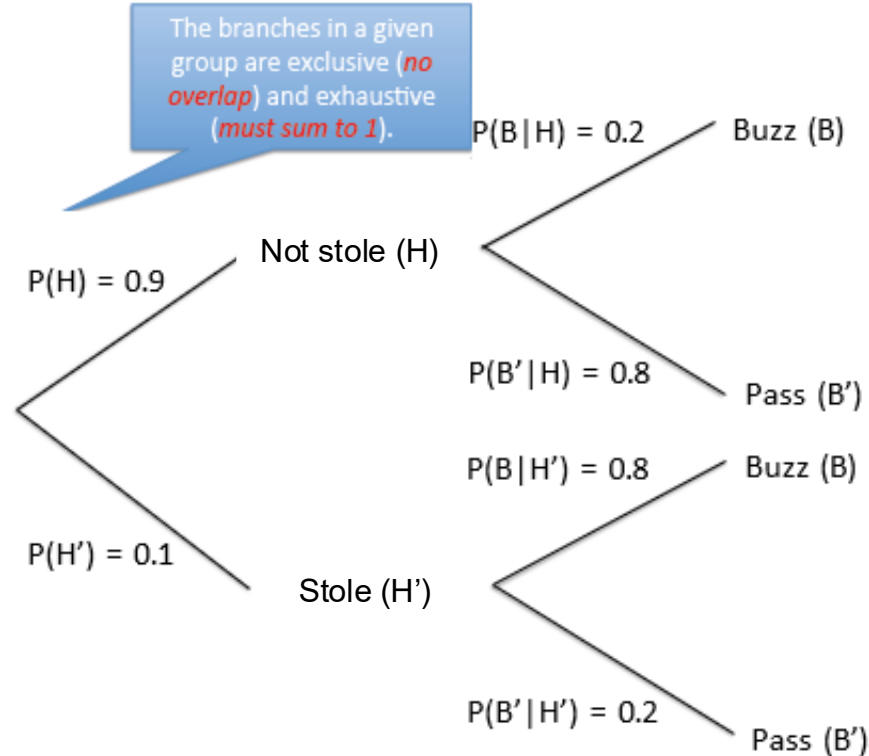- Bayes rule and law of total probability
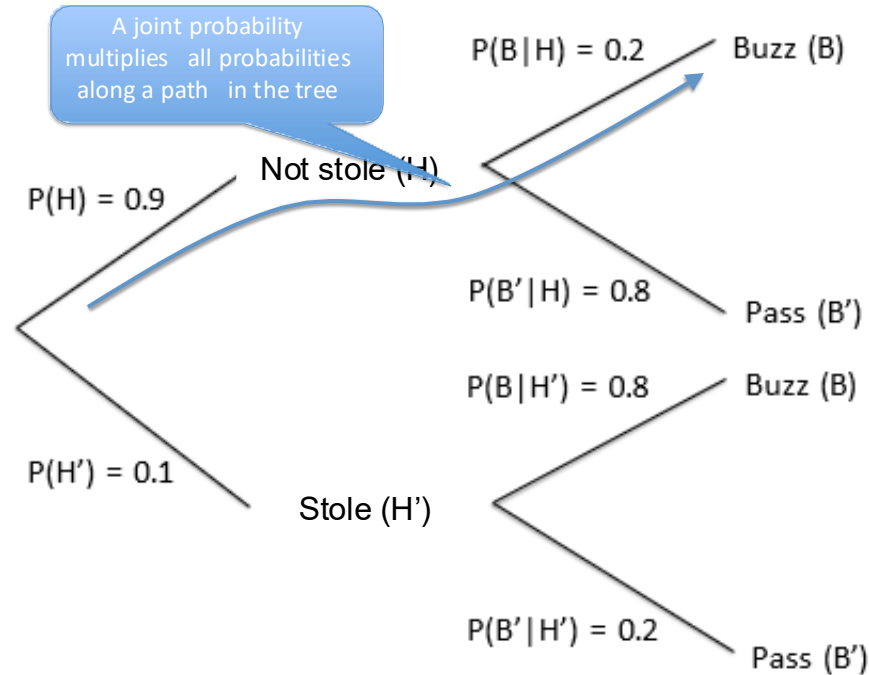- Independence

Probability trees: another useful tool
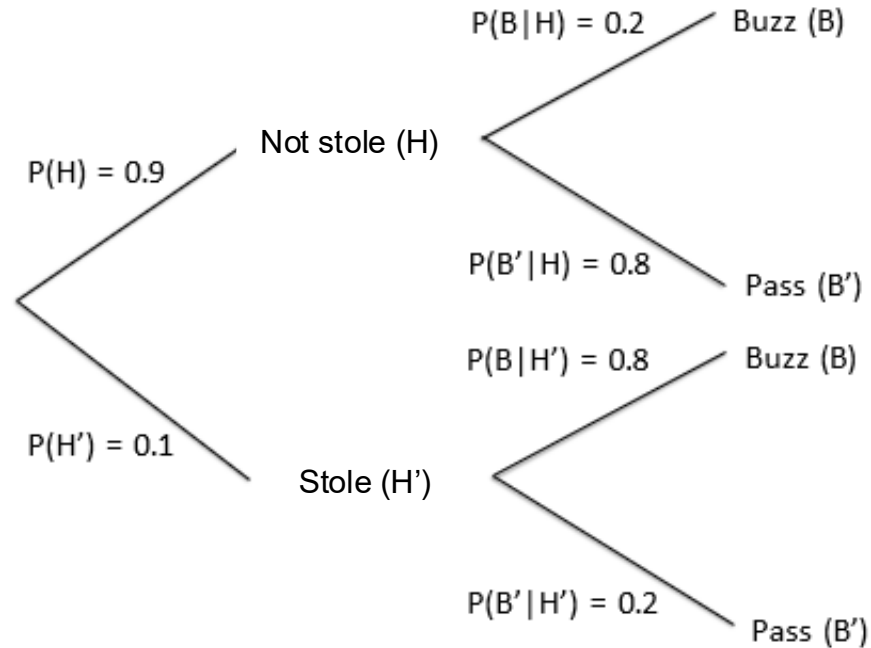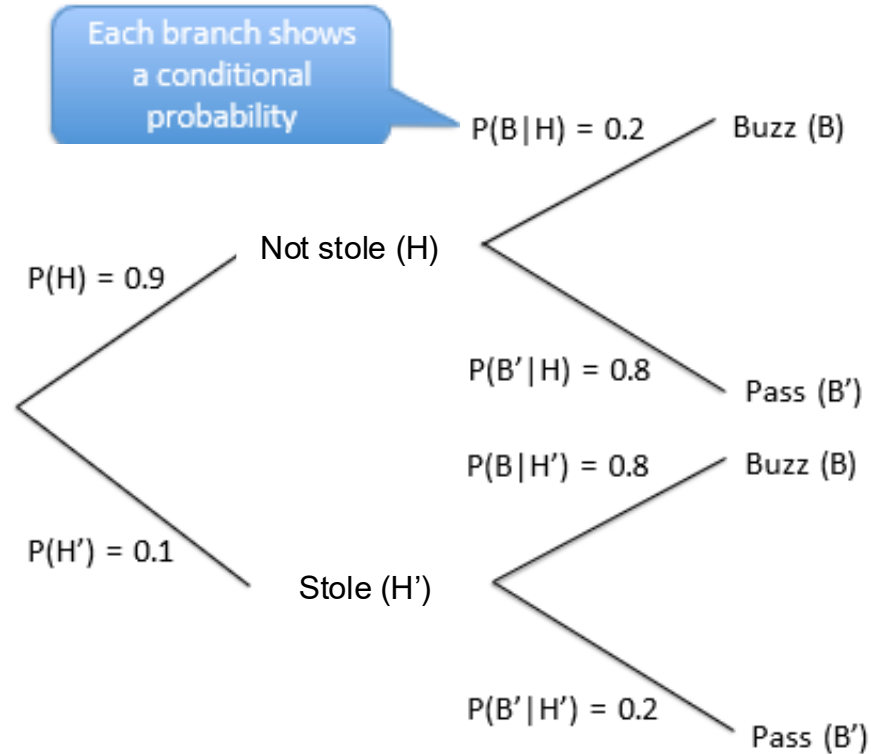
Probability trees: another useful tool

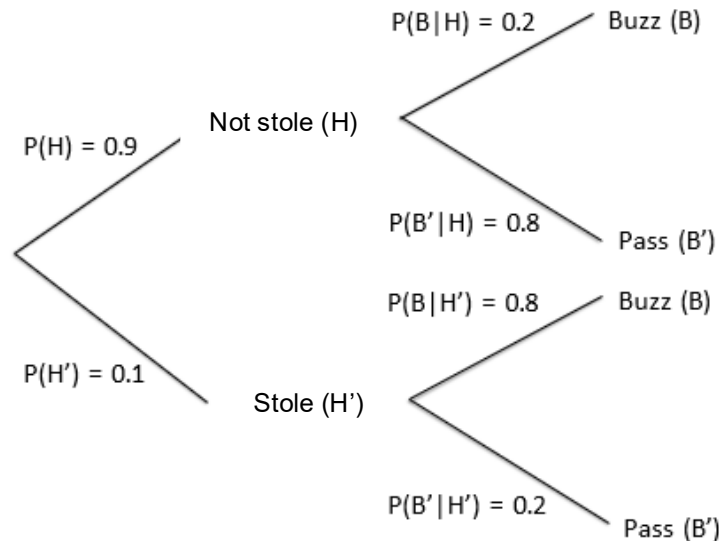# Probability trees: another useful tool

Probability trees: another useful tool

Probability trees: another useful tool

- What is $P(\text{Buzz}, \text{Stole})$?
  - 0.08

- $P(\text{Buzz})$?
  - Hint: which branches end up with buzzing?
  - 0.26 (0.08+0.18)

- $P(\text{Stole} \mid \text{Buzz})$?
  - Hint: which of the prev. branches contains the stole event?
  - 0.08 / 0.26

The Public Health Department gives us the following information:

- A test for the disease yields a positive result (+) 90% of the time when the disease is present (Y) $P(+ \mid Y) = 0.9$, "sensitivity" of the test

- A test for the disease yields a positive result 1% of the time when the disease is not present (N)
$P(+ \mid N) = 0.01$

- One person in 1,000 has the disease. $P(Y) = 0.1\%$

Draw a probability tree and use it to answer: what is the probability that a person with positive test has the disease?
$P(Y \mid +)?$

In-class activity: COVID test

- Goal: calculate P(Y | +)
- Two branches are associated

with positive test results +

- What are the associated events?
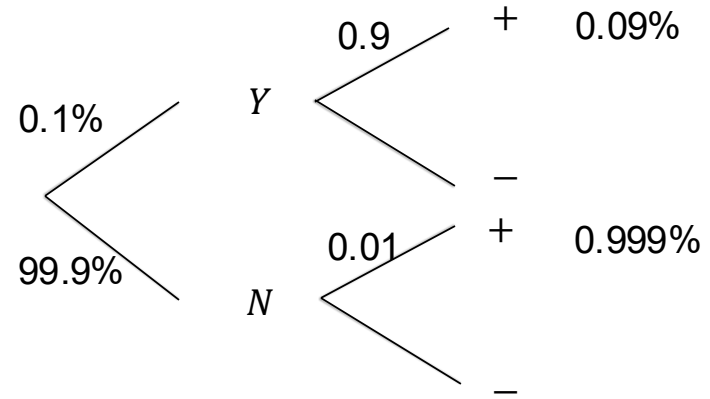
$$P(+, Y) = P(+ \mid Y)P(Y) = 0.09\%$$

$$P(+, N) = P(- \mid N)P(N) = 0.999\%$$

$$P(Y \mid +) = \frac{P(+, Y)}{P(+)} = \frac{0.09\%}{0.09\% + 0.999\%} \approx \frac{1}{12}$$

- Conclusion: being tested positive does not mean much..

P(+ | Y) = 0.9
P(+ | N) = 0.01
P(Y) = 0.001

0.1%
0.9   +   0.09%
Y
−
99.9%
0.01   +   0.999%
N
−

In-class activity: COVID test

Probabilistic reasoning tells as how does seeing a new evidence affect our prior belief about an event.

- Prior probability: one person in 1,000 has the disease: P(Y) = 0.1%

- New evidence: seen a person is tested positive

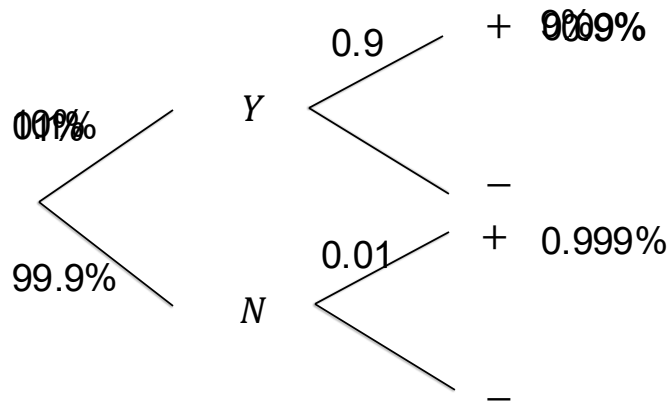- Posterior probability: a person with positive test has the disease:

$$P(Y \mid +) = \frac{0.09\%}{0.09\% + 0.999\%} \approx \frac{1}{12}$$

COVID test: additional insights

- What would P(Y | +) look like, if instead:
  - 1 in 100 people have COVID?
  - 1 in 10?

$$P(Y \mid +)= \frac{P(+,Y)}{P(+)} = \frac{0.09\%}{0.09\%+0.999\%} \approx \frac{1}{12}$$

0.9   +   0.09%

Y

0.1%

99.9%

N

0.01   +   0.999%

−

−

Note: this branch's value is imprecise, but roughly stays close to 1%!

- Insight: *base rate $P(Y)$* significantly affects P(Y | +), hence the conclusions we draw

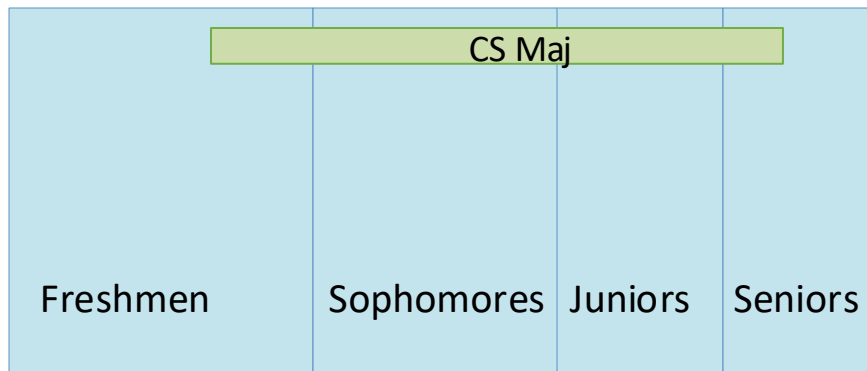# Conditional probability: additional note

- The rules of probability also applies to the rules of conditional probability
- Just replace $P(E), P(F)$ with $P(E|A), P(F|A)$
  - But, need to condition on the same A in the same equation

Rules of Probability

1. **Non-negativity:** All probabilities are between 0 and 1 (inclusive)

2. **Unity of the sample space:** $P(S) = 1$

3. **Complement Rule:** $P(E^C) = 1 - P(E)$

4. **Probability of Unions:**
   (a) *In general,* $P(E \cup F) = P(E) + P(F) - P(E \cap F)$
   (b) *If E and F are disjoint, then* $P(E \cup F) = P(E) + P(F)$

Some examples

- $P(S|A) = 1$ A: CS major
- $P(E|A) + P(E^C|A) = 1$
- $P(E|A) + P(F|A) = P(E \cup F|A)$ for disjoint E and F

Bayes rule

Reversing conditional probabilities

- Is $P(A \mid B) = P(B \mid A)$ in general?

- Let's see..

$$P(A, B) = P(A \mid B) \cdot P(B) = P(B \mid A) \cdot P(A)$$

- Equal only when $P(A)$ and $P(B)$ are equal

- Let's take a look at a real-world example when they are unequal…

Reversing conditional probabilities

Q: Hearing a French accent means someone is French?

Event A: A person is from France.

Event B: A person speaks English with a French accent.

- In a diverse city, only 5% of people are from France
- Of those from France, 80% speak English with a French accent: $P(B|A)$
- Of those not from France, only 2% speak English with a French accent (due to schooling, mimicry, or neighboring countries)

What is $P(A), P(B)$ and $P(A|B)$?

Reversing conditional probabilities

What is $P(A), P(B)$ and $P(B|A)$?

- $P(A) = 0.05$
- $P(B) = P(A, B) + P(A^c, B) = P(B|A) \cdot P(A) + P(B|A^c) \cdot P(A^c) = 0.8 \cdot 0.05 + 0.02 \cdot (1 - 0.05) = 0.04 + 0.019 = 0.059$
- $P(A \mid B) = P(A, B)/P(B) = 0.04/0.059 \approx 0.678$

So $P(A) \neq \mathrm{P}(B)$, also hearing a French accent doesn't guarantee someone is French: a ~68% chance

Bayes rule

Bayes rule For events A, B,

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

- Very easy to derive from the chain rule, so remember that first.
- Named after Thomas Bayes (1701-1761), English philosopher & pastor

Bayes rule

Bayes rule For events A, B,

Prior probability   Support of evidence

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

Posterior probability   Probability of evidence

Examples:

- A: I have COVID, B: my test shows positive
- A: employee stole B: the detector buzzes
- A: student is CS major B: student is a senior

Bayes rule: another example

- In a class, 16% of the students are Nutrition Science majors, 55% students are female. Of the Nutrition Science majors, 54% are female.

- What proportion of female students in the class are Nutrition Science majors?

- What is the probability tree of this?

- We are looking for $P(N \mid F)$
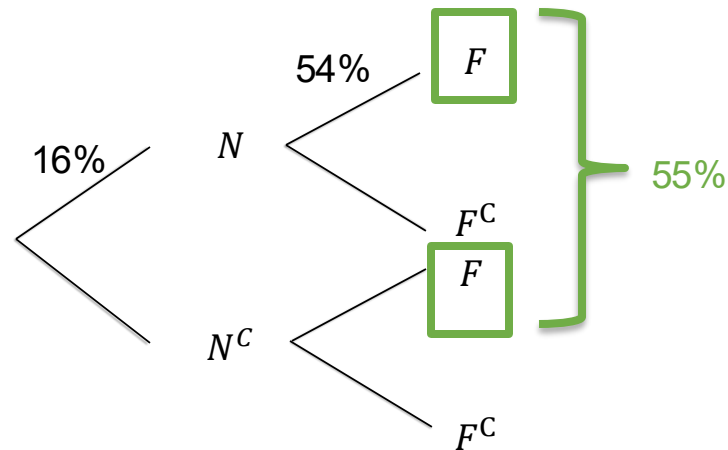
Bayes rule: another example

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

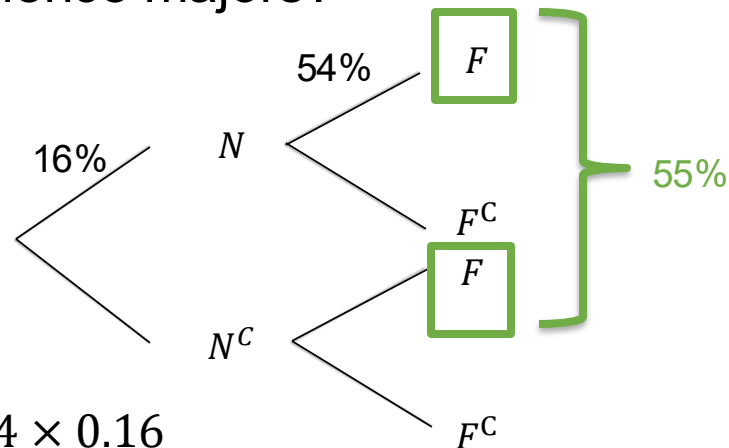- We can use $P(N \mid F) = \frac{P(N,F)}{P(F)}$
  - We know $P(F) = 0.55$

- Can we obtain $P(N,F)$?
  - We can use $P(N,F) = P(F \mid N) \cdot P(N) = 0.54 \times 0.16$
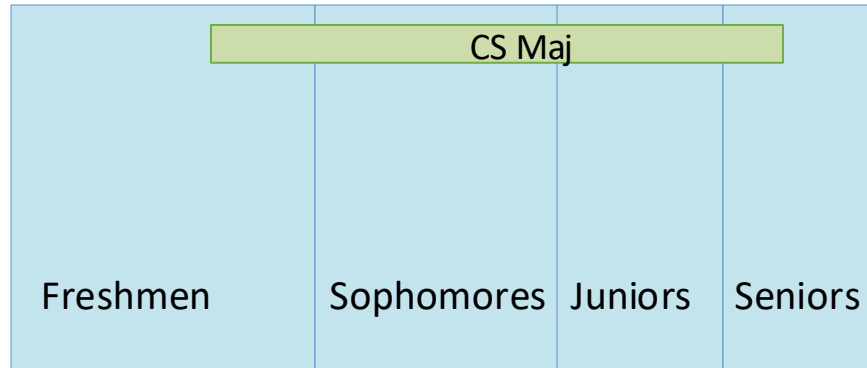
- Altogether, we have
$$P(N \mid F) = \frac{P(F \mid N) \cdot P(N)}{P(F)} = \frac{0.54 \times 0.16}{0.55}$$

Bayes rule and Law of Total Probability

Bayes rule (equivalent form) For event A and $B_1, \ldots, B_n$ forming a partition of S,

$$P(B_i \mid A) = \frac{P(A \mid B_i) \cdot P(B_i)}{\sum_{j=1}^{n} P(A \mid B_j) \cdot P(B_j)} \quad \longleftarrow \quad P(A)$$

CS Maj

Freshmen   Sophomores   Juniors   Seniors

....

Example Suppose UA has an equal number of students in the 4 class years, and the fraction of CS major in these 4 class years are 10% ($P(C|B_1)$), 10%, 20%, 80% respectively.

We have previously calculated that $P(C) = 30\%$

If we see a CS major student, what is their most likely year class?

$$P(B_1 \mid C), \ldots, P(B_4 \mid C) \text{ -> maximum?}$$

Bayes rule and Law of Total Probability

- Let's draw a probability tree..
- After learning that the student is CS major:

$$P(B_1 \mid C) = \frac{0.25 \times 0.1}{P(C)}$$

...

$$P(B_4 \mid C) = \frac{0.25 \times 0.8}{P(C)}$$

- So most likely, this student is a senior
- Equivalent form: $P(B_i \mid C) \propto P(B_i)P(C \mid B_i)$
  - $\propto$: proportional to
  - $P(C)$ can be viewed as a *normalization factor*