Computer Science

CSC380: Principles of Data Science

Probability 2

Chicheng Zhang

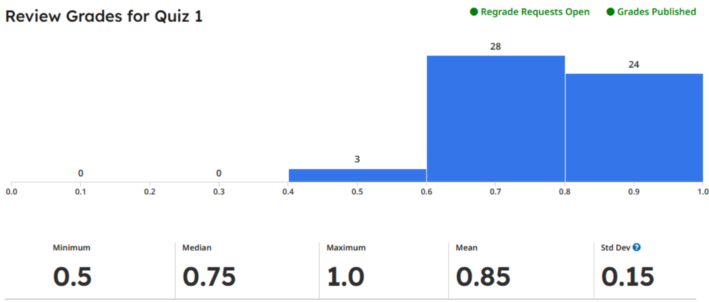
Announcements 1/29

- We will have a quiz next Monday (2/3)
  - Based on this week's lectures
- Quiz 1: grading in progress
  - Groups of 2 may only see one member's grade in gradescope
  - Don't worry, we will update on D2L
- HW1 due 1/30 3:30pm
  - Submit code in "HW1 code" entry
  - One submission per group

  (I enabled group submission of max size 2 in gradescope)
- HW2 will be out around 1/31 (due 2/10)

Announcements 2/3

- We have synced Quiz 1 grades on D2L
  - Let us know if you have any questions



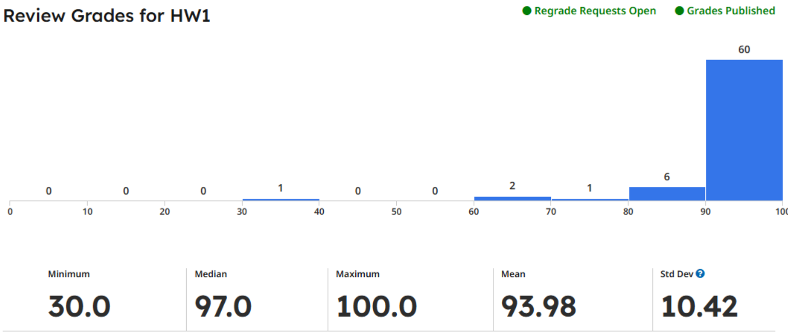
- HW2 due 2/11
- Simulating n 2-dice rolls:

```
res_dice1 = np.random.randint(1,6+1,size=n)
res_dice2 = np.random.randint(1,6+1,size=n)
res = [(res_dice1[i], res_dice2[i]) for i in range(len(res_dice1))]
```

- Can I make future HW announcements on Piazza?

Announcements 2/5

- We graded HW1 & posted in D2L



- Add'l HW policies:
  - Only one submission per group
  - All group members are expected to contribute equally for every qn
- For all regrade requests & D2L grade errors:
  - we need your reports within a week of the grade announcement times
- Quiz 3: 2/12 (Wed)

Quiz 2/3

- You conduct an experiment for which *one* run consists of flipping a coin three times. Define E_1 to be the event that the *first* flip is heads, and E_2 to be the event that the *second* flip is heads.
  1. Write down all outcomes in the full sample space, *S*.
     (e.g. we can represent the outcome of 3 heads as HHH)
  2. Write down all outcomes in E_1.
  3. Write down all outcomes in $E_1 \cup E_2$.
  4. Write down all outcomes in $E_1 \cap E_2$.
  5. Write down all outcomes in E_1^C .

Quiz 2/3

- You conduct an experiment for which *one* run consists of flipping a coin three times. Define E_1 to be the event that the *first* flip is heads, and E_2 to be the event that the *second* flip is heads.

  Write down all outcomes in the full sample space, *S*.

  {HHH, HHT, HTH, …, TTT}

  Write down all outcomes in E_1.

  {HHH, HHT, HTH, HTT}

  Write down all outcomes in $E_1 \cup E_2$.

  $E_2 = \{HHH, HHT, THH, THT\}$

  Write down all outcomes in $E_1 \cap E_2$.

  {HHH, HHT}

  Write down all outcomes in E_1^C .

  {THH THT, TTH, TTT}

Summary: calculating probabilities

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

We will use combinatorics to do counting

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

Number of elements in event set ← $|E|$

Number of possible outcomes (e.g. 36) ← $|S|$

- If |E| is hard to calculate directly, we can try using the rules of probability

- If this is still challenging, we can try using 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 $P(E \cup F) = P(E) + P(F)$

Overview

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

Conditional Probability

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

- 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?
  - Chance that event B happens to them now?

Example: Seat Belts

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

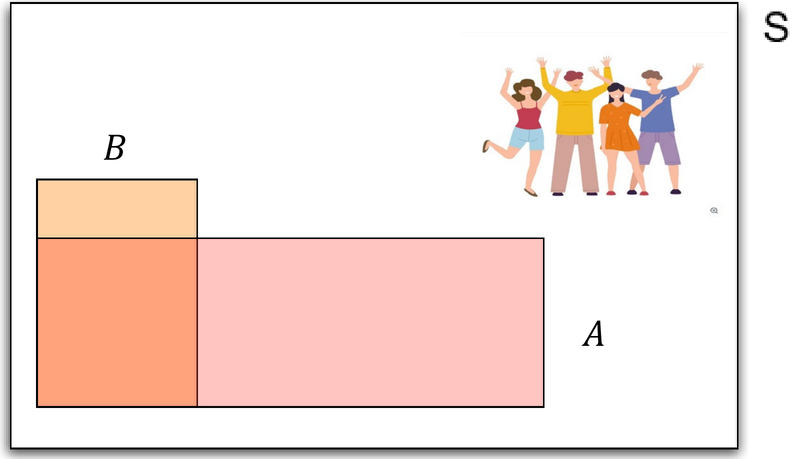
Table: Probability Estimates for Seat Belt Status

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 ("given that") Parent is Buckled?

Relative area

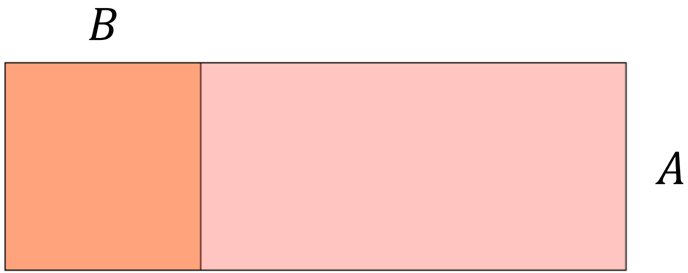
- A: antigen A present B: antigen B present
- Given that A happens, what is the chance of B happening?



- 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..



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

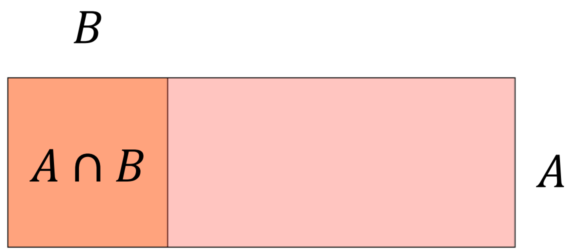
Conditioning changes the sample space

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

Conditional Probability

- 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

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

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

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

Terminology

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	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

- 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.	Child: 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

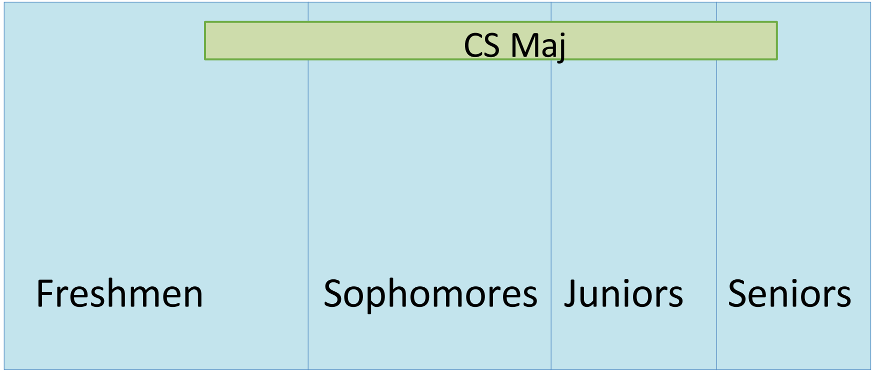
Suppose we pick a family from the US at random:

- $P(C) = 0.58$
- $P(C \mid A) = \frac{P(C \cap A)}{P(A)} = \frac{0.48}{0.60} = 0.8$ 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, revisited

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

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



Law of Total Probability, revisited

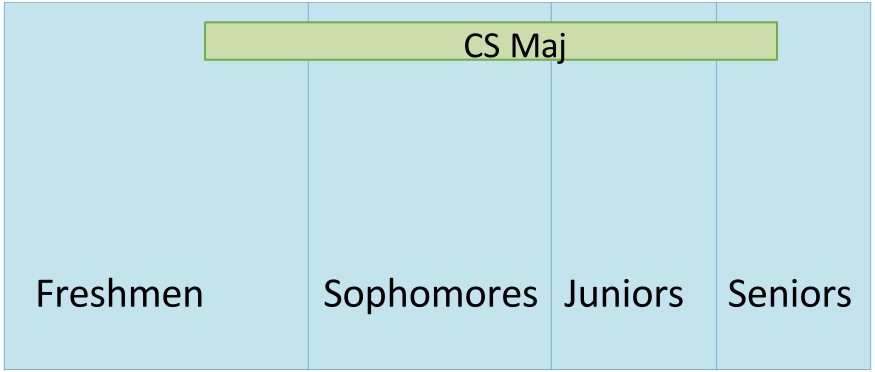
Expanding each $P(A, B_i) = P(A \mid B_i)P(B_i)$, we have:

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

A: student in CS major

B_i : student in class year i

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?

Soln

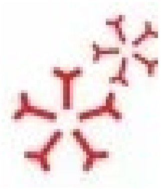
- $P(B_1) = P(B_2) = P(B_3) = P(B_4) = 0.25$
- $P(C \mid B_1) = 0.1, \ldots, P(C \mid B_4) = 0.8$
- We can now calculate $P(C)$ by:

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

Probabilistic reasoning

Probabilistic reasoning

- We have some prior belief of an event A happening
  - $P(A)$, prior probability
  - E.g. me infected by COVID



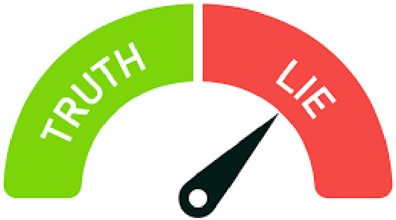
- We see some new evidence B
  - E.g. I test COVID positive

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

Another example: 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?



Another example: lie detector

- Suppose that 10% of employees stole, but 100% say they didn't.

H: employee is honest $P(H) = 0.9$

- The lie detector buzzes 80% of the time that someone lies, and 20% of the time that someone is telling the truth.

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

B: lie detector buzzes

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

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

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

Lie detector analysis: Probability table

- Aka contingency table, two-way table

		Lie Detector Result		
		Pass	Buzz!!	Marginal
Employee	Honest	??		$P(H) = 0.9$
	Dishonest!!			
	Marginal			

Table: Lie Detector Probabilities

- $P(H) = 0.9$
- $P(B \mid H^C) = 0.8$, $P(B \mid H) = 0.2$
- If we can compute all entries of the probability table, we can get $P(H^C \mid B)$

Lie detector analysis: Probability table

		Lie Detector Result		
		Pass	Buzz!!	Marginal
Employee	Honest			$P(H) = 0.9$
	Dishonest!!			$P(H^C) = 0.1$
	Marginal			

Table: Lie Detector Probabilities

- $P(H) = 0.9$
- $P(B \mid H^C) = 0.8$, $P(B \mid H) = 0.2$
- Let's try to fill in the table..

Lie detector analysis: Probability table

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

		Lie Detector Result		
		Pass	Buzz!!	Marginal
Employee	Honest	0.72	0.18	$P(H) = 0.9$
	Dishonest!!	0.02	0.08	$P(H^C) = 0.1$
	Marginal	0.74	0.26	1

Table: Lie Detector Probabilities

- $P(H) = 0.9$
- $P(B \mid H^C) = 0.8$, $P(B \mid H) = 0.2$
- Let's try to fill in the table..

Lie detector analysis: Probability table

		Lie Detector Result Pass	Buzz!!	Marginal
Employee	Honest	0.72	0.18	$P(H) = 0.9$
	Dishonest!!	0.02	0.08	$P(H^C) = 0.1$
	Marginal	0.74	0.26	1

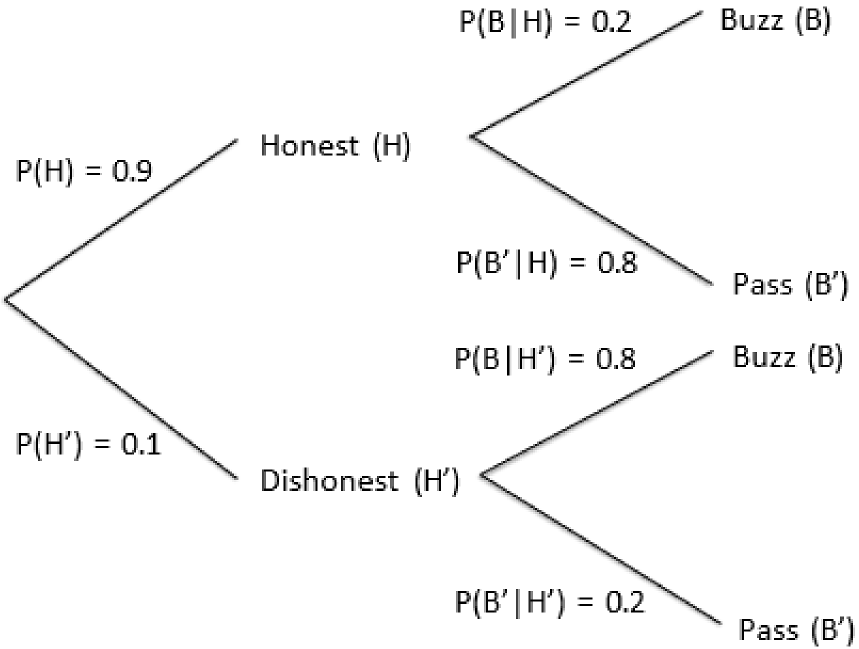
Table: Lie Detector Probabilities

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

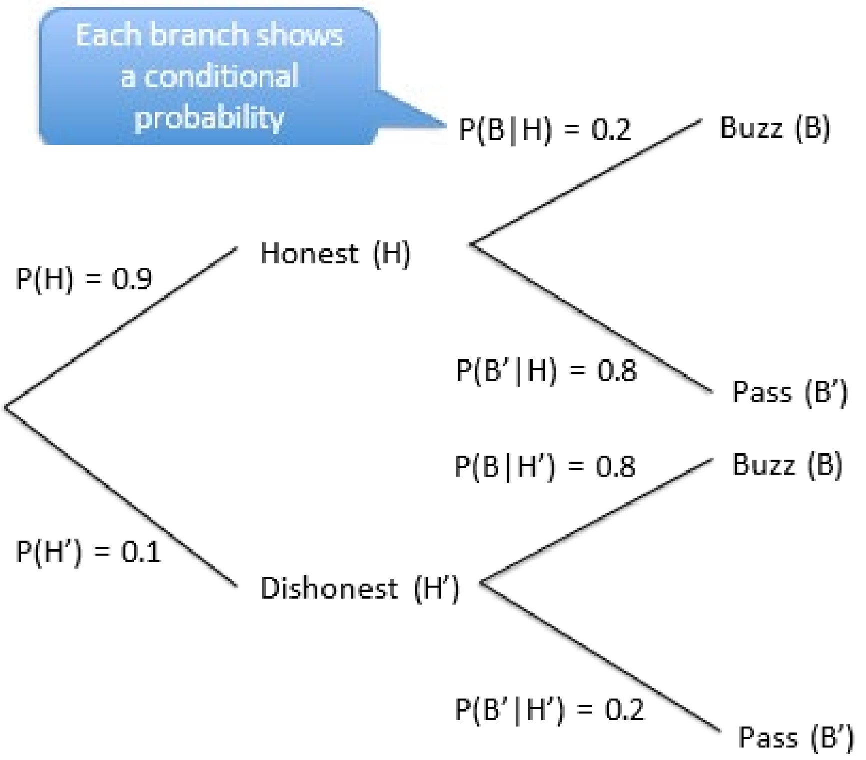
$$P(H^C \mid B) = \frac{P(H^C, B)}{P(B)} \qquad \frac{0.08}{0.26} = 0.307$$

- It seems like the lie detector is not very reliable..

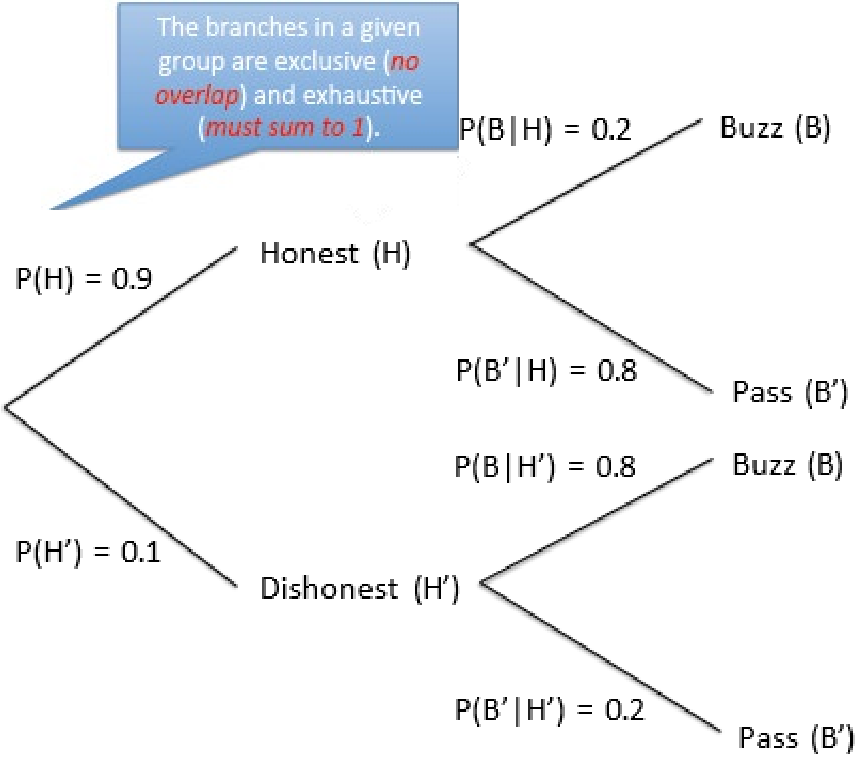
Probability trees: another useful tool



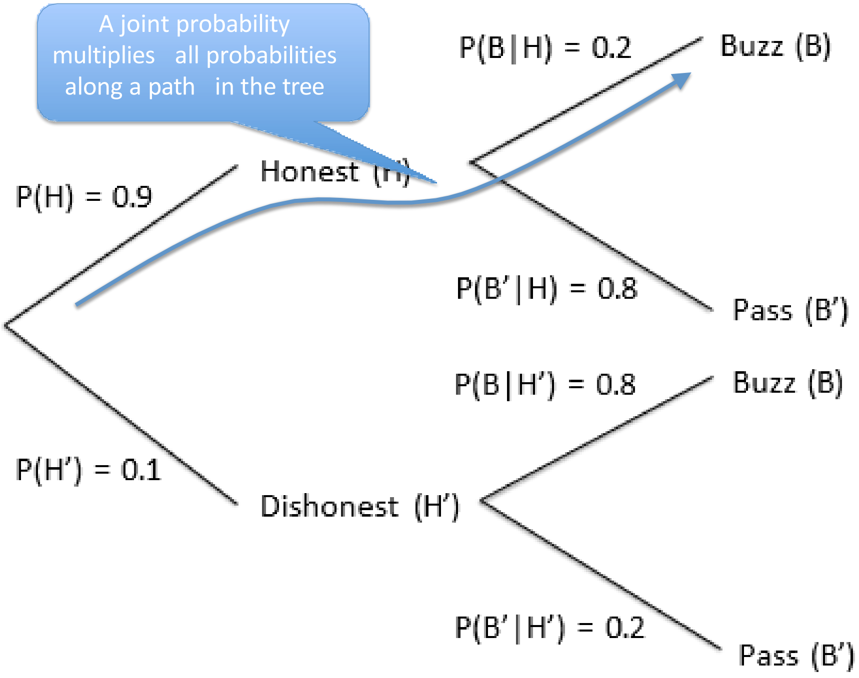
Probability trees: another useful tool



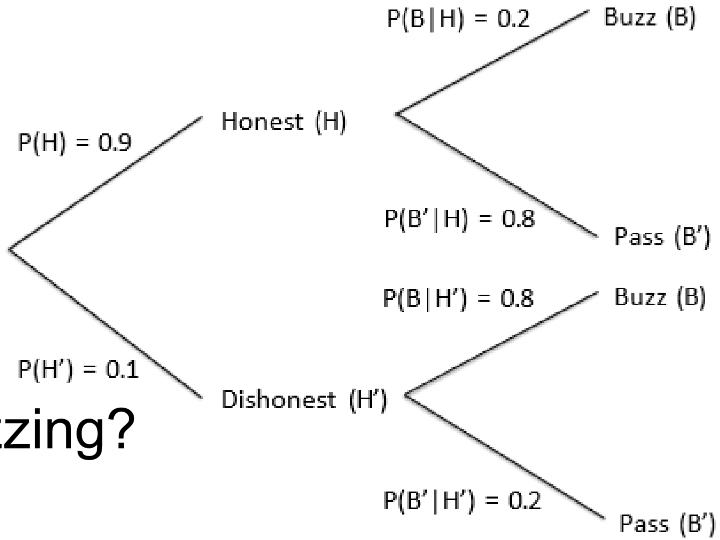
Probability trees: another useful tool



Probability trees: another useful tool



Probability trees: another useful tool



Let's use this to answer:

- What is $P(\text{Buzz}, \text{Dishonest})$?
  - 0.08
- $P(\text{Buzz})$?
  - Hint: which branches end up with buzzing?
  - 0.26
- $P(\text{Dishonest} \mid \text{Buzz})$?
  - Hint: which of the prev. branches contains the Dishonest event?
  - 0.08 / 0.26
  - We will soon see tools that can simplify this calculation a bit..

In-class activity: COVID test

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)　　$P(+ \mid N) = 0.01$

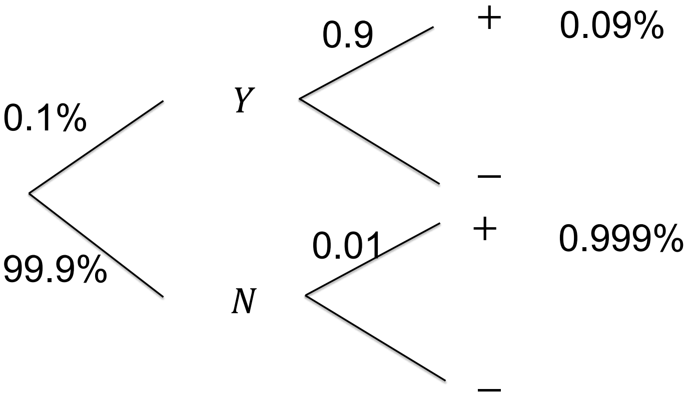
- One person in 1,000 has the disease.　　$P(Y) = 0.1\%$

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?



- $P(+,Y) = P(+|\ Y\ )P(Y) = 0.09\%$
- $P(+,N) = P(-|\ N\ )P(N) = 0.999\%$

P(+ | Y) = 0.9

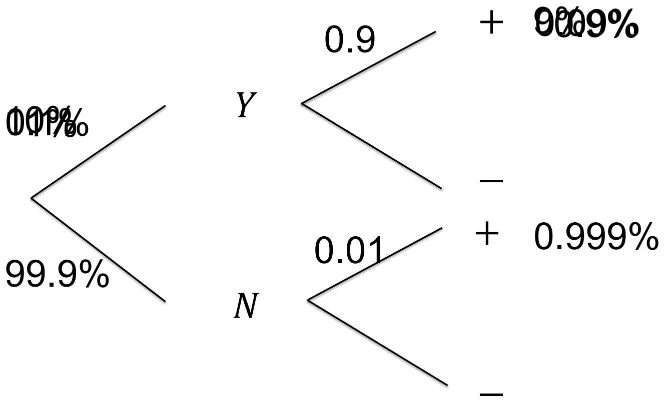
P(+ | N) = 0.01

P(Y) = 0.001

- $P(Y\mid+) = \frac{P(+,Y)}{P(+)} = \frac{0.09\%}{0.09\%+0.999\%} \approx \frac{1}{12}$
- Conclusion: being tested positive does not mean much..

COVID test: additional insights

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



Note: this branch's value is imprecise, but roughly stays at 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

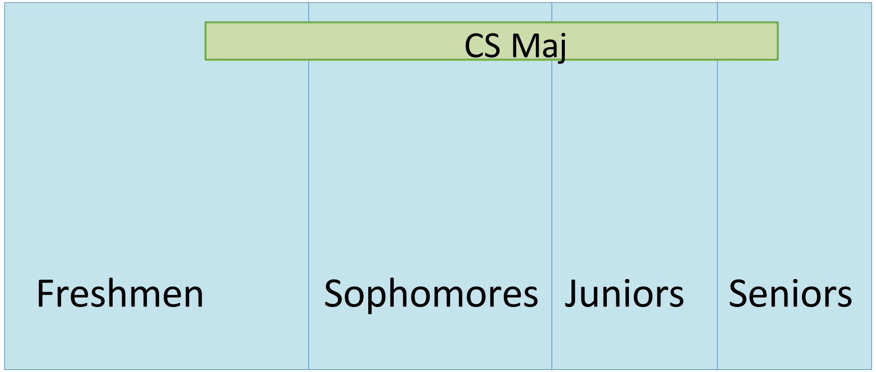
Rules of Probability

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

- $P(S|A) = 1$
- $P(E|A) + P(E^C|A) = 1$
- $P(E|A) + P(F|A) = P(E \cup F|A)$

A: CS major



Bayes rule

Reversing conditional probabilities

- Is $P(A \mid B) = P(B \mid A)$ in general?
- Let's see..

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

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

- Can you find a real-world example when they are unequal?
  - A: nurses; B: healthcare professionals

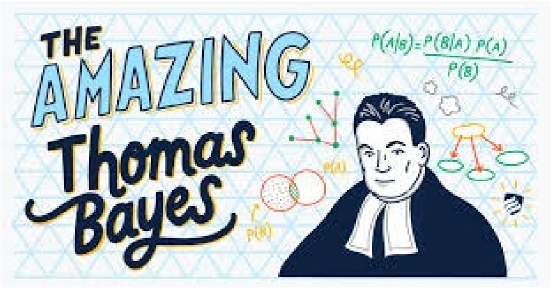
- Can I write one in terms of the other?

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

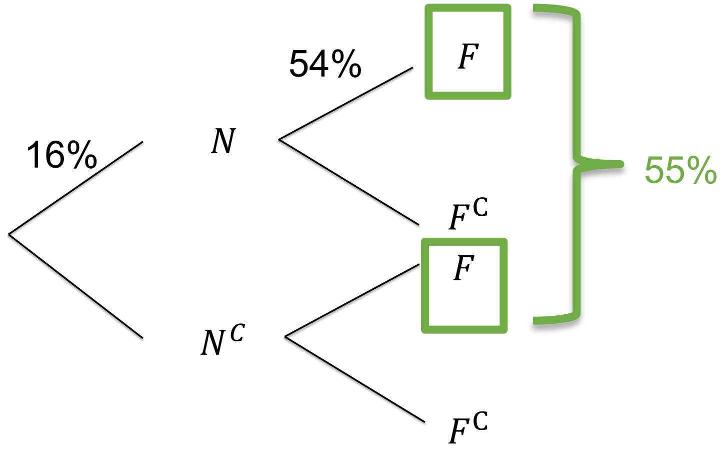
- Examples:
- A: I have COVID, B: my test shows positive
- A: employee lies B: the lie detector buzzes
- A: student is CS major B: student is a senior

Bayes rule: another example

- 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?

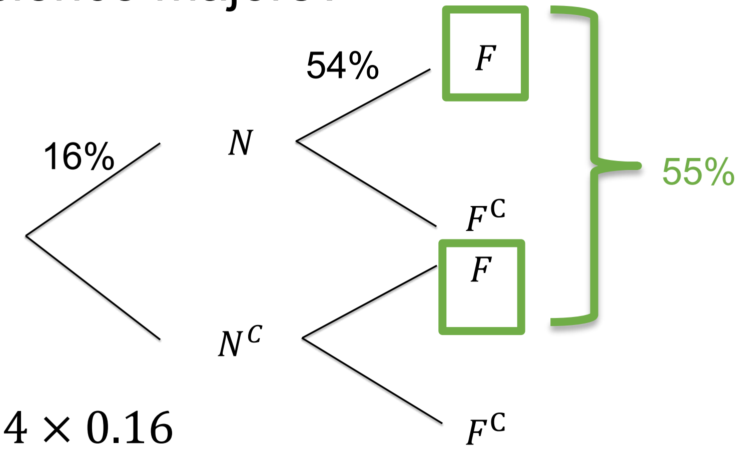
- What is the probability tree of this?

- We are looking for $P(N \mid F)$



Bayes rule: another example

- 16% of the students are Nutrition Science majors and 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: another example

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.

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

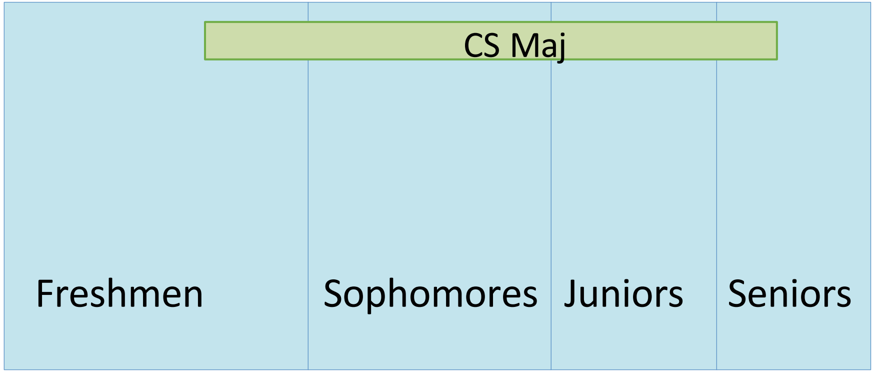
If we see a CS major student, what is their most likely year class?

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

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,

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



....

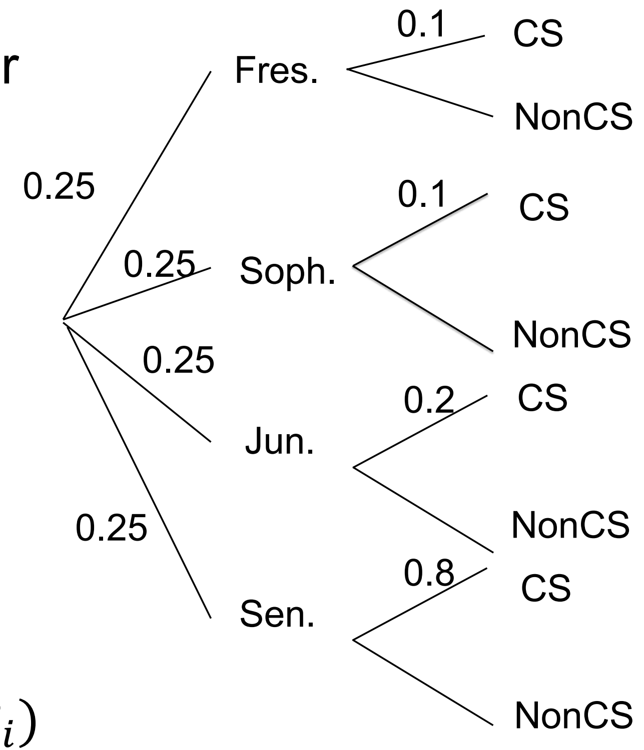
Bayes rule and Law of Total Probability

- 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
- Side note: $P(C)$ can be viewed as a *normalization factor*
- Equivalent form: $P(B_i \mid C) \propto P(B_i)P(C \mid B_i)$
  - $\propto$: proportional to

Independence

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

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

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

Probabilistic Independence

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

- Using Bayes' rule, we have

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

- So independence is indeed a symmetric notion

Independence: equivalent statement

- If A, B are independent, then their joint probability has a simple form:

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

- This is an equivalent characterization of independence

Independence (version 2)

If A and B are independent events, then

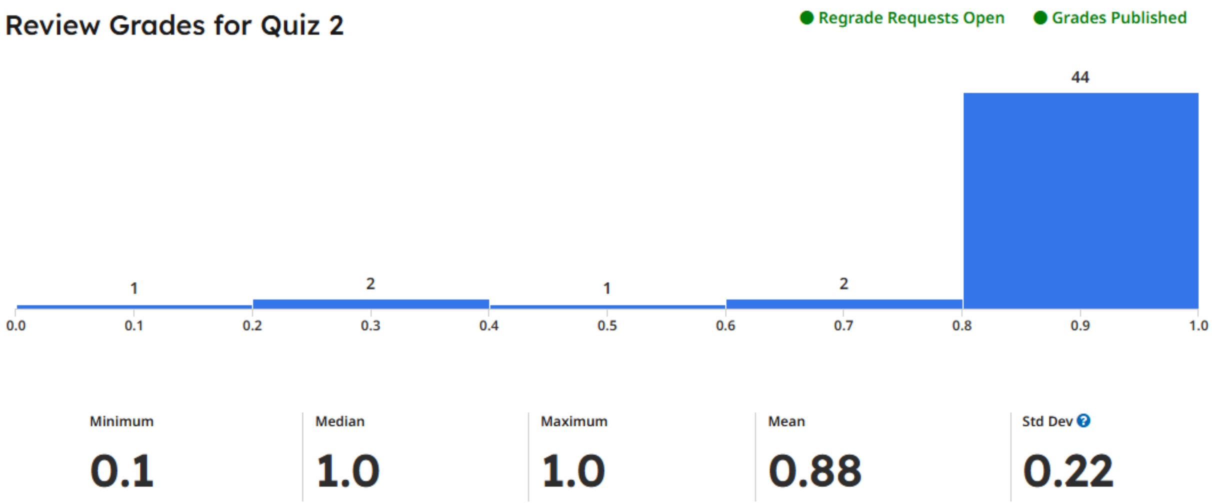
$$P(A \cap B) = P(A)P(B)$$

Are these independent events?

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

Announcements 2/10

- Quiz 2 graded; let us know if you have questions!



- We will have quiz 3 this Wed (2/12)

Recap: conditional probability

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

The "Chain Rule" of Probability

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

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

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

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

Extension: chain rule for conditional probability

- If we deal with more than 3 events happening together, we can apply the chain rule of probability repeatedly:

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

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

Recap: probability 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

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

Independence (version 2)

If A and B are independent events, then

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

Independence of several events

- We can generalize the notion of independence from two events to more than two.
  - E.g. A: employee is honest; B: rain tomorrow, C: stock price up

- Events $A_1, \dots, A_n$ are independent if for any subsets $A_{i_1}, \dots, A_{i_j}$,

$$P\left(A_{i_1}, \dots, A_{i_j}\right) = P\left(A_{i_1}\right) \cdot \dots \cdot P(A_{i_j})$$

Independence of several events

- E.g. if events A, B, C are independent, then
  - $P(A, B, C) = P(A) \cdot P(B) \cdot P(C)$
  - $P(A, C) = P(A) \cdot P(C)$
  - $P(B, C) = P(B) \cdot P(C)$
  - …

Probabilistic independence

- What's the sample space for rolling a die three times?
- How can we find the probability of the sequence (1, 2, 3)?

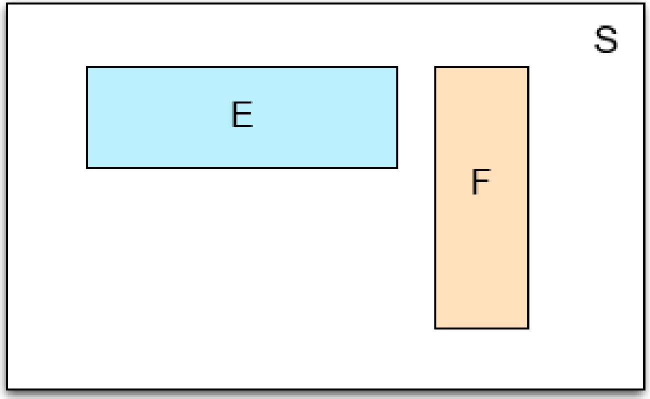
- What's the probability of the event that the sequence starts with 1? 1/6
- What's the conditional probability that we start with (1, 2) given that the first roll was a 1?
- We assume the rolls are independent, so the conditional probability is the same as the probability of rolling a 2. 1/6

- So what's the unconditional probability of a sequence beginning with (1, 2)? 1/6 x 1/6
- How about of the sequence (1, 2, 3)? 1/6 x 1/6 x 1/6

Independent vs. Disjoint Events

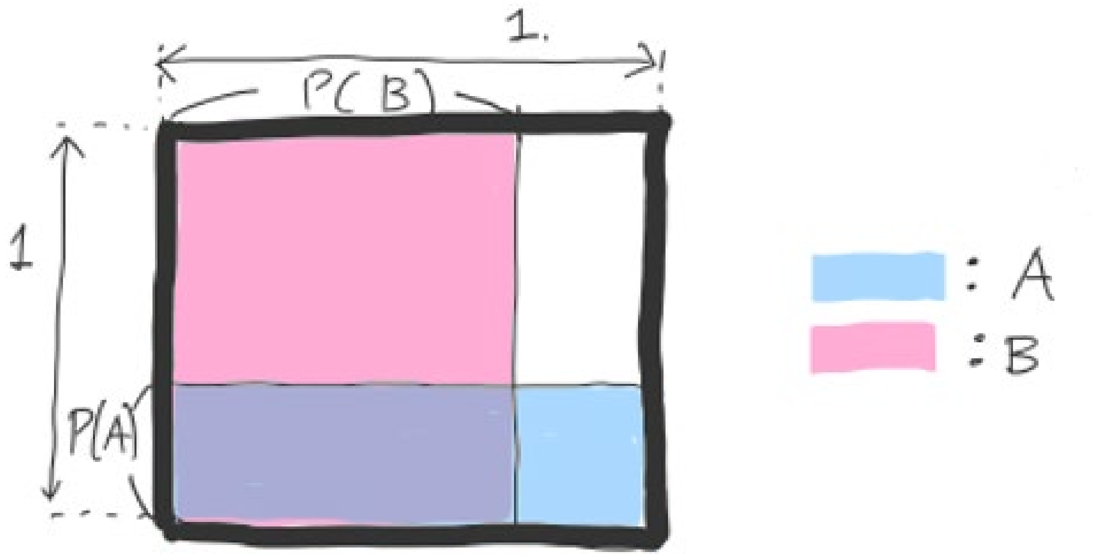
- Many people confuse independence with disjointness.
- They are very different!
- What does it mean for two events to be disjoint?
- If A and B are disjoint, then they cannot occur simultaneously; they are *mutually exclusive*; their intersection is the empty set.

- What does the Venn diagram look like?



Independent vs. Disjoint Events

- If A and B are independent, then P(B|A) = P(B).
- What does the Venn Diagram look like?



Independent vs. Disjoint Events

- If A and B are disjoint, what is P(B|A)?

$$P(B \mid A) = \frac{P(A,B)}{P(A)} = 0!$$

- Disjointness is practically the opposite of independence: if A occurs, you have all the information about whether B will occur.
    - Specifically, B doesn't occur

Independent vs. Disjoint Events

- Defining property of independent events:

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

- Defining property of disjoint events:

$$P(A \cap B) = 0$$

In-class activity: the absent-minded diners

- Three friends decide to go out for a meal, but they forget where they're going to meet.

- Fred decides to throw a coin. If it lands heads, he'll go to the diner; tails, and he'll go to the Italian restaurant.
- George throws a coin, too: heads, it's the Italian restaurant; tails, it's the diner.
- Ron decides he'll just go to the Italian restaurant because he likes the food.

- What's the probability that all three friends meet? 0.5 * 0.5 = 0.25
- What's the probability that one of them eats alone? 1 - 0.25 = 0.75

Summary

Conditional Probability Summary

- Representing conditional probabilities using contingency tables, Venn diagrams, and probability trees.
- The chain rule
- Bayes rule
- The law of total probability
- Independent events
- Disjoint events

Probability and Combinatorics

Probability and Combinatorics

- Combinatorics (in CSc144) are useful in calculating probabilities
- Recall: when all outcomes are equally likely:

We will use combinatorics to do counting

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

Number of elements in event set

Number of possible outcomes (e.g. 36)

- We will also see its another usage in a popular example: repeated independent trials (Bernoulli trials)

Example 1: sampling without replacement

- A president (P) and a treasurer (T) chosen from 20 people including Alice, Bob.
- Probability that Alice is president and Bob is treasurer?

- #ways to select (P,T) = $20 \times 19 = 380$
- P(E) = $\frac{1}{380}$

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

Number of elements in event set

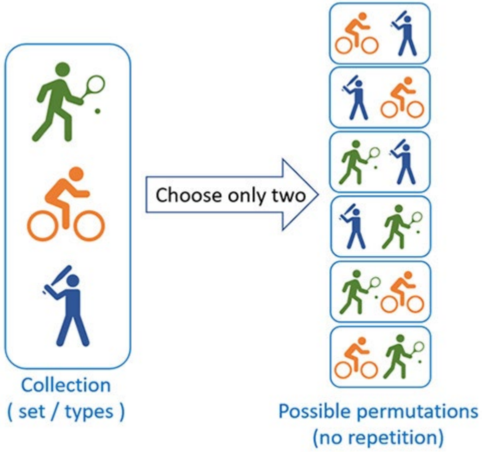
Number of possible outcomes (e.g. 36)

Permutation number

- If *ordered* selection of k items out of n is done without replacement, there are

$$n \times (n-1) \times \cdots \times (n-k+1) = \frac{n!}{(n-k)!}$$

outcomes



Example 2: Birthdays

- Probability that 2 in a group of 20 have same birthday?
- Sample space: $S = \{(n_1, \ldots, n_{20}) \colon n_1, \ldots, n_{20} \in \{1, \ldots, 365\}\}$
- What is $|S|$?
  - 365^{20}

- E: set of outcomes where two have same birthday, e.g. (5,5,176, .., 80)

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

Number of elements in event set

Number of possible outcomes (e.g. 36)

Example 2: Birthdays

- Let's try to calculate $|E|$
- It turns out that it is easier to calculate $|E^C|$

- E^C: all 20 birthdays are different

185	13	359	..	243	19

- $|E^C| = \frac{365!}{(365-20)!}$

- $P(E) = 1 - P(E^C) = 1 - \frac{365!}{(365-20)! \cdot 365^{20}} \approx 0.411$

This is quite high?! "Birthday paradox"

Repeated independent trials

Example The house or the player?

- 4 dice are rolled:
- House wins if at least one die is a 6, otherwise player wins.
- What is the probability that the house wins?

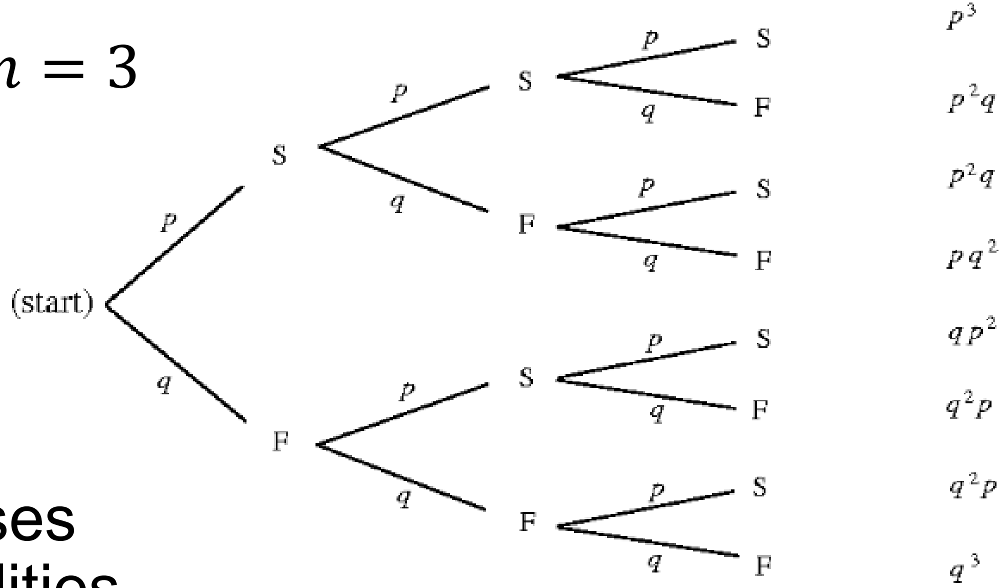
- What if I change the rule to “House wins if at least two die is a 6”?

Repeated independent trials (Bernoulli trials)

- In general, we are interested in the question:
- Suppose we repeatedly perform an experiment n independent times, each with success probability p, what is the probability that we succeed m times?
- The gambling example above: $n = 4$, $p = \frac{1}{6}$, $m = 1,2,3,4$
- Applications: sports analytics, gene mutations, etc.
- Named after Jacob Bernoulli (1655-1705)

Repeated independent trials: analysis

- Let's draw a probability tree!
- For simplicity let's look at $n = 3$
- Let $q \coloneqq 1 - p$



Observations:

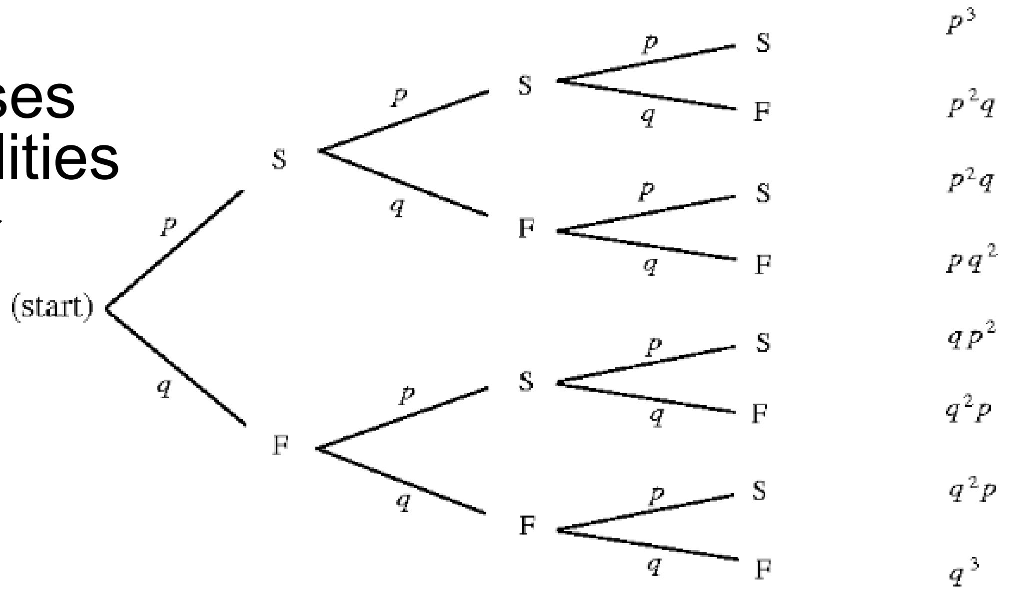
- $2^n = 8$ paths
- Paths with same #successes (m) have identical probabilities
  - They are equal to $p^m q^{n-m}$

Repeated independent trials: analysis

Observations:

- $2^n = 8$ paths
- Paths with same #successes (m) have identical probabilities
  - They are equal to $p^m q^{n-m}$

- How many paths have 3 successes?
  - 2?
  - 1?
  - 0?

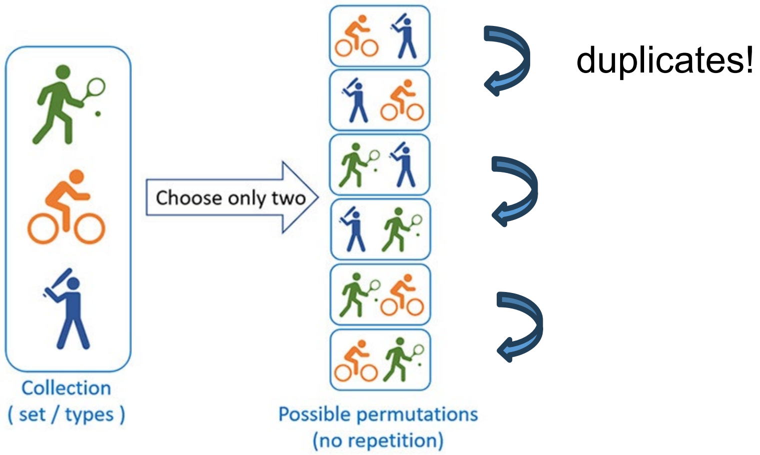


Combination number

- If *unordered* selection of k items out of n is done without replacement, there are

$$\frac{n!}{(n-k)!\ k!} =: \binom{n}{k}$$

outcomes



Repeated independent trials: analysis

- Out of all $2^3 = 8$ paths, the paths with 2 successes are:

  SSF, SFS, FSS

  # such paths is $\binom{3}{2}$

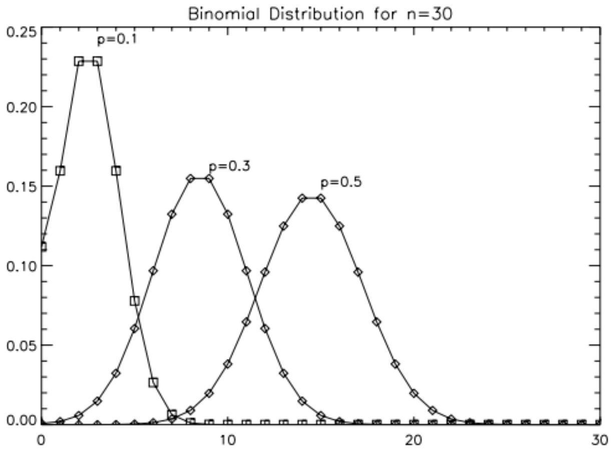
- In general, given n trials, #paths with m successes is $\binom{n}{m}$
  - select m different success positions out of n slots

- Thus, $P(m \text{ successes}) = \binom{n}{m} \cdot p^m q^{n-m}$

Repeated independent trials: conclusion

- In summary, in an experiment with n repeated independent trials with success probability p,

$$P(m \text{ successes}) = \binom{n}{m} \cdot p^m (1-p)^{n-m}$$



- The (random) number of successes is said to follow a *binomial distribution (more to come next lecture)*

.. Back to gambling

Example The house or the player?

- 4 dice are rolled:
- House wins if at least one die is a 6, otherwise player wins.
- What is the probability that the house wins?

- We have $n = 4$ repeated independent trials
- Here, "success" = "die is a 6"
- The asked probability is $P(\geq 1 \, \text{success})$

.. Back to gambling

- We do $n = 4$ repeated independent trials
- Here, "success" = "die is a 6"
- The asked probability is $P(\geq 1 \text{ successes})$

- $P(\geq 1 \text{ successes}) = \sum_{i=1}^{4} P(i \text{ successes})$

  $= \sum_{i=1}^{4} \binom{4}{i} \cdot \left(\frac{1}{6}\right)^{i} \left(\frac{5}{6}\right)^{4-i} = 0.518$
- Take home message: the house always wins ☺

.. Back to gambling

- There is another easier way to think about this problem..

$$P(\geq 1 \text{ successes}) = 1 - P(0 \text{ successes})$$

Complementary rule

$$P(0 \text{ successes}) = P(\text{Fail}_1, \ldots, \text{Fail}_4) = \left(\frac{5}{6}\right)^4$$

Independence of the 4 dice rolls

Backup

Application of Bayes rule: COVID test

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)

  $P(+ \mid Y) = 0.9$

- 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. $P(Y) = 0.001$

Q: What is the probability that a person with positive test has the disease? $P(Y \mid +)$?

Application of Bayes rule: COVID test

- We could solve this problem by filling in probability table, or probability tree

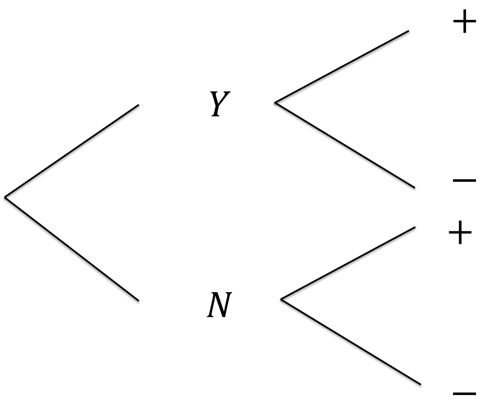
P(+ | Y) = 0.9

P(+ | N) = 0.01

P(Y) = 0.001

- Let's try Bayes rule this time..

$$P(Y \mid +) = \frac{P(+ \mid Y)P(Y)}{P(+)}$$

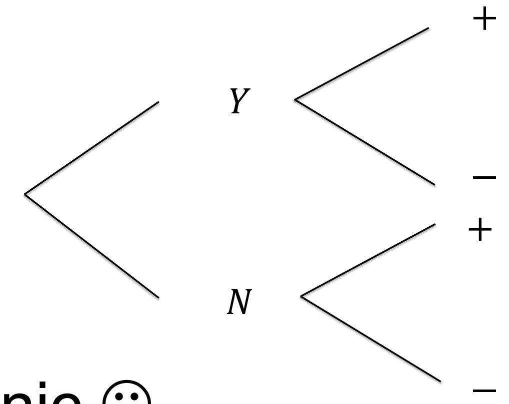


- $P(+ \mid Y)P(Y) = 0.9 \cdot 0.001 = 0.0009$
- What about $P(+)$?

Application of Bayes rule: COVID test

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

- $P(+) = P(+|Y)P(Y) + P(+|N)P(N)$
  - First part, $P(+|Y)P(Y)$, was calculated before (0.0009)
  - Second part, $P(+|N)P(N) = 0.01 \cdot 0.999 = 0.00999$



Therefore,

$$P(Y \mid +) = \frac{0.0009}{0.0009+0.00999} \approx \frac{1}{1+11} = 0.083$$

Conclusion: if tested positive, no need to panic ☺

Announcements

HW1 has been out

- D2L -> Content
- Due next Friday, Jan 26 by 11:59 pm

Participation policy (5 points)

- Each office hour: + 1 point
- Answering question in the lecture: + 1point
- Answering question on Piazza: + 1 point
- Asking question (related to course materials) on Piazza: +0.5 point

Note: It is your responsibility to ensure the TA or instructor enter your participation points on gradescope during the office hour or after each lecture. Instructors will not award you these points at a later date, do not email instructors about getting points at a later date (for example, if you forget to ask the TA to enter your office hour points on gradescope).

Review

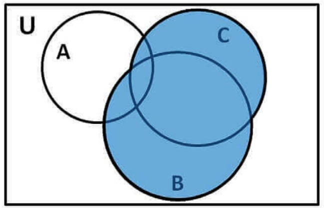
- What is probability?
- Axioms
- Event = set $\Rightarrow$ use set theory!
- Set theory + axiom 3 is quite useful
- Draw diagrams
- Lots of jargons

- Make your own cheatsheet.

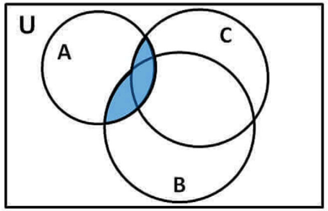
Review

- $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

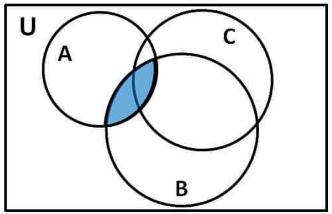
distributive law by Venn diagram



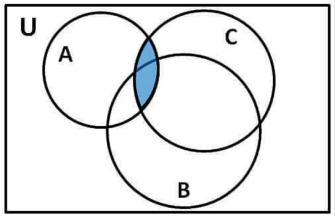
(B ∪ C)



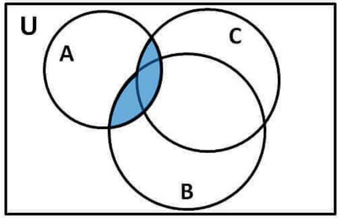
A ∩ (B ∪ C)



(A ∩ B)

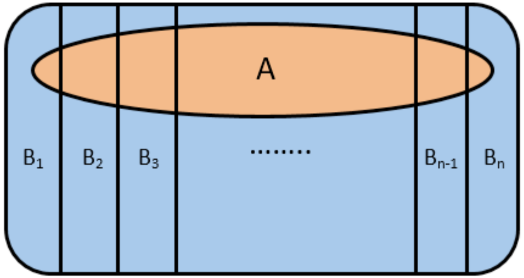


(A ∩ B)



(A ∩ B) ∪ (A ∩ C)

Review



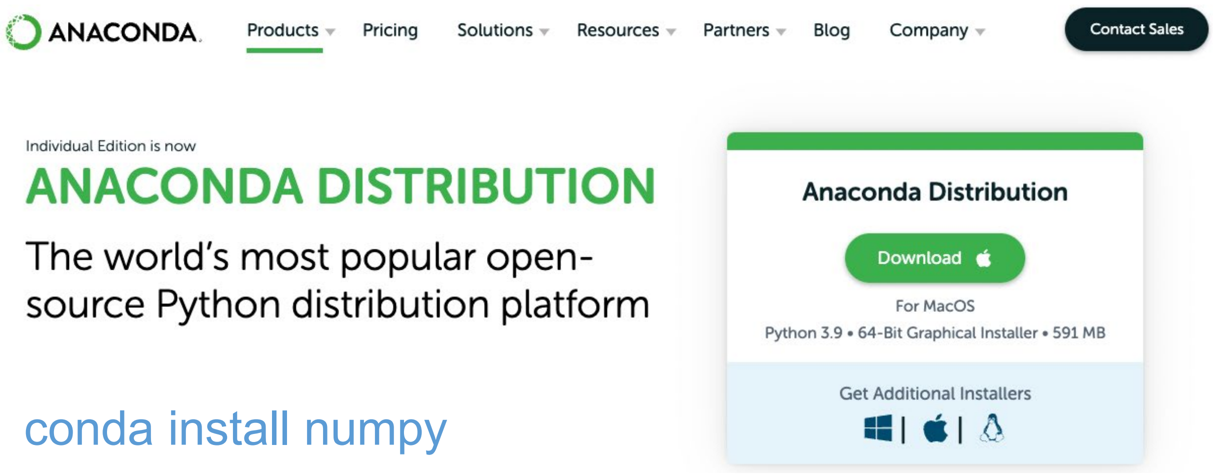
- $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
- $A = A \cap \Omega = A \cap (\cup_i B_i) = \cup_i (A \cap B_i)$
  $= A \cap (B_1 \cup B_2 \cup B_3 \ldots \cup B_n)$
  $= (A \cap B_1) \cup (A \cap B_2) \cup (A \cap B_3) \ldots \cup (A \cap B_n)$

<u>Law of total probability</u>*:* Let A be an event. For any events $B_1, B_2, \ldots$ that partitions Ω, we have

$$P(A) = \sum_i P(A \cap B_i)$$

Numpy Library

Package containing many useful numerical functions…



conda install numpy

If you use pip:
pip install numpy

…we are interested in numpy.random at the moment

numpy.random

numpy.random.randint

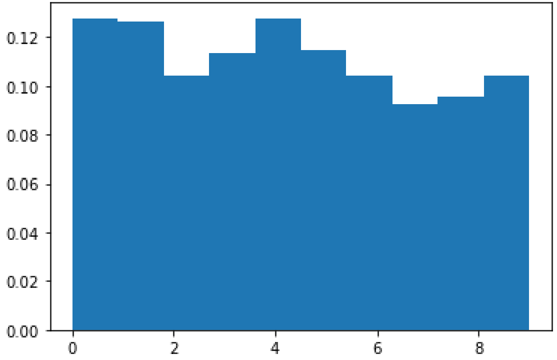
numpy.random.randint(*low*, *high=None*, *size=None*, *dtype='l'*)

Return random integers from *low* (inclusive) to *high* (exclusive).

Return random integers from the "discrete uniform" distribution of the specified dtype in the "half-open" interval [*low*, *high*). If *high* is None (the default), then results are from [0, *low*).

Sample a discrete uniform random variable,

```
import matplotlib.pyplot as plt
X = np.random.randint(0,10,1000)
count, bins, ignored = plt.hist(X, 10, density=True)
plt.show()
```



- **Caution** Interval is [low,high) and upper bound is **exclusive**
- `Size` argument accepts tuples for sampling ndarrays (multidimentional arrays)

numpy.random

Allows sampling from many common distributions

Set (global) random seed as,

```python
import numpy as np

seed = 12345
np.random.seed(seed)
```

- ☺ easier to debug (otherwise, you may have 'stochastic' bug)
- ☹ can be risky

  E.g., buy into the result based on a particular seed, publish a report.
  … turns out, you get a widely different result if you use a different seed!

  Recommendation: change the seed every now and then

Random Events and Probability

Consider: What is the probability of having two numbers sum to 6?

```
import numpy as np
for n in [10,100,1_000,10_000,100_000]:
    res_dice1 = np.random.randint(1,6+1,size=n)
    res_dice2 = np.random.randint(1,6+1,size=n)
    res = [(res_dice1[i], res_dice2[i]) for i in range(len(res_dice1))]

    cnt = len(list(filter(lambda x: x[0] + x[1] == 6, res)))
    print("n=%6d, result: %.4f " % (n, cnt/n))
```

```
n=      10, result: 0.1000
n=     100, result: 0.1200
n=    1000, result: 0.1350
n=   10000, result: 0.1365
n=  100000, result: 0.1388
n= 1000000, result: 0.1385
```

```
n=      10, result: 0.1000
n=     100, result: 0.1900
n=    1000, result: 0.1540
n=   10000, result: 0.1366
n=  100000, result: 0.1371
n= 1000000, result: 0.1394
```

every time you run, you get a different result

however, the number seems to **converge** to 0.138-0.139

There seems to be a precise value that it will converge to.. what is it?

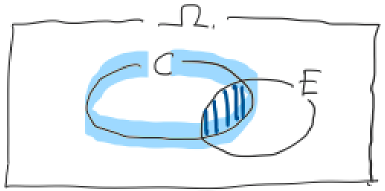
Conditional Probability

- Suppose I roll two dice secretly and tell you that <u>one of the dice is 2</u>. C
- **<u>In this situation</u>**, find the probability of two dice summing to 6. E

```
import numpy as np
for n in [10,100,1000,10_000,100_000, 1_000_000]:
    res_dice1 = np.random.randint(6,size=n) + 1
    res_dice2 = np.random.randint(6,size=n) + 1
    res = [(res_dice1[i], res_dice2[i]) for i in range(len(res_dice1))]

    conditioned = list(filter(lambda x: x[0] == 2 or x[1] == 2, res))
    n_eff = len(conditioned)

    cnt = len(list(filter(lambda x: x[0] + x[1] == 6, conditioned)))
    print("n=%9d, n_eff=%9d, result: %.4f " % (n, n_eff, cnt/n_eff))
```



Without conditioning, it was 0.138.

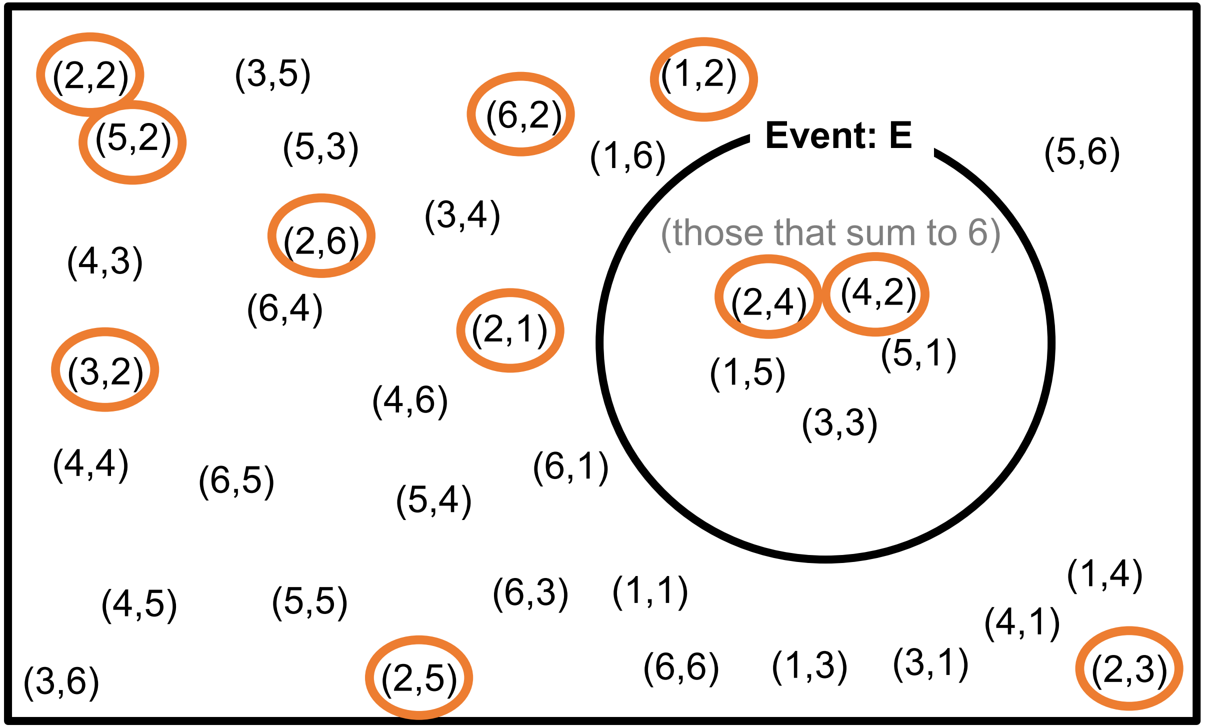
```
n=       10, n_eff=        4, result: 0.0000
n=      100, n_eff=       32, result: 0.2500
n=     1000, n_eff=      300, result: 0.1733
n=    10000, n_eff=     3002, result: 0.1742
n=   100000, n_eff=    30590, result: 0.1823
n=  1000000, n_eff=   305616, result: 0.1818
```

```
n=       10, n_eff=        3, result: 0.3333
n=      100, n_eff=       32, result: 0.0625
n=     1000, n_eff=      343, result: 0.2245
n=    10000, n_eff=     3062, result: 0.1897
n=   100000, n_eff=    30651, result: 0.1811
n=  1000000, n_eff=   305580, result: 0.1808
```

There seems to be a precise value that it will converge to.. what is it?

Random Events and Probability

What is the probability of having two numbers sum to 6 given one of dice is 2?



Each outcome is equally likely.
by the **independence**
(will learn this concept later)
=> 1/36

sum to 6:
=> 5

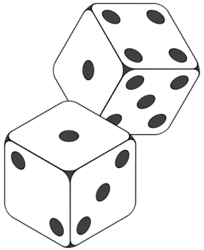
one of dice is 2:
=> 11

sum to 6 and one of dice is 2:
=> 2

answer:
2/11 = 0.181818….

Conditional Probability

Two fair dice example

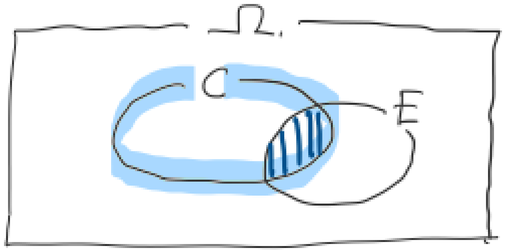


- Find the probability of one of the dice is 2 (event C) and two dice summing to 6 (E)

$$P(E \cap C)$$

- I secretly tell you one of the dice is 2, find the probability of two dice summing to 6.

$$\frac{P(E \cap C)}{P(C)}$$



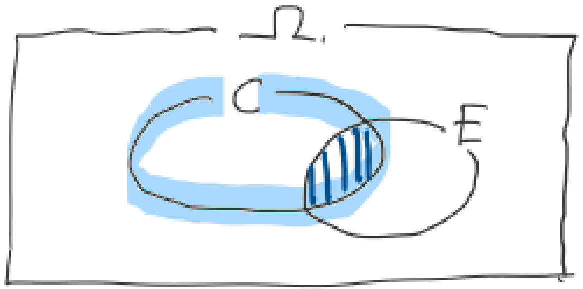
Conditional Probability

- Two fair dice example:
  - Suppose I roll two dice and secretly tell you that one of the dice is 2. C
  - **<u>In this situation</u>**, find the probability of two dice summing to 6. E

- Turns out, such a probability can be computed by $\frac{P(E \cap C)}{P(C)}$
- It's like "zooming in" to the condition.
- This happens a lot in practice, so let's give it a notation:

$$P(E|C) := \frac{P(E \cap C)}{P(C)}$$

Say: probability of "E given C", "E conditioned on C"



"it's the ratio"

Conditional Probability

Q: Conditional probability P(A|B) could be undefined. When?

- A: The denominator can be 0 already. In this case, numerator is also 0!

Note $P(A|B) \neq P(B|A)$ in general!

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

E.g., throw a fair die. X := outcome. A = {X=4}, B = {X is even}
Question: P(A | B) = P(B | A) ?

- $\mathrm{P(A)} = 1/6$
- $\mathrm{P(B)} = 1/2$
- $\mathrm{P}\ (\mathrm{A} \cap \mathrm{B}) = 1/6$
- Therefore, $\mathrm{P(A|B)} = 1/3, \quad \mathrm{P(B|A)} = 1$

Conditional Probability

Chain rule

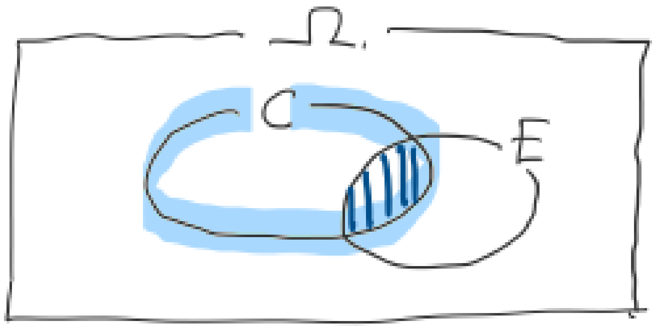
- $P(A \cap B) = P(A|B)P(B)$ ←just a rearrangement of definition: $P(A|B) := \frac{P(A \cap B)}{P(B)}$

- $P(A \cap B \cap C) = P(A|B \cap C)P(B \cap C) = P(A|B \cap C)P(B|C)P(C)$

- $P(E_1 \cap E_2 \cap \cdots \cap E_n) = P(E_1) \prod_{i=2}^{n} P(E_i | \cap_{j=1}^{i-1} E_j)$ valid for any ordering!

Conditional Probability

- $P(E \cap C) = P(E|C)P(C) = P(C|E)P(E)$

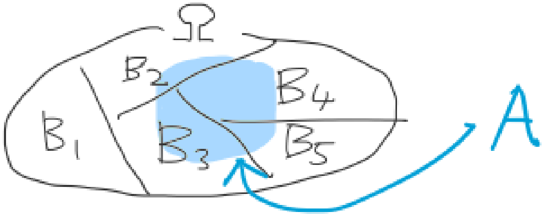


"it's the ratio"

Conditional Probability

Recall: let A be an event. For events $\mathrm{B}_1, \mathrm{B}_2, \ldots$ that partitions Ω, we have

$$P(A) = \sum_i P(A \cap B_i)$$



Law of total probability: If $A \in \mathcal{F}$ and $\{B_i \in \mathcal{F}\}_i$ partitions Ω, then

$$P(A) = \sum_i P(A, B_i) = \sum_i P(B_i) P(A | B_i)$$

$$= \sum_i P(A) P(B_i | A)$$

(by definition)

Shortcut:
P(A,B) := P(A ∩ B)

Conditional Probability

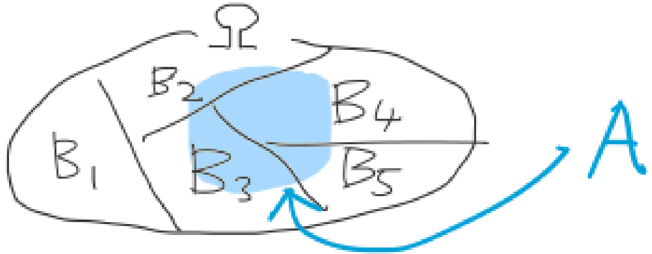
Law of total probability: If $A \in \mathcal{F}$ and $\{B_i \in \mathcal{F}\}_i$ partitions Ω, then

$$P(A) = \sum_i P(A, B_i) = \sum_i P(B_i)P(A|B_i)$$

$$= \sum_i P(A)P(B_i|A) \quad \text{(by definition)}$$

If we divide both sides by P(A):

$$1 = \sum_i P(B_i|A)$$



Conditional Probability: an example

$$P(A) = \sum_i P(A \cap B_i)$$

A: customer (100)

B: fill gas

- B_1: unleaded (30)
- B_2: mid grade (30)
- B_3: premium (40)

30, 1/3 | 30, 1/2 | 40, 1/8

Q: what's the probability that the customer is a student?

$$\begin{aligned}
&\mathrm{P}(\mathrm{A} = \text{student}) \\
&= \mathrm{P}(\mathrm{A} = \text{student}, \mathrm{B} = \mathrm{B}_1) + \mathrm{P}(\mathrm{A} = \text{student}, \mathrm{B} = \mathrm{B}_2) + \mathrm{P}(\mathrm{A} = \text{student}, \mathrm{B} = \mathrm{B}_3) \\
&= \mathrm{P}(\mathrm{A} = \text{student} | \mathrm{B} = \mathrm{B}_1)\mathrm{P}(\mathrm{B} = \mathrm{B}_1) + \mathrm{P}(\mathrm{A} = \text{student} | \mathrm{B} = \mathrm{B}_2)\mathrm{P}(\mathrm{B} = \mathrm{B}_2) + \\
&\mathrm{P}(\mathrm{A} = \text{student} | \mathrm{B} = \mathrm{B}_3)P(\mathrm{B} = \mathrm{B}_3)
\end{aligned}$$

Conditional Probability: an example

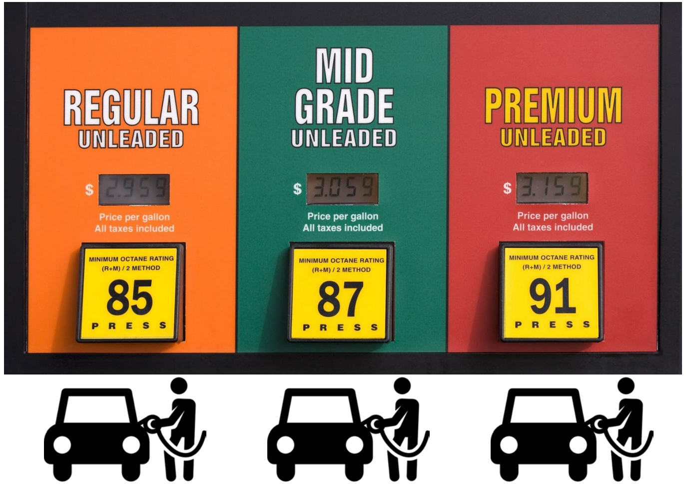
- $P(A) = \sum_i P(A, B_i) = \sum_i P(B_i) P(A|B_i)$

$\mathrm{P(A = student)}$
$\mathrm{= P(A = student|B = B_1)P(B = B_1) + P(A = student|B = B_2)P(B = B_2) + P(A = student|B = B_3)}P\mathrm{(B = B_3)}$

$\mathrm{P(A = student)}$
$= 1/3 \times 30/100 + 1/2 \times 30/100 + 1/8 \times 40/100$

- $\sum_i P(B_i|A) = 1$

$\mathrm{P(B_1|A = student) + P(B_2|A = student) + P(B_3|A = student)}$
$= \frac{10}{10+15+5} + \frac{15}{10+15+5} + \frac{5}{10+15+5} = 1$



30, ⅓: 10   30, ½: 15   40, ⅛: 5

Conditional Probability

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)

$P(+ \mid Y) = 0.9$

- A test for the disease yields a positive result 1% of the time when the disease is not present (N)

$P(+ \mid N) = 0.01$

$P(Y) = 0.001$

- One person in 1,000 has the disease.

$P(Y \mid +)$?

Q: What is the probability that a person with positive test has the disease?

Pick a person uniformly at random from the population. Apply the test. When test=+, what is the probability of this person having the disease (Y) ?

Conditional Probability

What we know:

P(+ | Y) = 0.9

P(+ | N) = 0.01

P(Y) = 0.001

⇒

P(- | Y) = 0.1

P(- | N) = 0.99

P(N) = 0.999

Question: $P(Y \mid +) = \dfrac{P(Y,+)}{P(+)}$

$$P(+) = P(+,Y) + P(+,N)$$

$$P(+,Y) = P(+|\,Y)P(Y)$$

$$P(+,N) = P(+|\,N)P(N)$$

Law of total probability

$$P(A) = \sum_i P(A,B_i) = \sum_i P(B_i)P(A|B_i)$$

The answer is 0.0826…

Terminology

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 $P(A^c)$

Conditional Probability

$P(+ \mid Y) = 0.9$
$P(+ \mid N) = 0.01$
$P(Y) = 0.001$

Tip: Make a table of **joint probabilities**

Each cell is P(column event ∩ row event) = $P(T=t \cap D=d)$ = $P(T=t \mid D=d)\,P(D=d)$

	Test = +	Test = -	
Disease=Y	$0.9 \cdot 0.001 = 0.0009$	$0.1 \cdot 0.001 = 0.0001$	0.001
Disease=N	$0.01 \cdot 0.999 = 0.00999$	$0.99 \cdot 0.999 = 0.98901$	0.999
	0.01089	0.98911	

$P(test = +)$

Workflow:

- make a table, then fill in the cells.
- write down the target P(A|B) all in terms of joint probabilities and marginal probabilities.

Conditional Probability

We can directly calculate:

$$P(Y|+) = \frac{P(Y,+)}{P(+)} = \frac{P(+|Y)P(Y)}{P(+)}$$

Bayes rule

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

proof: definition and definition!

⟹ particularly useful in practice: infer P(A|B) given P(B|A)!

P(A): **prior** probability

P(A|B): **posterior** probability

e.g., A='dice sum to 6', B='one of the die is 2'

e.g., A='disease=Y', B='test=+'

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:

- $P(C) = 0.58$
- $P(C \mid A) = \frac{P(C \cap A)}{P(A)} = \frac{0.48}{0.60} = 0.8$ Larger than $P(C)$
- Suppose that we see a buckled parent, it is more likely that we see their child buckled

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

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