Class starts after this song

TOTO – Africa (1982)

requested by Austin Huang (TA-of-CM8)

I enjoy playing many sports including basketball, tennis, and running. I studied Chinese in Taiwan over the summer and loved the local food. I enjoy pineapple on pizza.

- Remember *not to discuss* Exam 2 until the entire class has taken it and the grades are announced (ETA *Thursday*)

- Shao-Heng's Zoom consulting hours for assignment makeups
  - Thursdays 11am-noon
  - Tuesdays 4-5pm
  - Suitable for: if you have many past assignments to validate at a time
  - It's about time to patch the holes in your past assignments so that they don't pile up on you around LDoC

CS230 Spring 2024
Module 08: Probability

Duke

Why probability?

- Unlikely
- Likely
- Guaranteed to
- Certainly
- Who knows if
- Chances are
- Typically
- Frequently
- Most of the time
- …

Outcomes vs. Events

- Outcomes are disjoint possibilities
- Events let us talk about many possibilities together
- Both are artificial and manually-defined
  - Think about the weather tomorrow
  - We can define the outcomes as $\text{No Rain}, \text{Rain}$
  - Or we can define the outcomes as $\text{No Rain}, \text{Light Rain}, \text{Heavy Rain}$ and still can talk about the event $\textbf{Rain} = \{\text{Light Rain}, \text{Heavy Rain}\}$

Terminology Musing

- Events with only one outcome
  - Example: $\mathbf{No\ Rain} = \{\text{No Rain}\}$
  - $\{6\}$ when the experiment is throwing a six-face die
  - $\{6\}$ is an event with a single outcome; 6 is just an outcome
  - Some past instructors call such events “singleton events”
  - MFADM also calls all the outcomes “elementary events”
    - Now this is a misnomer that doesn’t well distinguish $\{6\}$ and 6
    - We will discourage using the term “elementary events”

Terminology Musing

- Expressing a probability distribution/measure:
  - MFADM uses $\mathrm{Pr}(\cdot)$
  - Other people/books use $\mathrm{P}(\cdot)$, $\mathrm{p}(\cdot)$, or $\mathrm{Prob}(\cdot)$
  - It doesn't matter which one to use as long as the context is clear (it is THE probability distribution you're referring to)
  - The notation $\mathrm{Pr}(\cdot)$ or $\mathrm{Prob}(\cdot)$ when used should NOT be italicized
    - Much more okay for $\mathrm{P}(\cdot)$ and $\mathrm{p}(\cdot)$ to be italicized

Duke

“Intuition of equal probability”

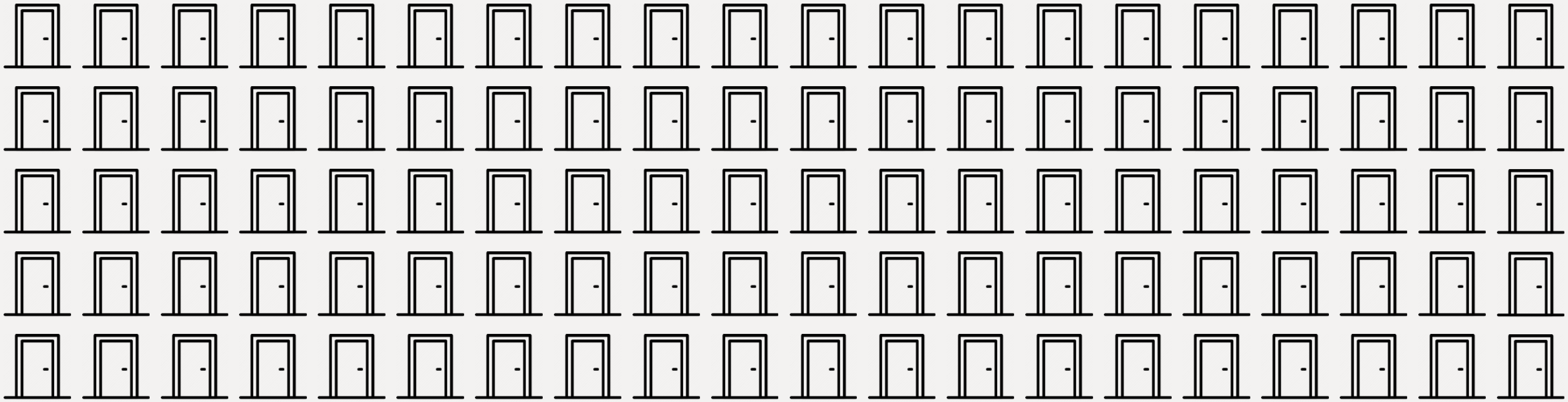
- It is intuitive to think that “without further information, all outcomes in an experiment have equal probability”
- Problem: outcomes are artificial
  - When the outcomes are No Rain, Rain: 50% probability to see rain?
  - When the outcomes are No Rain, Light Rain, Heavy Rain: 66.6%?
- Equal probability is an assumption, needs to be acknowledged
  - Whether it makes sense for modeling the real world is subjective

“Intuition of equal probability”, Monty Hall

- Without further information,
  we reasonably assume that the car is behind each door with equal probability before one picks the original door
- Without careful analysis,
  some people believe the car is behind the original door (no change) and the other remaining door (yes change) with equal probability
  - Both are just assumptions, strictly speaking
- The Monty Hall problem is counterintuitive because
  those two assumptions cannot hold at the same time

Monty Hall to the extreme

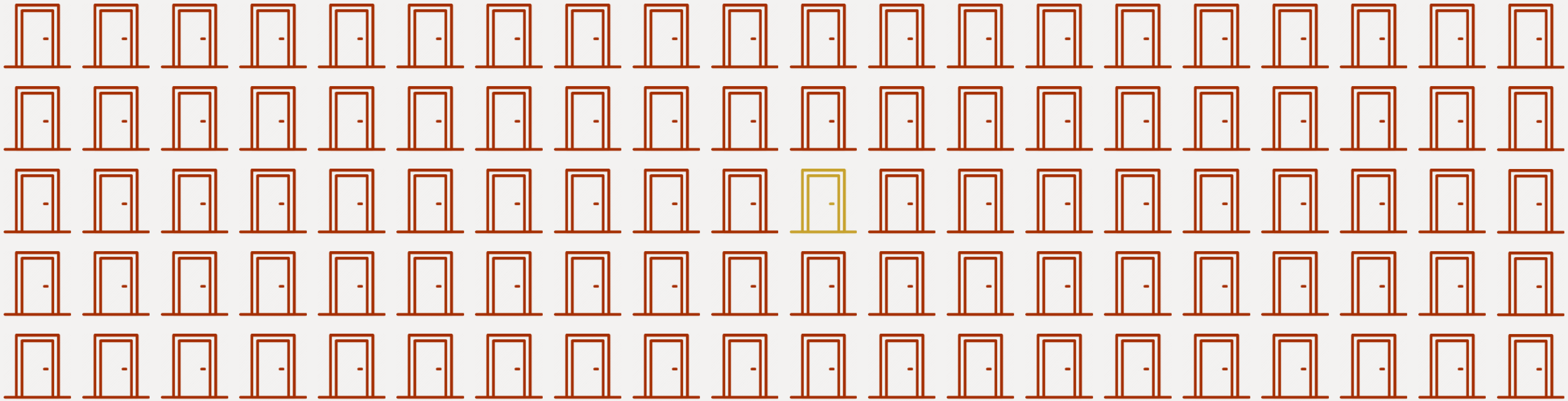
- Same scenario, but 100 doors (host will open 98 out of 99)



Duke

Monty Hall to the extreme

- Reasonable assumption: there is a $1/100$ chance to select the right door



Monty Hall, a conditional probability perspective

- Reasonable assumption: there is a $1/100$ chance to select the correct door
- $\Pr(\text{win car with no change}|\text{selected correct door}) = 1$
- $\Pr(\text{win car with door change}|\text{selected correct door}) = 0$
- $\Pr(\text{win car with no change}|\text{selected incorrect door}) = 0$
- $\Pr(\text{win car with door change}|\text{selected incorrect door}) = 1$

- $$\begin{aligned}\Pr(\text{win with no change}) &= \Pr(\text{win with no change}|\text{correct}) \cdot \Pr(\text{correct}) \\ &+ \Pr(\text{win with no change}|\text{incorrect}) \cdot \Pr(\text{incorrect}) \\ &= 1 \cdot \frac{1}{100} + 0 \cdot \frac{99}{100} = \frac{1}{100}\end{aligned}$$

Monty Hall, a conditional probability perspective

- Reasonable assumption: there is a $1/100$ chance to select the correct door
- $\Pr(\text{win car with no change}|\text{selected correct door}) = 1$
- $\Pr(\text{win car with door change}|\text{selected correct door}) = 0$
- $\Pr(\text{win car with no change}|\text{selected incorrect door}) = 0$
- $\Pr(\text{win car with door change}|\text{selected incorrect door}) = 1$

- $$\begin{aligned}\Pr(\text{win with door change}) &= \Pr(\text{win with door change}|\text{correct}) \cdot \Pr(\text{correct}) \\ &+ \Pr(\text{win with door change}|\text{incorrect}) \cdot \Pr(\text{incorrect}) \\ &= 0 \cdot \frac{1}{100} + 1 \cdot \frac{99}{100} = \frac{99}{100}\end{aligned}$$

Duke

Monty Hall, general case (n doors)

- Reasonable assumption: there is a $1/n$ chance to select the correct door
- $\Pr(\text{win with no change}) = \Pr(\text{win with no change} | \text{correct}) \cdot \Pr(\text{correct}) + \Pr(\text{win with no change} | \text{incorrect}) \cdot \Pr(\text{incorrect})$

$$= 1 \cdot \frac{1}{n} + 0 \cdot \frac{n-1}{n} = \frac{1}{n}$$

- $\Pr(\text{win with door change}) = \Pr(\text{win with door change} | \text{correct}) \cdot \Pr(\text{correct}) + \Pr(\text{win with door change} | \text{incorrect}) \cdot \Pr(\text{incorrect})$

$$= 0 \cdot \frac{1}{n} + 1 \cdot \frac{n-1}{n} = \frac{n-1}{n}$$

Duke

Monty Hall, classical case (3 doors)

- Reasonable assumption: there is a $1/3$ chance to select the correct door
- $\Pr(\text{win with no change}) = \Pr(\text{win with no change}|\text{correct}) \cdot \Pr(\text{correct})$
  $+ \Pr(\text{win with no change}|\text{incorrect}) \cdot \Pr(\text{incorrect})$
  $= 1 \cdot \frac{1}{3} + 0 \cdot \frac{2}{3} = \frac{1}{3}$

- $\Pr(\text{win with door change}) = \Pr(\text{win with door change}|\text{correct}) \cdot \Pr(\text{correct})$
  $+ \Pr(\text{win with door change}|\text{incorrect}) \cdot \Pr(\text{incorrect})$
  $= 0 \cdot \frac{1}{3} + 1 \cdot \frac{2}{3} = \frac{2}{3}$

Duke

Monty Hall (classical), a Bayesian perspective

- A priori: there is a $1/3$ chance to select the correct door
- $\Pr(\text{win car with no change}|\text{selected correct door}) = 1$
- $\Pr(\text{win car with door change}|\text{selected correct door}) = 0$
- $\Pr(\text{win car with no change}|\text{selected incorrect door}) = 0$
- $\Pr(\text{win car with door change}|\text{selected incorrect door}) = 1$

- $$\begin{aligned}\Pr(\text{win with door change}) &= \Pr(\text{win with door change}|\text{correct}) \cdot \Pr(\text{correct}) \\ &+ \Pr(\text{win with door change}|\text{incorrect}) \cdot \Pr(\text{incorrect}) \\ &= 0 \cdot \frac{1}{3} + 1 \cdot \frac{2}{3} = \frac{2}{3}\end{aligned}$$
- Therefore a posteriori: there is a $2/3$ chance to win if we switch

"School of thoughts" of probability

"Abstract" / "pure"

- $\Pr(\cdot)$ is just a mathematical function with beautiful properties that we can study
- It may or may not model the real world, and we might not care

Bayesian

- Everything is subjective; probability models our beliefs about things
- Data/evidence/events let us update our beliefs

Frequentist

- Everything is deterministic; uncertainties are just due to not being able to observe everything
- Data/evidence/events will converge to the "true underlying distribution"

Terminology Musing

- At random does not well-specify the probability model
- There is a bag with 3 balls: red, blue, green.
  We draw a ball from the bag at random.
  - $\Pr(\text{red}) = \Pr(\text{blue}) = \Pr(\text{green}) = \frac{1}{3}$?
  - $\Pr(\text{red}) = 56.1\%, \Pr(\text{blue}) = 7.1\%, \Pr(\text{green}) = 36.8\%$?
- The correct wording is uniformly (at) random
  - When you see at random, assume equal probability and state this assumption

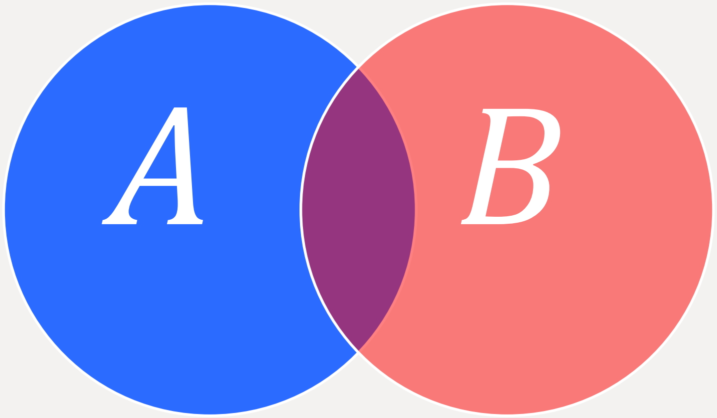
“Bayes’ Theorem”

- Past 230 students showed “overreliance” on Bayes’ Theorem:

$$\Pr(B|A) = \frac{\Pr(A|B) \cdot \Pr(B)}{\Pr(A)}$$

  - This includes trying to apply the theorem everywhere, including where the theorem does not help

"Bayes' Theorem"

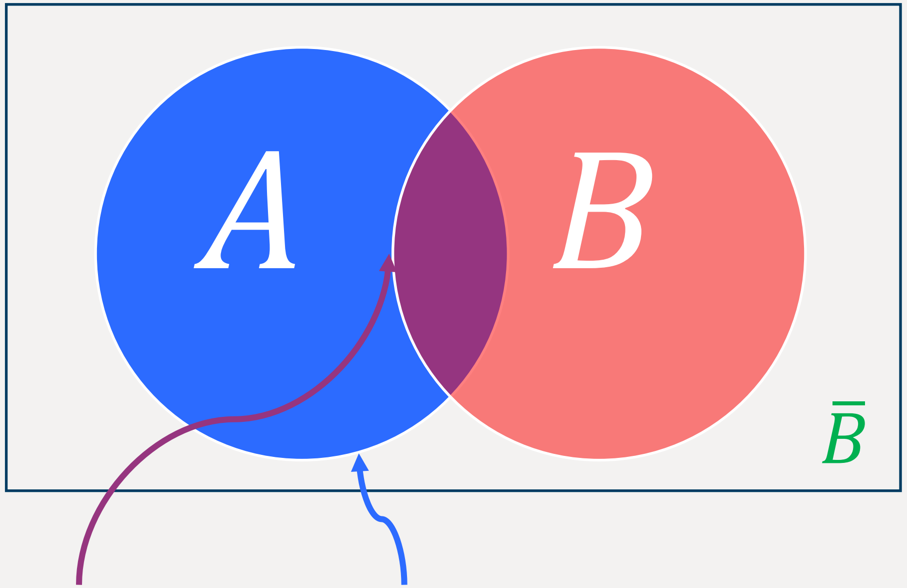


$$A \cap B = B \cap A$$
$$\Pr(A \cap B) = \Pr(B \cap A)$$
$$\Pr(A|B) \cdot \Pr(B) = \Pr(B|A) \cdot \Pr(A)$$

$$\Pr(B|A) = \frac{\Pr(A|B) \cdot \Pr(B)}{\Pr(A)}$$

Duke

"Bayes' Theorem"



$$\Pr(A) = \Pr(A \cap B) + \Pr(A \cap \bar{B})$$
$$= \Pr(A|B) \cdot \Pr(B) + \Pr(A|\bar{B}) \cdot \Pr(\bar{B})$$

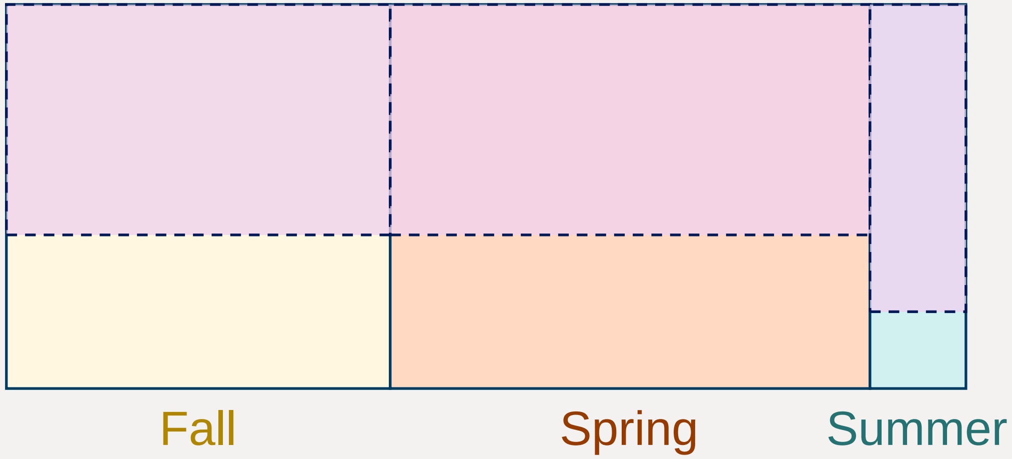
$$\Pr(B|A) = \frac{\Pr(A|B) \cdot \Pr(B)}{\Pr(A|B) \cdot \Pr(B) + \Pr(A|\bar{B}) \cdot \Pr(\bar{B})}$$

Duke

WOTO: NOT Bayes' Theorem

- For a general Duke CS student, there is a 60% probability to fall in love with CS230 after taking it in a regular semester. This probability is 80% for taking it in the summer.
  - 40% of those who take CS230 take it in the fall
  - 50% of those who take CS230 take it in the spring
  - The rest 10% take it in the summer
- Given a student fell in love with CS230 after taking it, what is the probability that they took it in the spring?

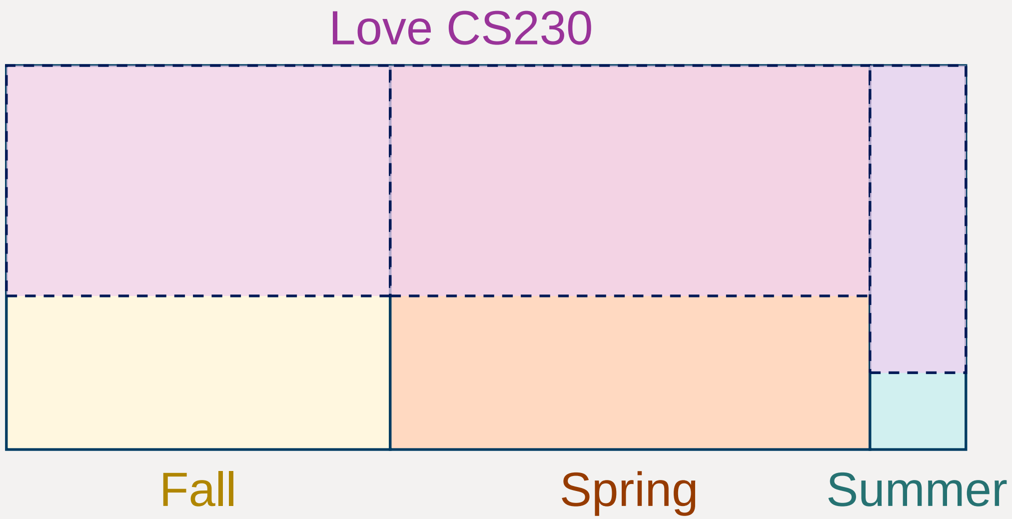
WOTO: NOT Bayes' Theorem



$$\Pr(L)$$
$$= \Pr(L \cap Fa) + \Pr(L \cap Sp) + \Pr(L \cap Su)$$
$$= \Pr(L|Fa) \cdot \Pr(Fa) + \Pr(L|Sp) \cdot \Pr(Sp)$$
$$+ \Pr(L|Su) \cdot \Pr(Su)$$
$$= 0.6 \times 0.4 + 0.6 \times 0.5 + 0.8 \times 0.1$$
$$= 0.62$$

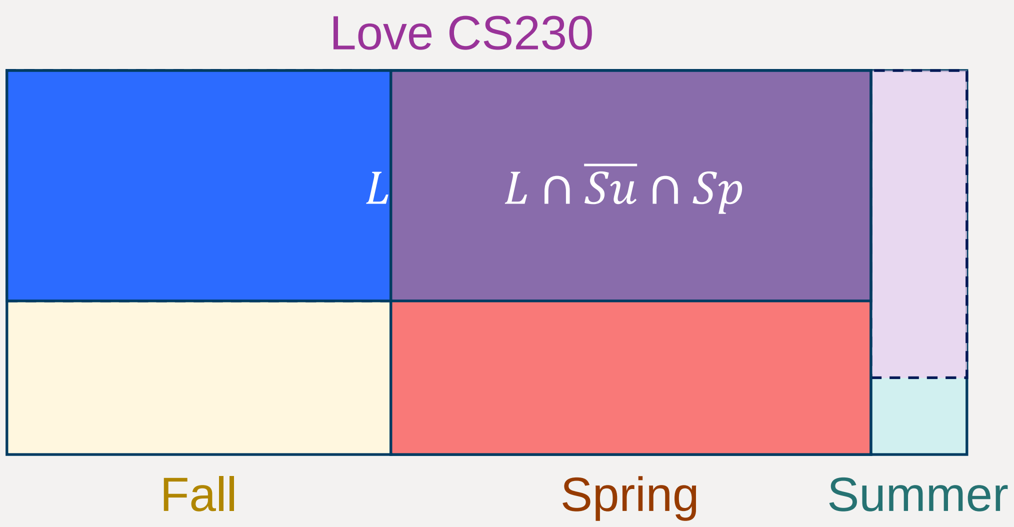
$$\Pr(Sp|L) = \frac{\Pr(Sp \cap L)}{\Pr(L)} = \frac{0.6 \times 0.5}{0.62} \approx 48.4\%$$

PI: Independence and Mutual Exclusivity



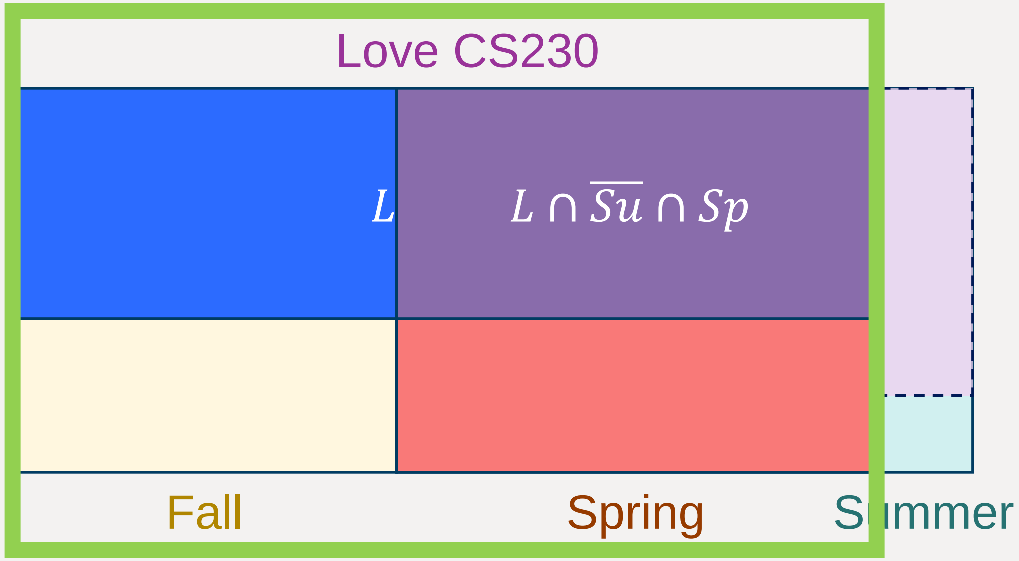
- Events $E_1, E_2, \ldots, E_k$ are mutually exclusive if and only if for every pair $i \neq j$ we have $\Pr(E_i \cap E_j) = 0$
- Events A, B are independent if and only if $\Pr(A \cap B) = \Pr(A) \times \Pr(B)$

PI: Independence and Mutual Exclusivity



- Events A, B are independent if and only if $\Pr(A \cap B) = \Pr(A) \times \Pr(B)$
  - For the third option $A = L \cap \overline{Su}$, $B = Sp$
  - $\Pr(A) = \Pr(L \cap \overline{Su}) = 0.6 \times 0.9 = 0.54$
  - $\Pr(B) = \Pr(Sp) = 0.5$
  - $\Pr(A \cap B) = \Pr(L \cap \overline{Su} \cap Sp)$
    $= 0.6 \times 0.5 = 0.3$
  - $\Pr(A) \times \Pr(B) = 0.54 \times 0.5 = 0.27 \neq 0.3$
- A lot of probability is counterintuitive

PI: Independence and Mutual Exclusivity

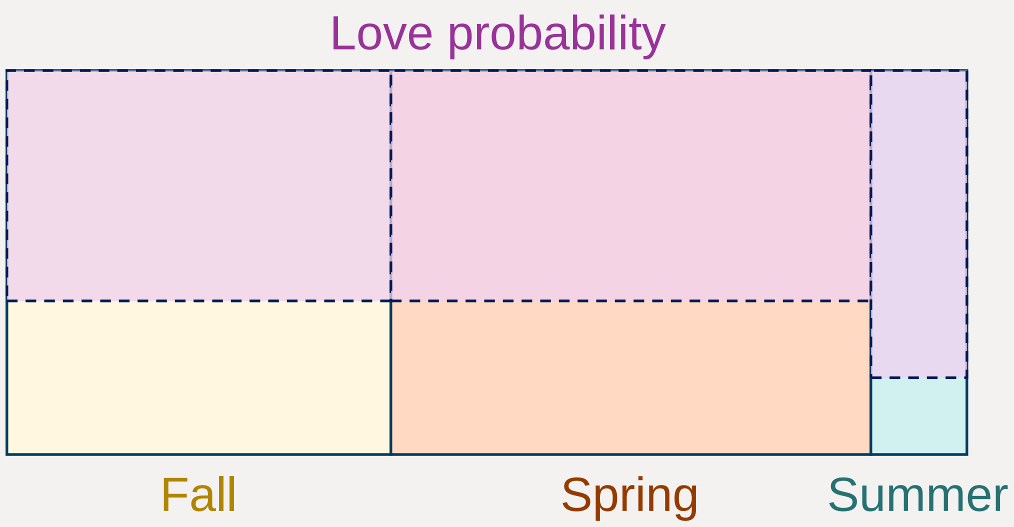


- The "intuition" should be captured by "conditional independence":
- Given that the student took CS230 in a regular semester, the probability that the student loves the class is independent with the semester

$$\Pr((L \cap Sp)|\overline{Su}) = \Pr(L|\overline{Su}) \times \Pr(Sp|\overline{Su})$$

Duke

PI: Independence and Mutual Exclusivity



- Events A, B are independent if and only if $\Pr(A \cap B) = \Pr(A) \times \Pr(B)$
  - If at least one of $\Pr(A)$, $\Pr(B)$ is zero, this is trivially satisfied as both sides are zero
  - If neither is zero, then we also have

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

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

Why define zero-probability events/outcomes?

- Remember that events and outcomes are artificial
- Even for finite discrete sample spaces Ω, we would sometimes still define zero-probability outcomes for simplicity/convenience
- Even if all outcomes have positive probability, there's always the empty event $\emptyset \subseteq 2^{\Omega}$, and

$$\Pr(\emptyset) = \sum_{\substack{\omega \in \Omega \\ \omega \in \emptyset}} \Pr(\omega) = 0$$

Class starts after this song

i don’t understand a single word in this song but it’s fire
---Bruno

AccuseFive – Finally (2021) requested by Bruno Makoto Tanabe de Lima (TA-of-CM8)

I'm a sophomore from Brazil studying CS and Math. I'm a big fan of rock climbing and I like playing Valorant in my free time. Fun fact: back in 9th grade, two friends and I ranked in the top 10 in Latin America for Brawl Stars.

Bounding probability values

- It is not always feasible to obtain precise probability values for complicated events in complicated models
- Sometimes we resort to bounding them (from below or above)
- Fundamental law of probabilities gives us:

$$0 \leq \Pr(A) \leq 1$$

  but we can do more!

Union bound: U for upper

$$\Pr(A_1 \cup A_2 \cup \cdots \cup A_n) \leq \Pr(A_1) + \Pr(A_2) + \cdots + \Pr(A_n)$$

- Intuition: think about inclusion-exclusion in counting: everything on the LHS is counted at least once on the RHS, and some may be overcounted multiple times
- Therefore: the equation holds when there's no overcounting (i.e., all events are mutually exclusive)

Lower bound

$$\min\{\Pr(A_1), \Pr(A_2), \ldots, \Pr(A_n)\} \leq \Pr(A_1 \cup A_2 \cup \cdots \cup A_n)$$

- Intuition: no matter what the minimum is, that set is a subset of the union
- The equation holds when all events are equal (i.e., $A_1 = A_2 = \cdots = A_n$)

Mutual and pairwise independence

- Events $A_1, A_2, \ldots, A_n$ are pairwise independent if and only if for all $i \neq j$,
$$\Pr(A_i \cap A_j) = \Pr(A_i) \times \Pr(A_j)$$
- Events $A_1, A_2, \ldots, A_n$ are mutually independent if and only if for all $S \subseteq \{A_1, A_2, \ldots, A_n\}$,
$$\Pr\left(\bigcap_{A \in S} A\right) = \prod_{A \in S} \Pr(A)$$
$$\Pr(A_1 \cap A_2 \cap \cdots \cap A_n) = \Pr(A_1) \times \Pr(A_2) \times \cdots \times \Pr(A_n)$$

Duke

PI: mutual and pairwise independence

First toss	Second toss	H_1	H_2	S	$H_1 \cap H_2$	$H_1 \cap S$	$H_2 \cap S$	$H_1 \cap H_2 \cap S$
Head	Head	✓	✓	✓	✓	✓	✓	✓
Head	Tail	✓						
Tail	Head		✓					
Tail	Tail			✓				

- H_1, H_2, S are pairwise independent: $\Pr(A_i \cap A_j) = \Pr(A_i) \times \Pr(A_j) = 1/4$
- H_1, H_2, S are NOT mutually independent:
  $$1/4 = \Pr(H_1 \cap H_2 \cap S) \neq \Pr(H_1) \times \Pr(H_2) \times \Pr(S) = 1/8$$

Random variable: neither random nor variable

- They are functions that map outcomes of random experiments to variables(numbers)
- The domain is the sample space
- The codomain is $\mathbb{R}$
- Each outcome is mapped (by the r.v.) to a real number
- When we say $\Pr(X = a)$ we really mean $\Pr(\{\omega \in \Omega | X(\omega) = a\})$

Random variable is also artificial

- Recall: outcomes and thus events are artificially designed (to model the real world)
- Therefore, random variables can also be artificially designed (again to model the real world)
- Designing them is just like designing an algorithm – it’s an art but there are common strategies

Indicator variable

- Indicator variable: $I_A : \Omega \to \{0,1\}$
  - When we care about an event A happening or not
  - $I_A(\omega) = 1$ if and only if $\omega \in A$
  - Use: can easily sum up lots of indicator variables
    to count the number of times A happens over a certain timespan

Indicator variable lets us be concise

Stimulus

Experiment 3

Let's still consider the card decks with four cards: $\{D, \mu, k, \varepsilon\}$. (We now know our deck is not faulty.) We will now draw **three** cards **with replacement**, with reshuffling in-between draws, so **every draw is uniformly at random** and **independent to every other draw** (i.e., the three draws are **mutually independent**).

Let L be the **random variable** denoting the **number of times we draw a letter card** (D or k).
Furthermore, let L_1, L_2, and L_3 be the **indicator variables** denoting whether or not the first, second, and third card we draw is a letter card, respectively. (In other words, $L_1 = 1$ if the first card is a letter, and $L_1 = 0$ if it is not a letter; same for the other two variables.)
Therefore, we have $L = L_1 + L_2 + L_3$.

1 Numeric 1 point

What is $\Pr(L = 0)$? (Answer with a decimal number.)

Without indicator variables, we would need to repeat "what is the probability of not drawing any letter card in any of the three draws" and "what is the probability of drawing exactly two letter cards in the three draws"

2

What is $\Pr(L = 2)$? (Answer with a decimal number.)
Hint: binomial distribution.

0.375

Level of formality (CM2) again

Excessive Symbols No words

Formal

- When to use a random variable/indicator variable is a subjective and stylistic choice
- For very straightforward tasks, using them might be an overkill
- For complicated problems, not using them might lead to wordy and inprecise communication
- Your call ultimately, but you need to at least acquire the tool

All words, no symbols

Proof by picture

Informal

Duke

Series of games

- Two sports team (say A and B) play a best-of-5 series.
  - In other words, they keep playing until one team accumulates 3 wins, at which time they are declared the winner of the series.
- Assumption: for every game independently, each team has a probability of $\frac{1}{2} = 50\%$ to win the game.
- What is the probability that team A wins the series?

Approach 1: “The intuition”

- “There is nothing indicating either team A or team B to be better than the other. Therefore, the probability that team A wins the series must be 50%.”

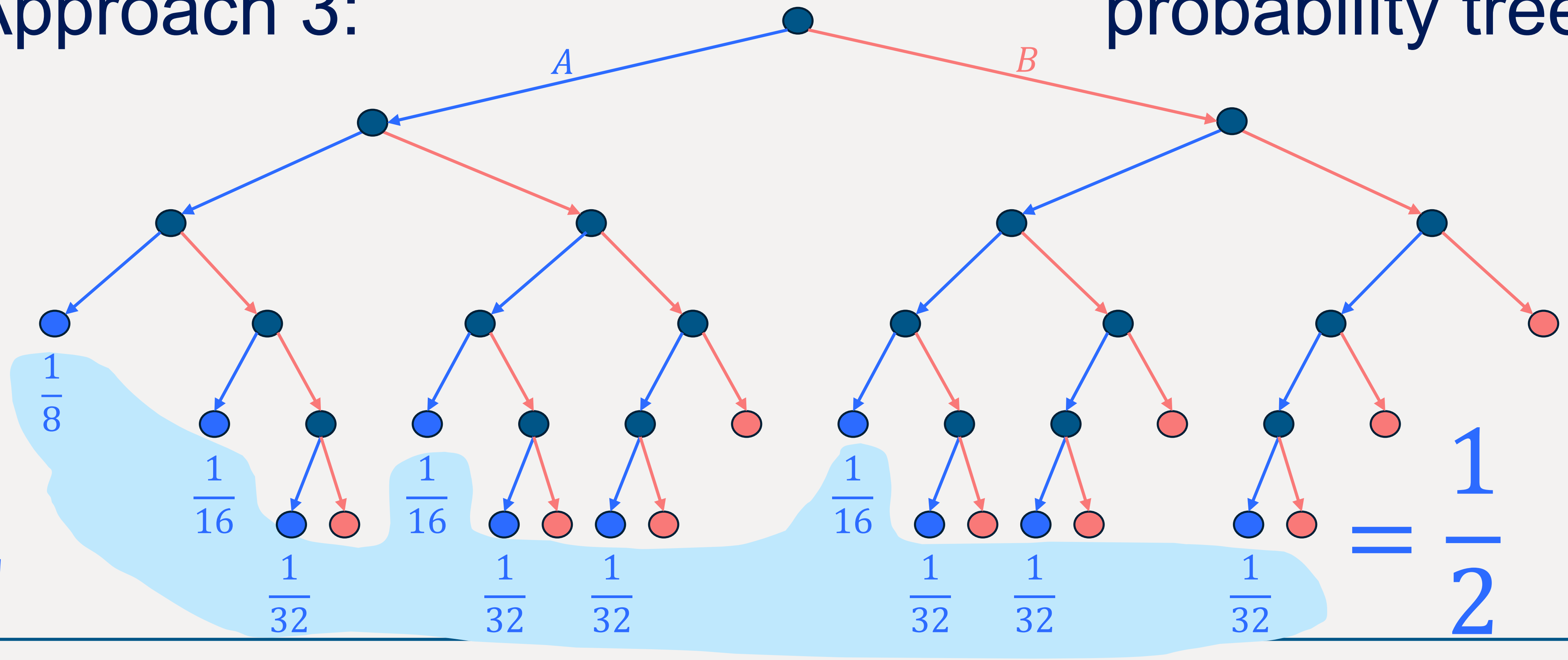
Approach 2: “counting equal chance outcomes”

- Let’s count the number of all outcomes as well as the number of outcomes in which team A wins the series:

$$AAA, AABA, ABAA, BAAA, AABBA, ABABA, BAABA, ABBAA, BABAA, BBAAA$$
$$BBB, BBAB, BABB, ABBB, BBAAB, BABAB, ABBAB, BAABB, ABABB, AABBB$$

- Therefore $\Pr(A \text{ wins the series}) = \frac{10}{20} = 50\%$

Approach 3: "probability tree"



$$\Sigma = \frac{1}{2}$$

Duke

Approach 4: Binomial distribution

- Each game is a Bernoulli trial with “success probability” 50%
- “Imagine” playing out all 5 games (including keeping playing after some team got 3 wins)
- Winning the series = winning at least 3 games out of 5
- $\sum_{i=3}^{5}\binom{5}{i}\left(\frac{1}{2}\right)^i\left(1-\frac{1}{2}\right)^{5-i} = \frac{\binom{5}{3}+\binom{5}{4}+\binom{5}{5}}{2^5} = \frac{\binom{5}{0}+\binom{5}{1}+\binom{5}{2}+\binom{5}{3}+\binom{5}{4}+\binom{5}{5}}{2\times 2^5} = \frac{2^5}{2^6} = \frac{1}{2}$

Approach 1: "The intuition" doesn't generalize

- "There is nothing indicating either team A or team B to be better than the other. Therefore, the probability that team A wins the series must be 50%."
- Very nice, and generalizes to best-of-7, best-of-9, …
- But what if team A is better than team B?
  Such that team A would win each game with probability 60%?
- Intuition only takes us so far

Duke

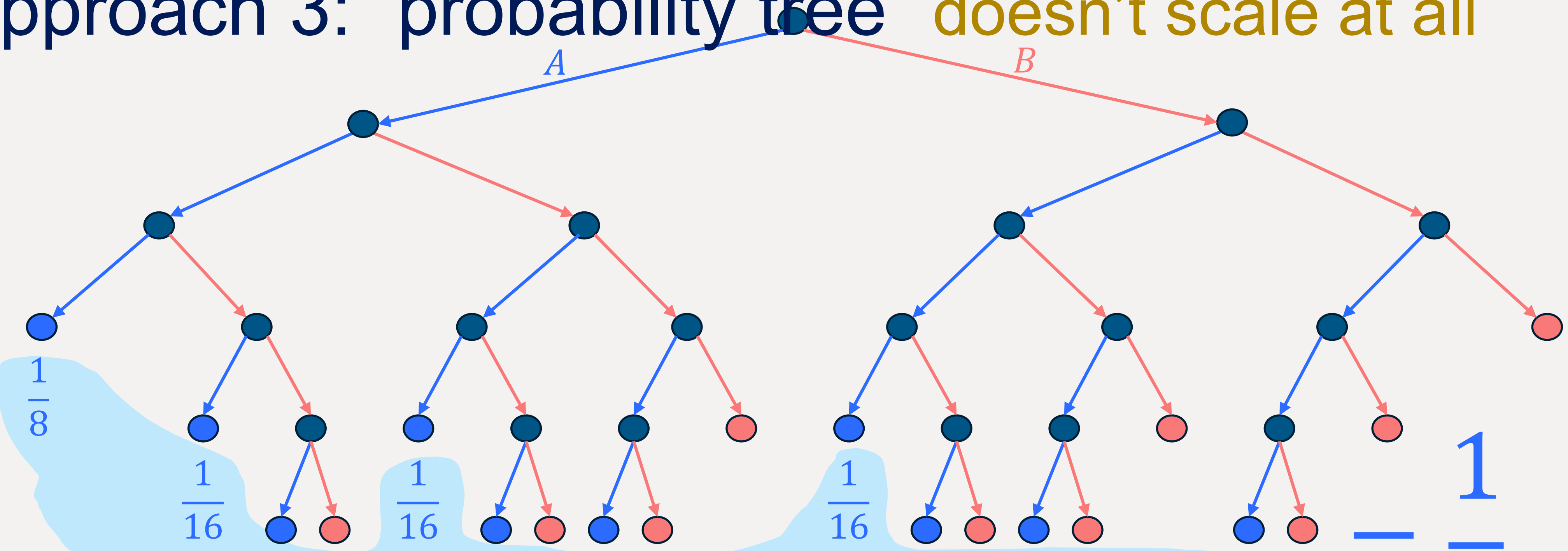
~~Approach 2: “counting equal chance outcomes”~~

- In case you didn’t notice, this was wrong!

$AAA, AABA, ABAA, BAAA, AABBA, ABABA, BAABA, ABBAA, BABAA, BBAAA$
$BBB, BBAB, BABB, ABBB, BBAAB, BABAB, ABBAB, BAABB, ABABB, AABBB$

- The outcomes did not have equal probability
- We lucked to the right answer simply because of symmetry, so of course this doesn’t generalize at all

Approach 3: "probability tree" doesn't scale at all



$$\sum \frac{1}{8} + \frac{1}{16} + \frac{1}{32} + \frac{1}{16} + \frac{1}{32} + \frac{1}{32} + \frac{1}{16} + \frac{1}{32} + \frac{1}{32} + \frac{1}{32} = \frac{1}{2}$$

Duke

Approach 4: Binomial distribution generalizes

- Each game is a Bernoulli trial with "success probability" **60%**
- "Imagine" playing out all **9** games (including keeping playing after some team got **5** wins)
- Winning the series = winning at least **5** games out of **9**
- $\sum_{i=5}^{9} \binom{9}{i} (0.6)^i (1 - 0.6)^{9-i}$
  (doesn't simply as nicely but still analytically simple)

WOTO: More complicated series

- Imagine a best-of-7 series where team A is the better team and will win each game independently with probability 75%
- But… they are down 1-2 after the first 3 games to team B
- Given this already happened, what is the probability that team A comes back and win the series?

WOTO: More complicated series

- Given team A down 1-2 after the first 3 games to team B, what is the probability they come back and win the series?
- Disregard the first 3 games that already happened. Team A now just needs to win 3 out of the rest 4 (the "imaginary extra game" still applies):
- $\sum_{i=3}^{4} \binom{4}{i} \left(\frac{3}{4}\right)^{i} \left(1 - \frac{3}{4}\right)^{4-i} = \frac{189}{256} \approx 73.8\%$

WOTO: More complicated series

- Given team A won the series (without other information), what is the probability that they were down 1-2 after the first 3 games to team B?
- $\frac{\sum_{i=3}^{4}\binom{4}{i}\left(\frac{3}{4}\right)^{i}\left(1-\frac{3}{4}\right)^{4-i}}{\sum_{j=4}^{7}\binom{7}{j}\left(\frac{3}{4}\right)^{j}\left(1-\frac{3}{4}\right)^{4-j}}$
- Wrong! What is wrong with it?

WOTO: More complicated series

- Given team A won the series (without other information), what is the probability that they were down 1-2 after the first 3 games to team B?

- $$\frac{\sum_{i=3}^{4}\binom{4}{i}\left(\frac{3}{4}\right)^{i}\left(1-\frac{3}{4}\right)^{4-i}}{\sum_{j=4}^{7}\binom{7}{j}\left(\frac{3}{4}\right)^{j}\left(1-\frac{3}{4}\right)^{4-j}}$$

This is $\Pr(A \text{ wins the series} \mid A \text{ down } 1-2 \text{ after } 3 \text{ games})$

What we should've put as the denominator is $\Pr(A \text{ wins the series} \cap A \text{ down } 1-2 \text{ after } 3 \text{ games})$, so we need to multiply the denominator by $\binom{3}{1}\left(\frac{3}{4}\right)^{1}\left(1-\frac{3}{4}\right)^{3-1}$

Joint distributions

Random variable $X, Y: \Omega \to \mathbb{R} \times \mathbb{R}$

$$f_{X,Y}(a, b) = \Pr((X = a) \land (Y = b))$$

Random variable $X: \Omega \to \mathbb{R}$

$$f_X(a) = \Pr(X = a)$$

Random variable $Y: \Omega \to \mathbb{R}$

$$f_Y(b) = \Pr(Y = b)$$

X, Y are independent random variables if and only if $f_{X,Y}(a, b) = f_X(a) f_Y(b)$ for all $a, b \in \mathbb{R} \times \mathbb{R}$

Marginalization

Random variable $X, Y: \Omega \to \mathbb{R} \times \mathbb{R}$

$$f_{X,Y}(a, b) = \Pr((X = a) \wedge (Y = b))$$

Random variable $X: \Omega \to \mathbb{R}$

$$f_X(a) = \Pr(X = a)$$

$$= \sum_b \Pr((X = a) \wedge (Y = b)) = \sum_b f_{X,Y}(a, b)$$

Random variable $Y: \Omega \to \mathbb{R}$

$$f_Y(b) = \Pr(Y = b)$$

$$= \sum_a \Pr((X = a) \wedge (Y = b)) = \sum_a f_{X,Y}(a, b)$$

Duke

Joint distributions, you all as an example

- The (empirical) probability distribution that 101 students used consulting hours and/or Ed discussions
- $\Omega = \{(\text{Yes CH}, \text{Yes Ed}), (\text{No CH}, \text{Yes Ed}), (\text{Yes CH}, \text{No Ed}), (\text{No CH}, \text{No Ed})\}$
- X indicator variable of using CH
- Y indicator variable of using Ed

Y \ X	**0**	**1**	f_Y
0	0.25	0.15	$\mathbf{0.4}$
1	0.25	0.35	$\mathbf{0.6}$
f_X	$\mathbf{0.5}$	$\mathbf{0.5}$	**1**

Confusing joint distributions: Duke players

- Let's now look at two Duke men's basketball alumni: AJ Griffin and Wendell Moore Jr. (2021-22 season)
- $\Omega = \{(2P, \text{Hit}), (3P, \text{Hit}), (2P, \text{Miss}), (3P, \text{Miss})\}$
- X random variable of point value of the shot
- Y indicator variable of making the shot
- For ease of understanding let us take the frequentist perspective and look at the counts

AJ Griffin

$Y \backslash X$	2	3	f_Y
0	62	88	150
1	75	71	146
f_X	137	159	296

Wendell Moore Jr.

$Y \backslash X$	2	3	f_Y
0	114	74	188
1	136	52	188
f_X	250	126	376

Confusing joint distributions: Duke players

- Let's now normalize/marginalize counts **by row** (since we care about the shooting percentages)
- In other words, we want to know $\Pr(Y = 1|X = 2)$ and $\Pr(Y = 1|X = 3)$ for both players
- After this marginalization, it looks like AJ Griffin was the better shooter than Wendell Moore Jr. since AJ had the higher 2P% and also 3P%
- But what if we look at the overall percentage?
- For further reading see Simpson's paradox

AJ Griffin

Y \ X	2	3
0	62/137	88/159
1	75/137	71/159
$f_{Y=1\|X=x}$	**54.7%**	**44.7%**

Wendell Moore Jr.

Y \ X	2	3
0	114/250	74/126
1	136/250	52/126
$f_{Y=1\|X=x}$	54.4%	35.8%