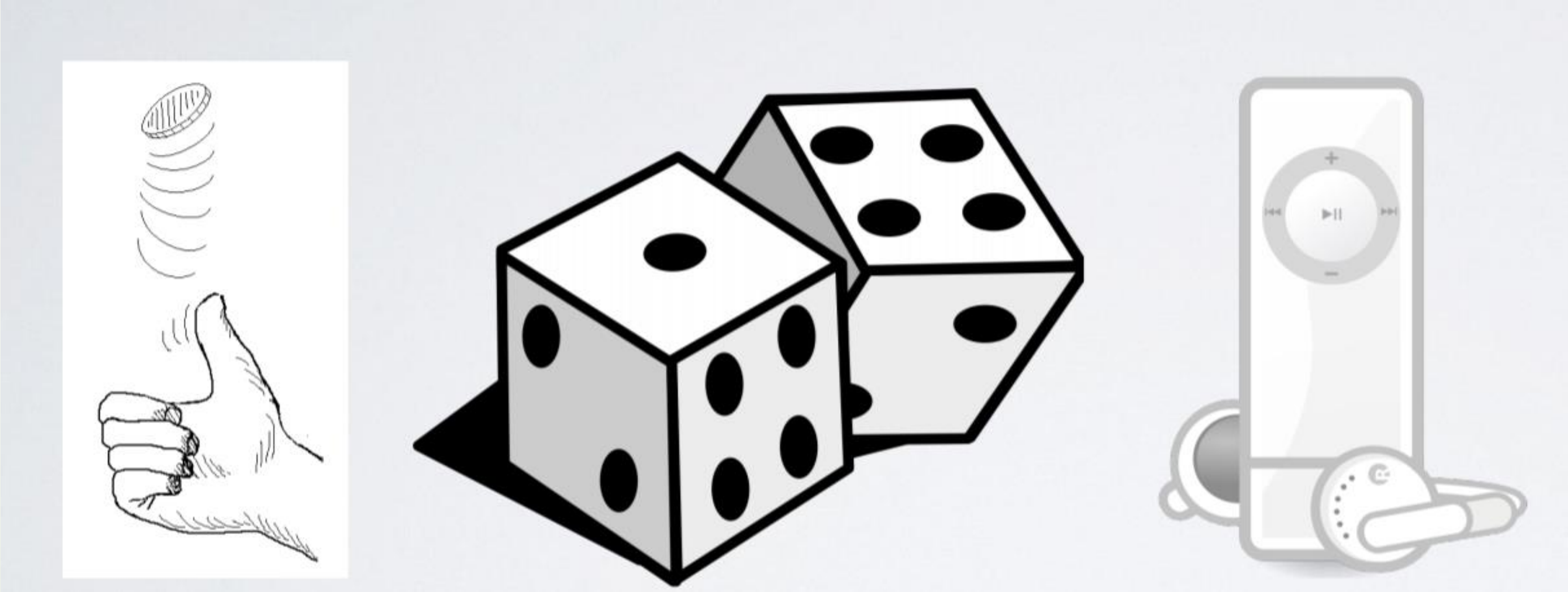
Probability and distrubution

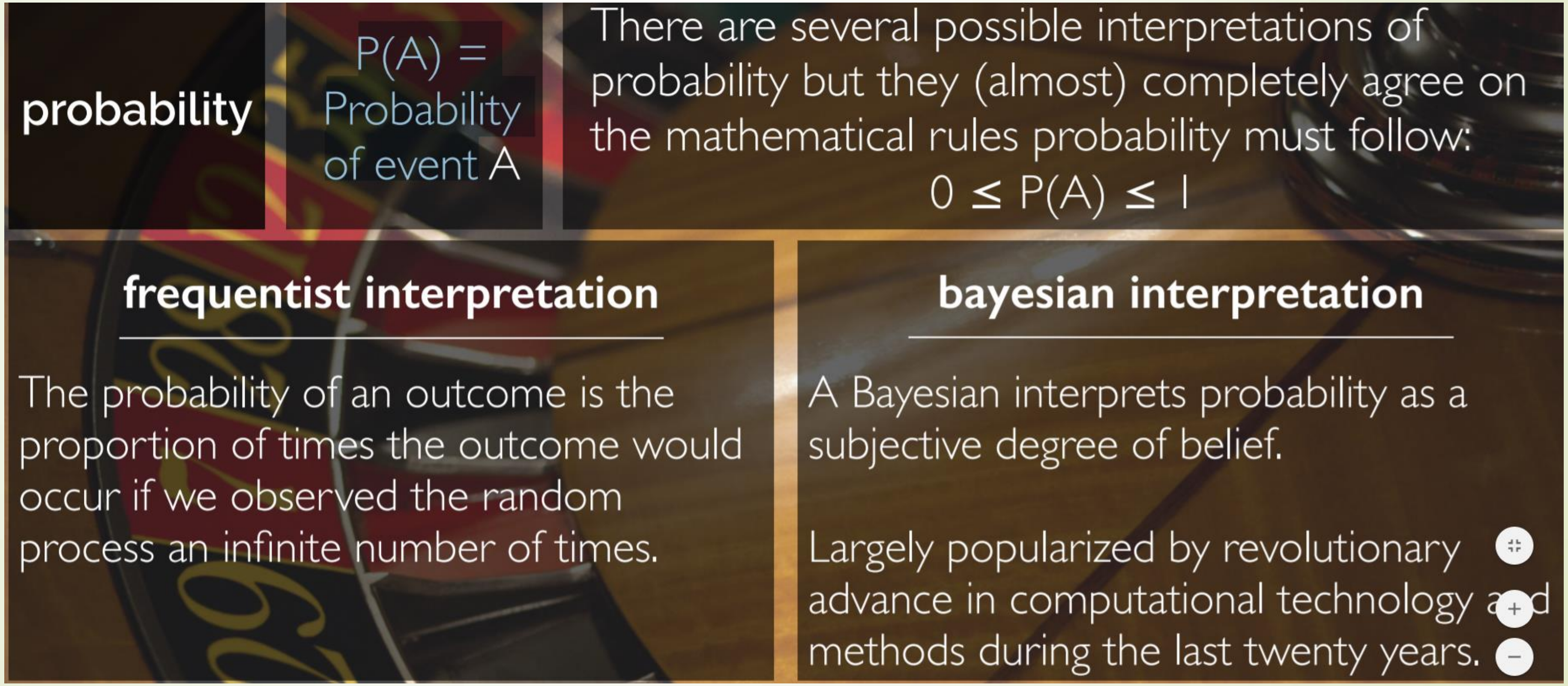
INTERNAL

Random process

- In a random process we know what outcomes could happen, but we don't know which particular outcome will happen.



Probability



probability

P(A) = Probability of event A

There are several possible interpretations of probability but they (almost) completely agree on the mathematical rules probability must follow:

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

frequentist interpretation

The probability of an outcome is the proportion of times the outcome would occur if we observed the random process an infinite number of times.

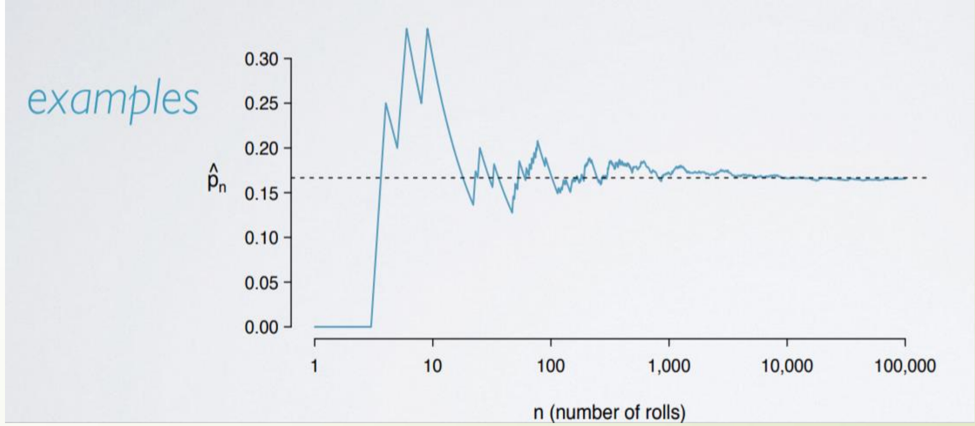
bayesian interpretation

A Bayesian interprets probability as a subjective degree of belief.

Largely popularized by revolutionary advance in computational technology and methods during the last twenty years.

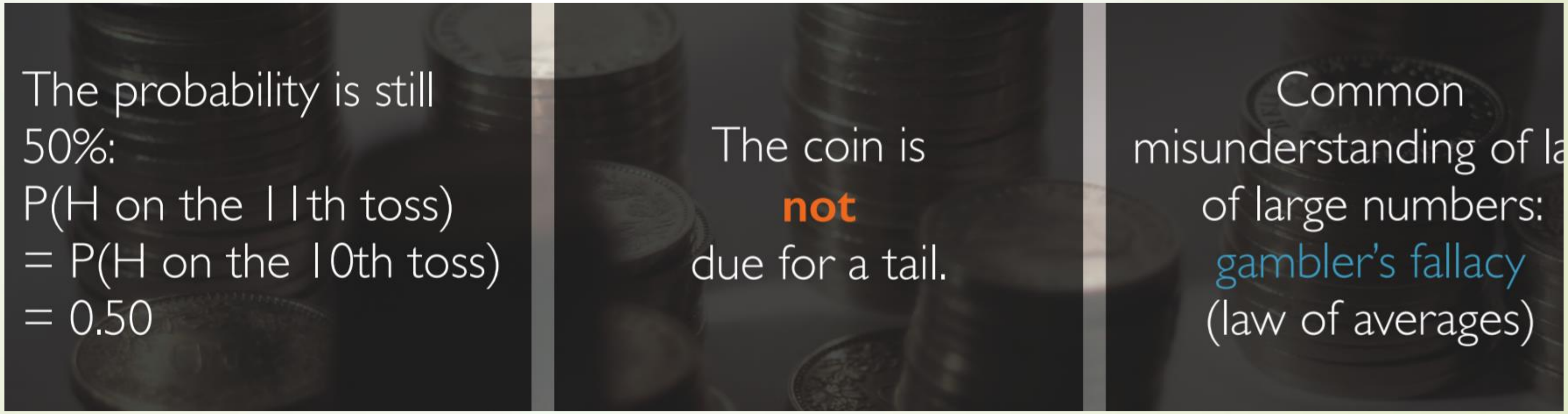
law of large numbers

- law of large numbers states that as more observations are collected, the proportion of occurrences with a particular outcome converges to the probability of that outcome.
- exactly 3 heads in 10 coin flips –
- exactly 3 heads in 100 coin flips –
- exactly 3 heads in 1000 coin flips



Example

- Say you toss a coin 10 times, and it lands on Heads each time. What do you think the chance is that another head will come up on the next toss? 0.5, less than 0.5, or more than 0.5?
- H H H H H H H H H H ?



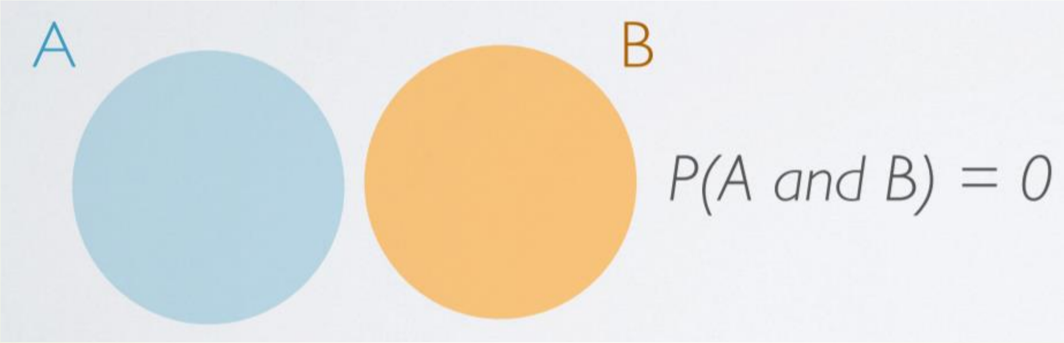
The probability is still 50%:
P(H on the 11th toss)
= P(H on the 10th toss)
= 0.50

The coin is
not
due for a tail.

Common
misunderstanding of la
of large numbers:
gambler's fallacy
(law of averages)

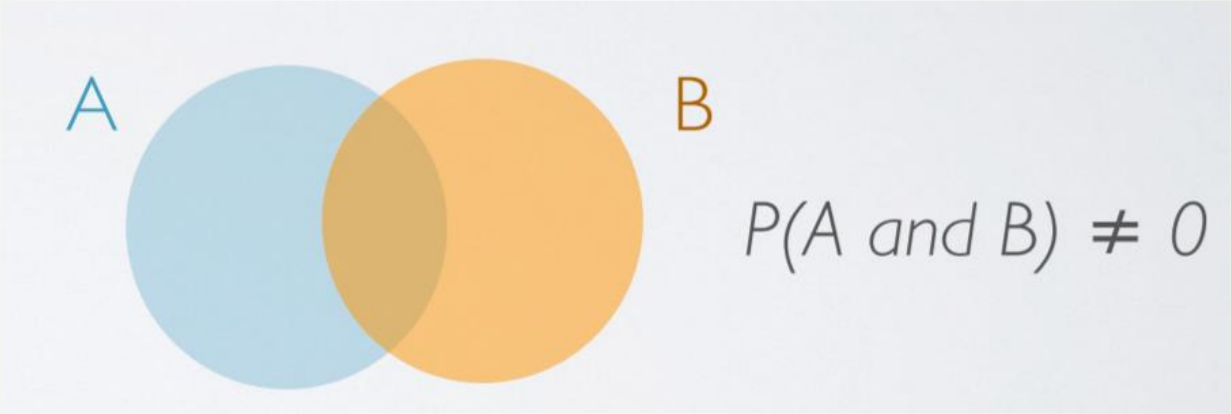
Disjoint events

- disjoint (mutually exclusive) events cannot happen at the same time.
- the outcome of a single coin toss cannot be a head and a tail.
- a student can't both fail and pass a class.
- a single card drawn from a deck cannot be an ace and a queen.



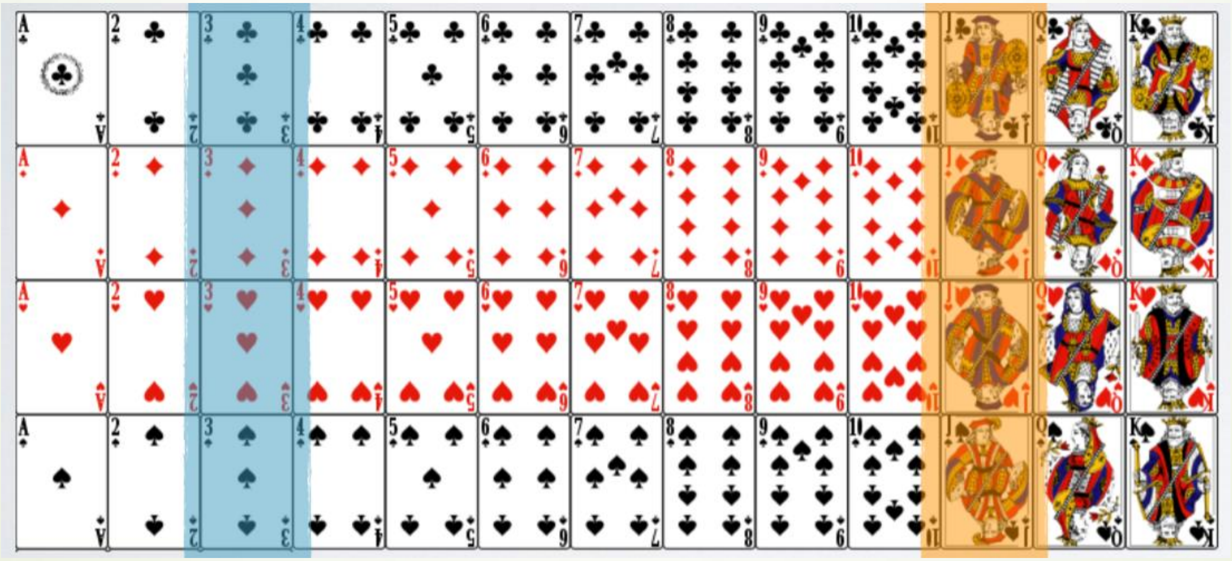
Non-Disjoint events

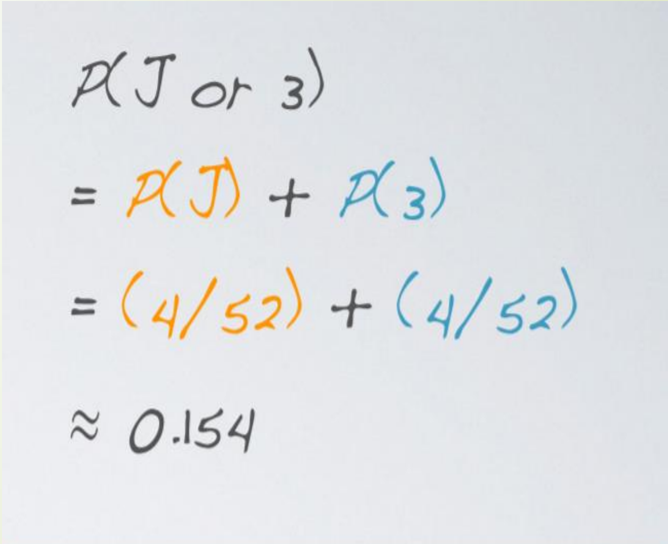
- non-disjoint events can happen at the same time.
- a student can get an A in Stats and A in Econ in the same semester.



union of disjoint events

- What is the probability of drawing a Jack or a three from a well shuffled full deck of cards?





$P(J \text{ or } 3)$

$= P(J) + P(3)$

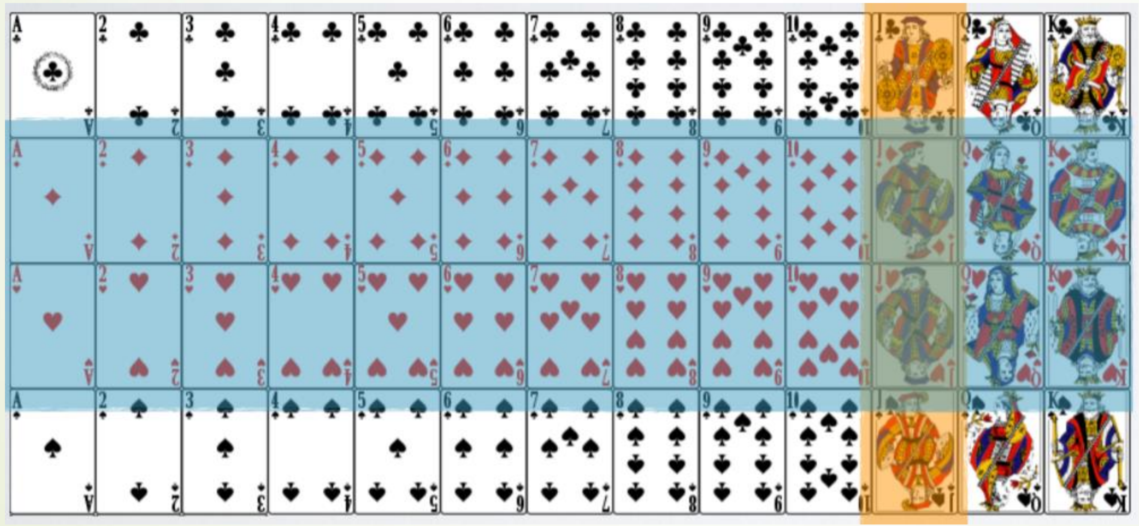
$= (4/52) + (4/52)$

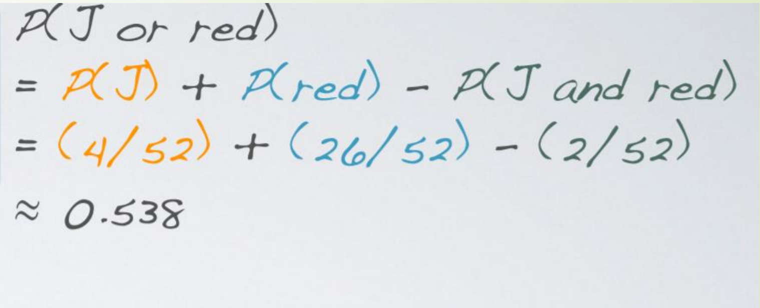
≈ 0.154

For disjoint events A and B,
P(A or B) = P(A) + P(B)

union of non-disjoint events

- What is the probability of drawing a Jack or a red card from a well shuffled full deck of cards?

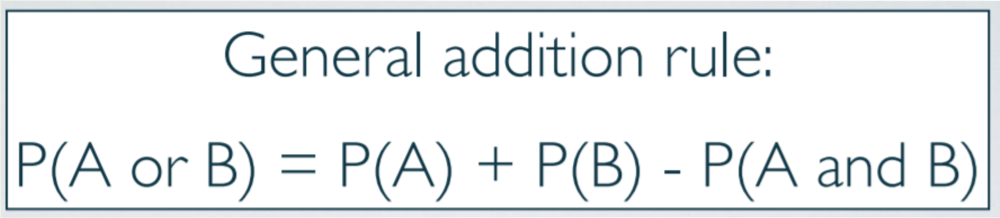




$$
\begin{aligned}
&P(J \text{ or } red) \\
&= P(J) + P(red) - P(J \text{ and } red) \\
&= (4/52) + (26/52) - (2/52) \\
&\approx 0.538
\end{aligned}
$$

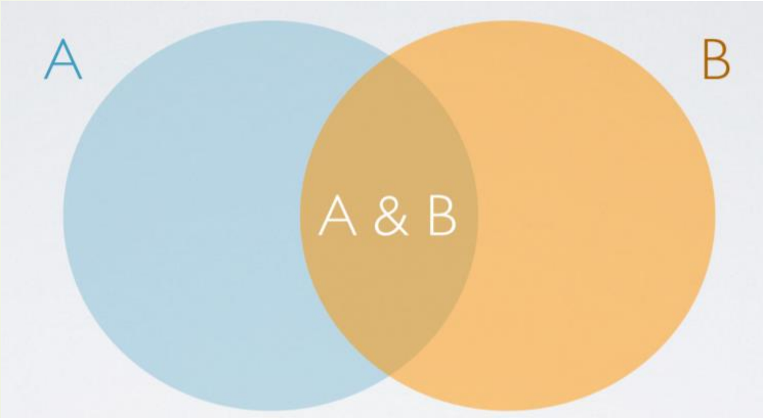
For non-disjoint events A and B,
P(A or B) = P(A) + P(B) - P(A and B)

General Addition rule



General addition rule:

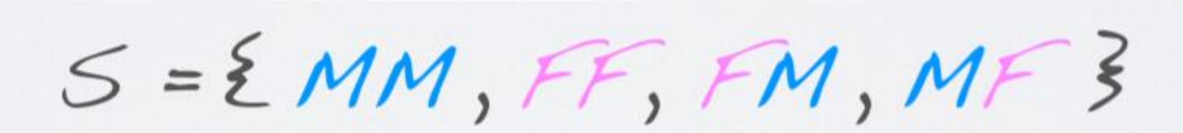
P(A or B) = P(A) + P(B) - P(A and B)



Note: When A and B are disjoint, P(A and B) = 0, so the formula simplifies to P(A or B) = P(A) + P(B).

Sample space

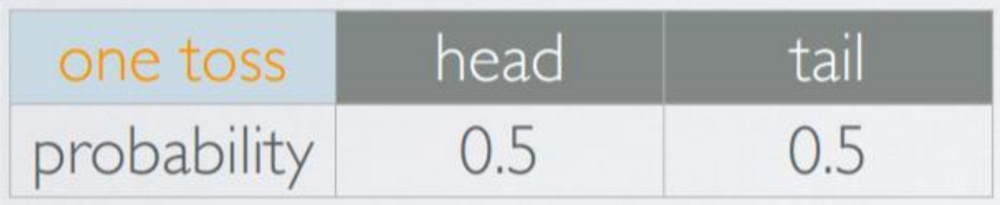
- a sample space is a collection of all possible outcomes of a trial.
- A couple has two kids, what is the sample space for the sex of these kids? For simplicity assume that sex can only be male or female?



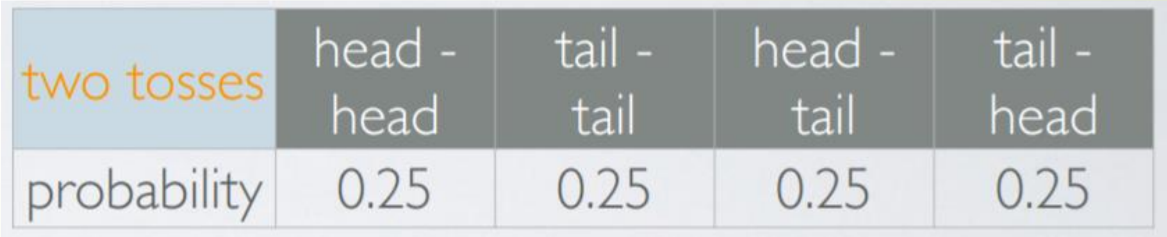
$S = \{ MM, FF, FM, MF \}$

Probability distributions

- rules 1. the events listed must be disjoint 2. each probability must be between 0 and 1 3. the probabilities must total 1



one toss	head	tail
probability	0.5	0.5

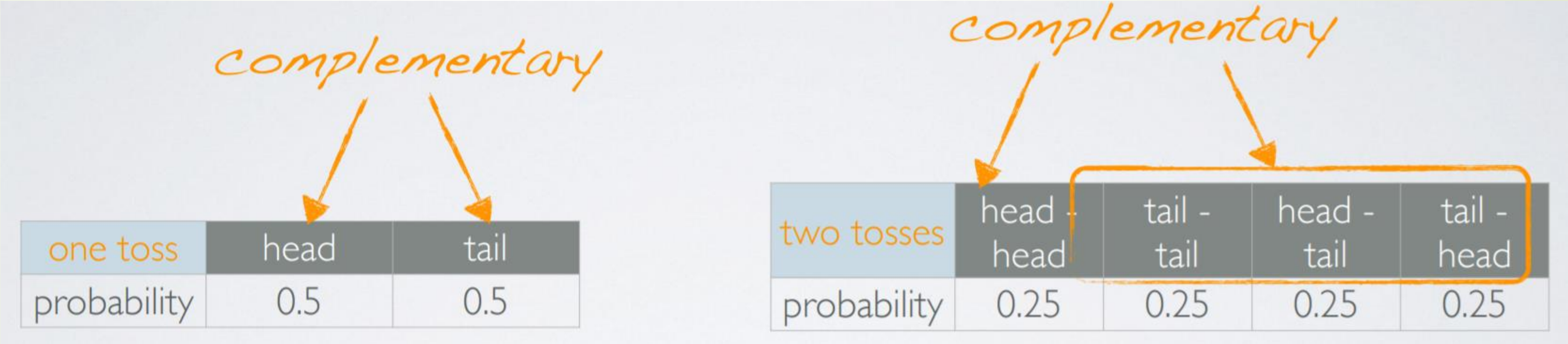


two tosses	head - head	tail - tail	head - tail	tail - head
probability	0.25	0.25	0.25	0.25

Rules:
1. the events listed must be disjoint
2. each probability must be between 0 and 1
3. the probabilities must total 1

complementary events

- complementary events are two mutually exclusive events whose probabilities add up to 1.



complementary

one toss	head	tail
probability	0.5	0.5

complementary

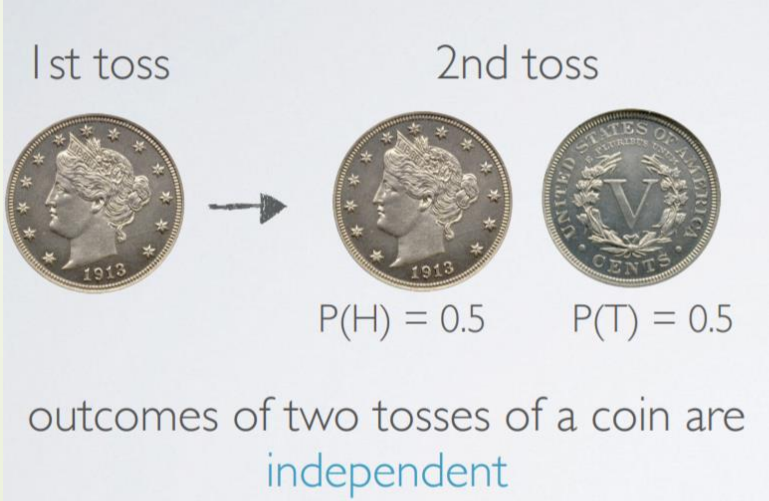
two tosses	head - head	tail - tail	head - tail	tail - head
probability	0.25	0.25	0.25	0.25

disjoint vs. complementary

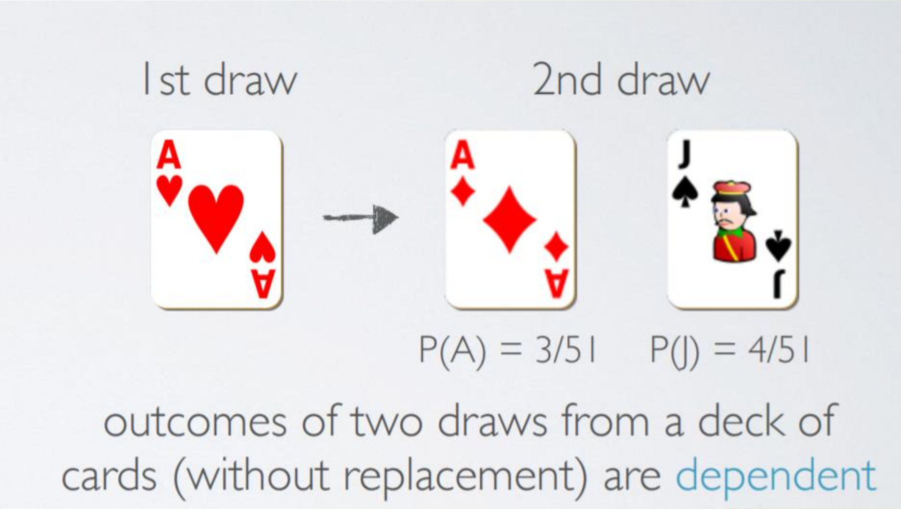
- Do the sum of probabilities of two disjoint outcomes always add up to 1?
- Not necessarily, there may be more than 2 outcomes in the sample space.
- Do the sum of probabilities of two complementary outcomes always add up to 1?
- Yes, that's the definition of complementary

independence

- two processes are independent if knowing the outcome of one provides no useful information about the outcome of the other.

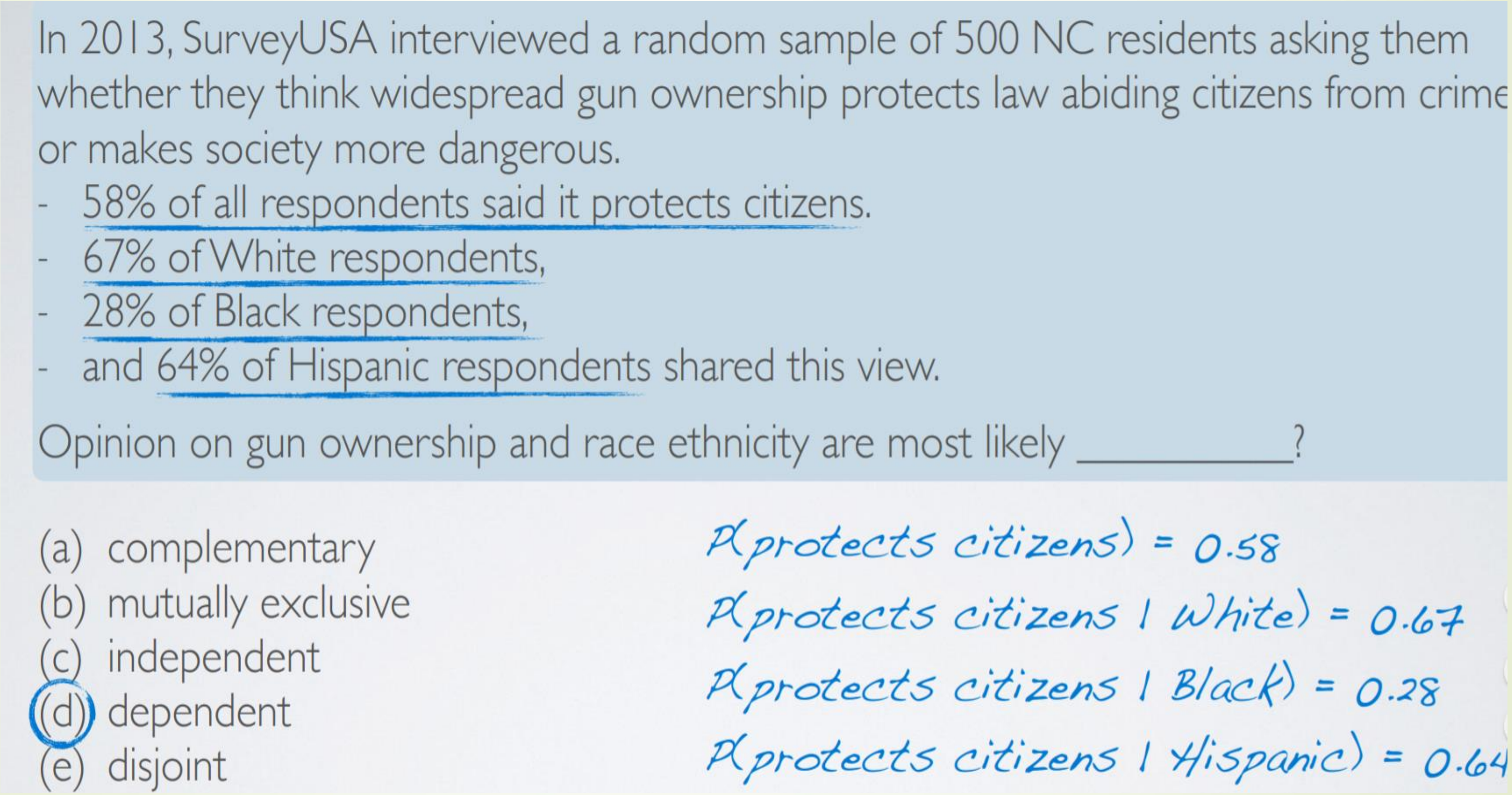


outcomes of two tosses of a coin are independent



outcomes of two draws from a deck of cards (without replacement) are dependent

Example



In 2013, SurveyUSA interviewed a random sample of 500 NC residents asking them whether they think widespread gun ownership protects law abiding citizens from crime or makes society more dangerous.

- 58% of all respondents said it protects citizens.
- 67% of White respondents,
- 28% of Black respondents,
- and 64% of Hispanic respondents shared this view.

Opinion on gun ownership and race ethnicity are most likely __________?

(a) complementary
(b) mutually exclusive
(c) independent
[✓ (d) dependent]
(e) disjoint

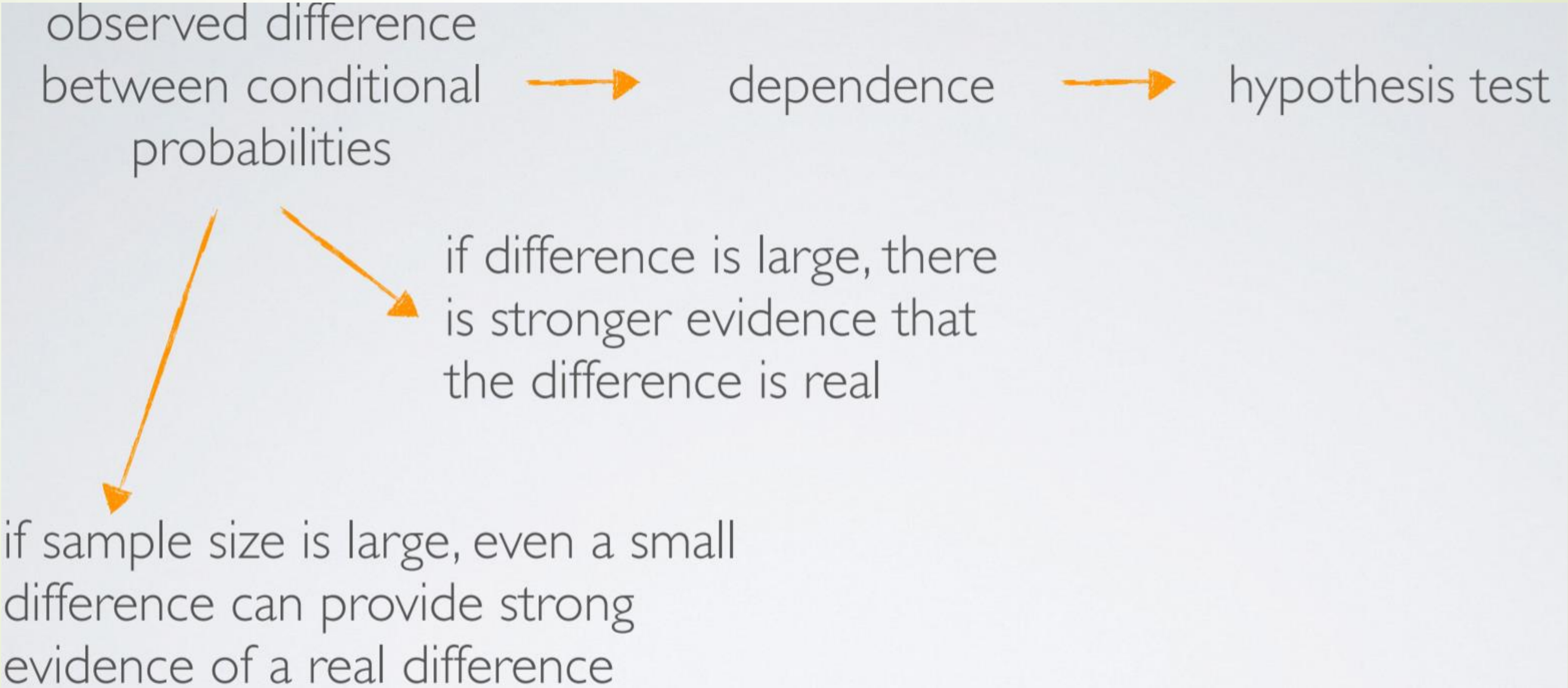
P(protects citizens) = 0.58

P(protects citizens | White) = 0.67

P(protects citizens | Black) = 0.28

P(protects citizens | Hispanic) = 0.64

determining dependence based on sample data

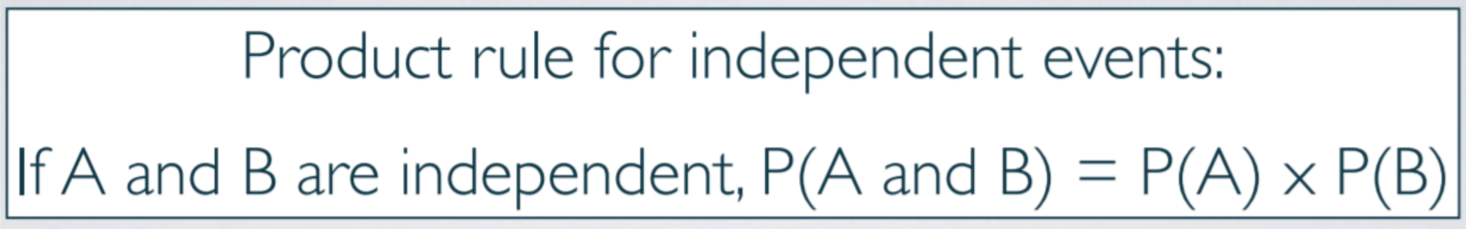


observed difference between conditional probabilities → dependence → hypothesis test

if difference is large, there is stronger evidence that the difference is real

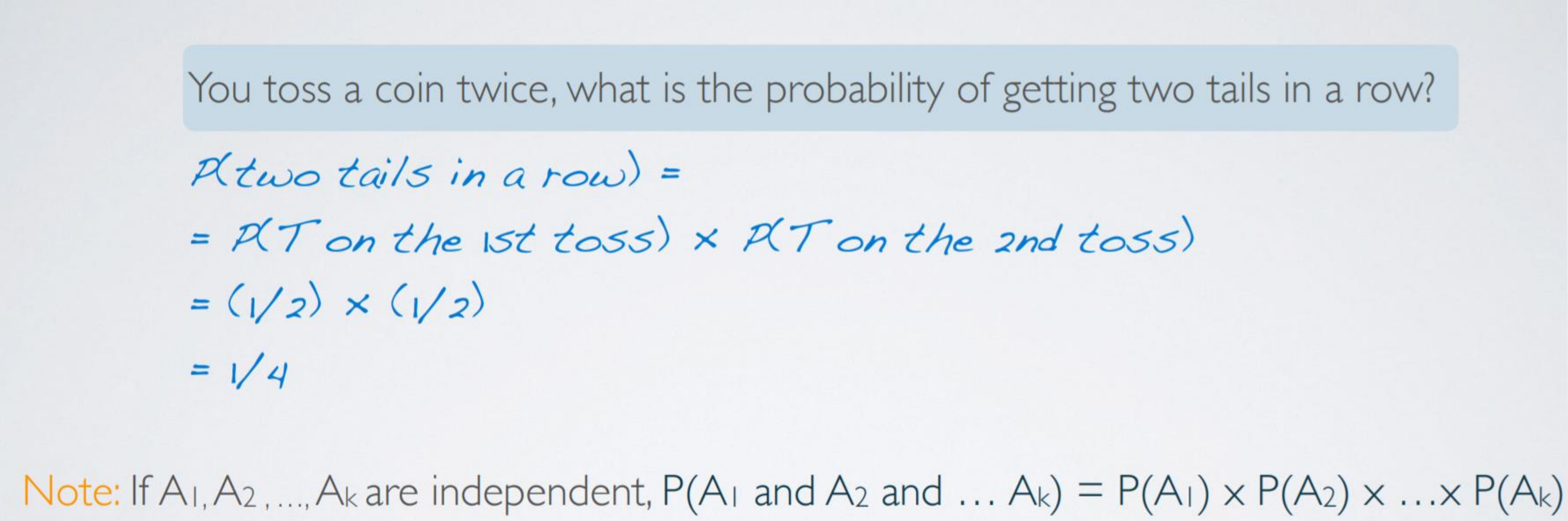
if sample size is large, even a small difference can provide strong evidence of a real difference

Product rule



Product rule for independent events:

If A and B are independent, P(A and B) = P(A) x P(B)



You toss a coin twice, what is the probability of getting two tails in a row?

P(two tails in a row) =

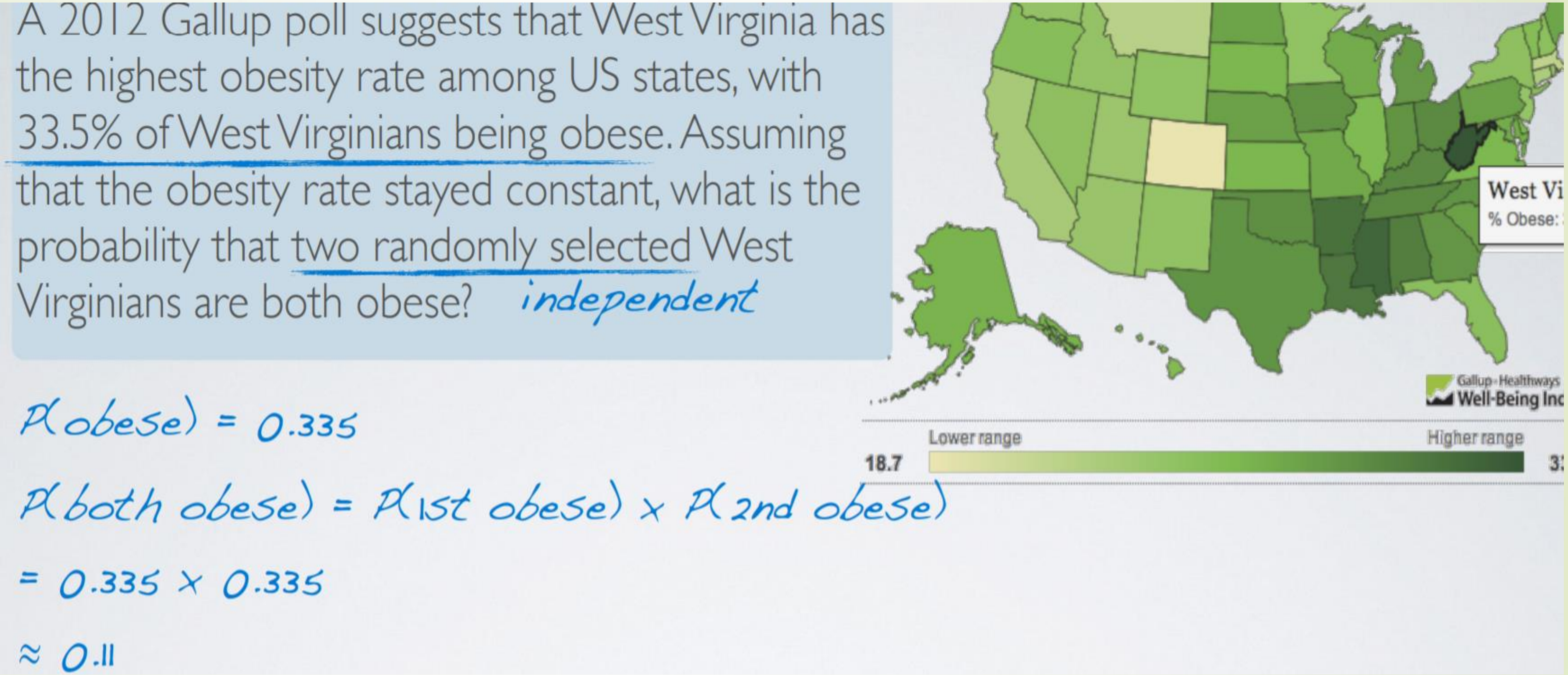
= P(T on the 1st toss) × P(T on the 2nd toss)

= (1/2) × (1/2)

= 1/4

Note: If $A_1, A_2, \ldots, A_k$ are independent, $P(A_1$ and A_2 and … $A_k) = P(A_1) \times P(A_2) \times \ldots \times P(A_k)$

Example



A 2012 Gallup poll suggests that West Virginia has the highest obesity rate among US states, with 33.5% of West Virginians being obese. Assuming that the obesity rate stayed constant, what is the probability that two randomly selected West Virginians are both obese? independent

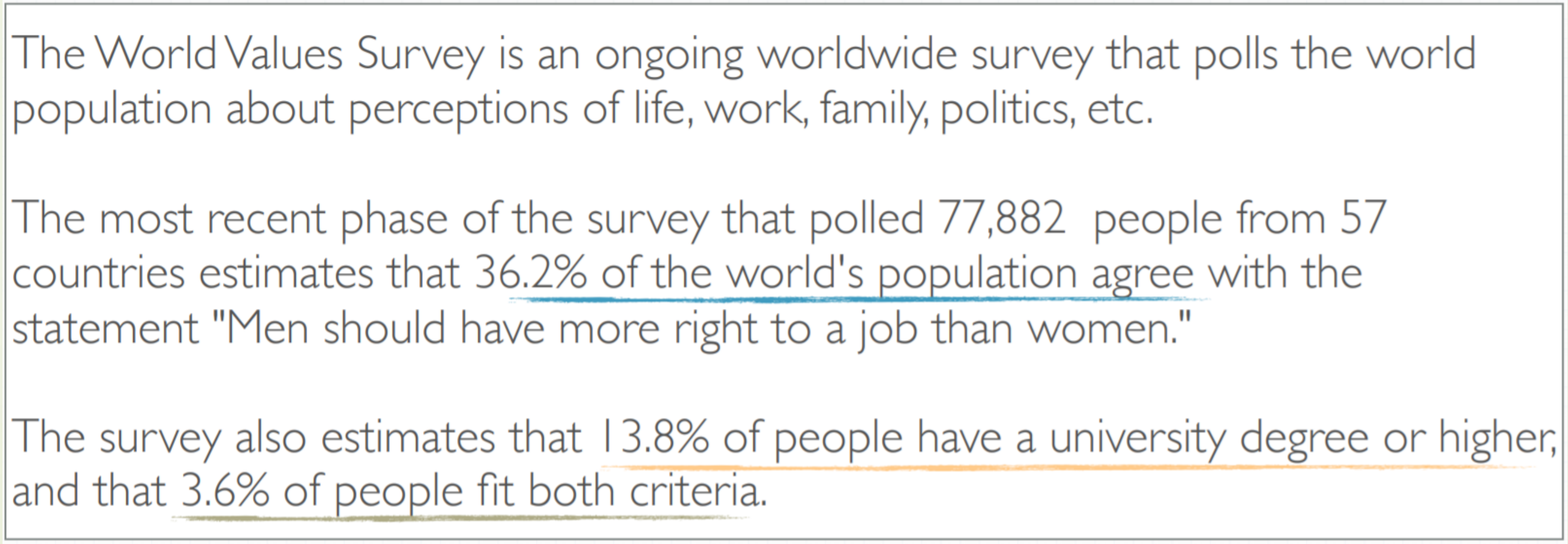
$P(\text{obese}) = 0.335$

$P(\text{both obese}) = P(\text{1st obese}) \times P(\text{2nd obese})$

$= 0.335 \times 0.335$

≈ 0.11

Example

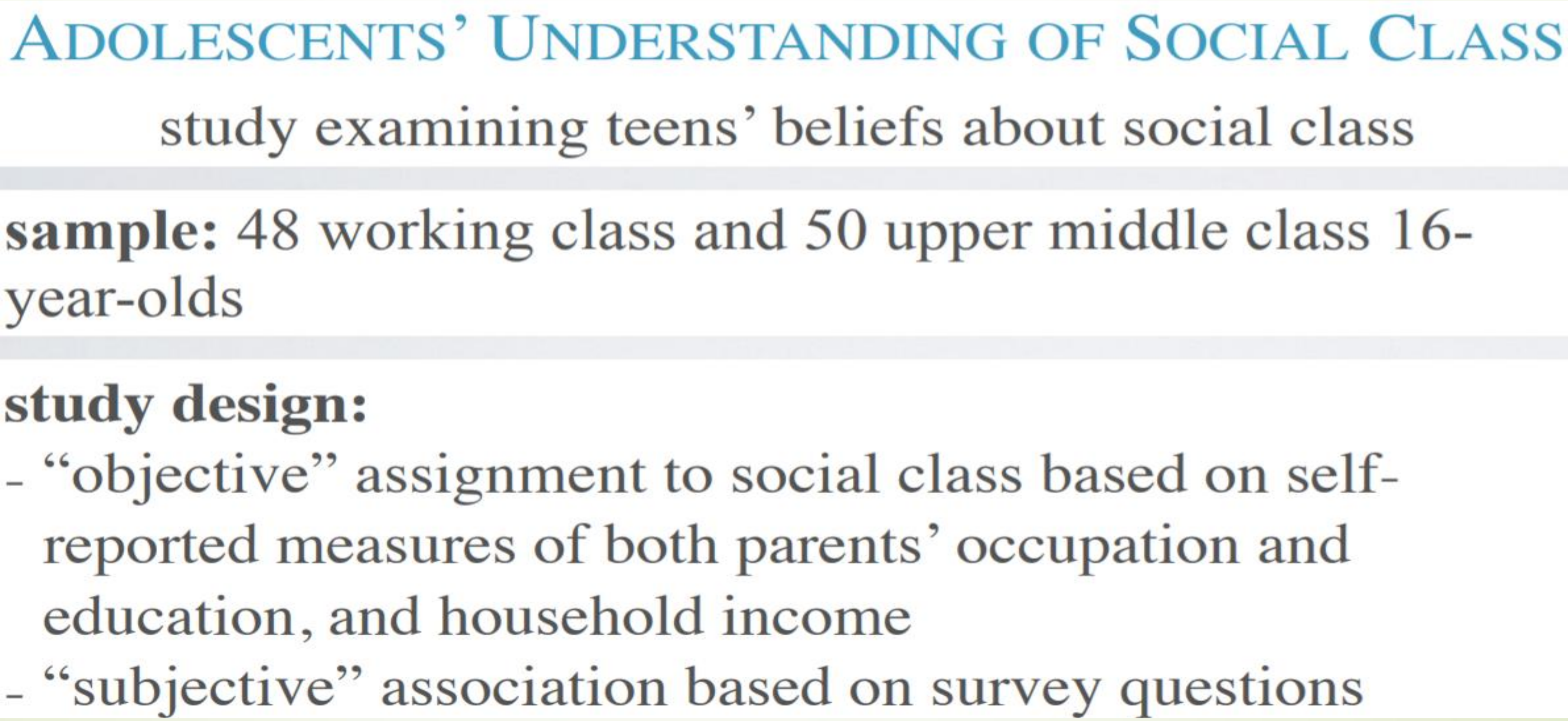


The World Values Survey is an ongoing worldwide survey that polls the world population about perceptions of life, work, family, politics, etc.

The most recent phase of the survey that polled 77,882 people from 57 countries estimates that 36.2% of the world's population agree with the statement "Men should have more right to a job than women."

The survey also estimates that 13.8% of people have a university degree or higher, and that 3.6% of people fit both criteria.

Conditional probability



Adolescents' Understanding of Social Class

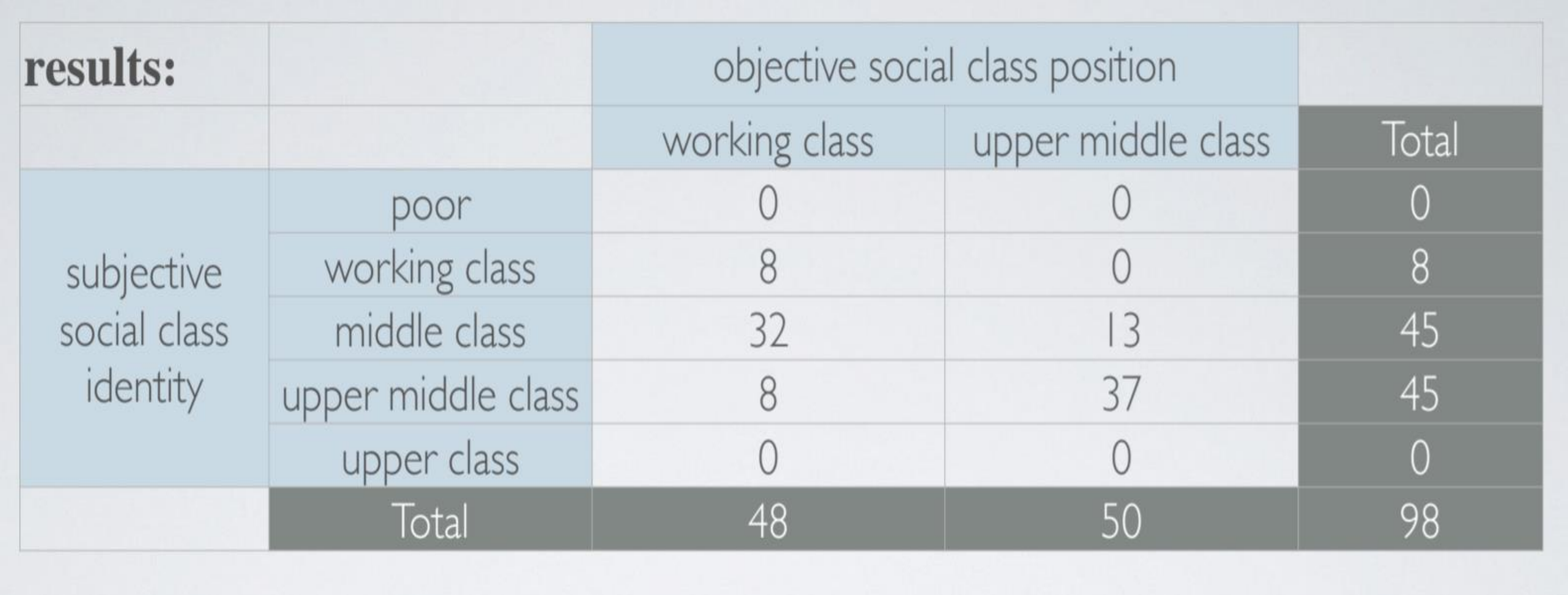
study examining teens' beliefs about social class

sample: 48 working class and 50 upper middle class 16-year-olds

study design:

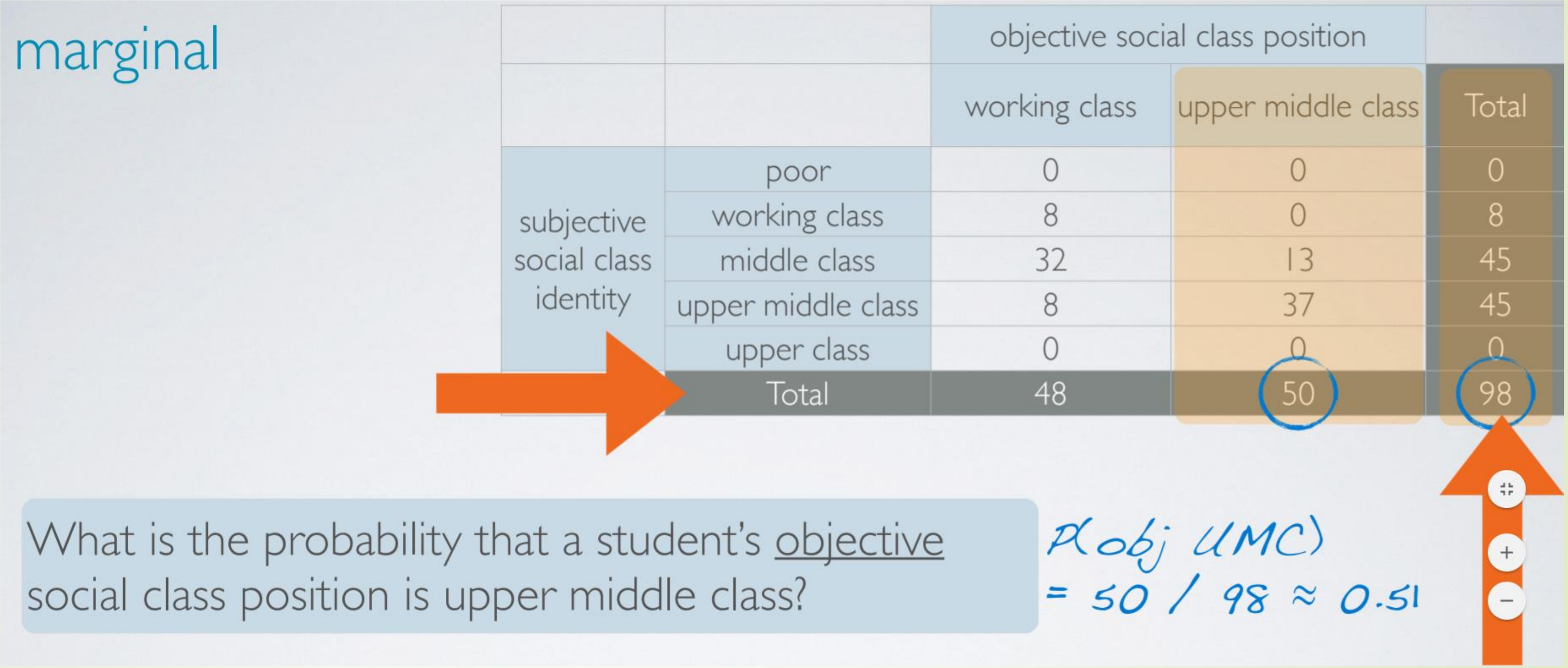
- "objective" assignment to social class based on self-reported measures of both parents' occupation and education, and household income
- "subjective" association based on survey questions

Example



results:		objective social class position		
		working class	upper middle class	Total
subjective social class identity	poor	0	0	0
	working class	8	0	8
	middle class	32	13	45
	upper middle class	8	37	45
	upper class	0	0	0
	Total	48	50	98

Maginal



marginal

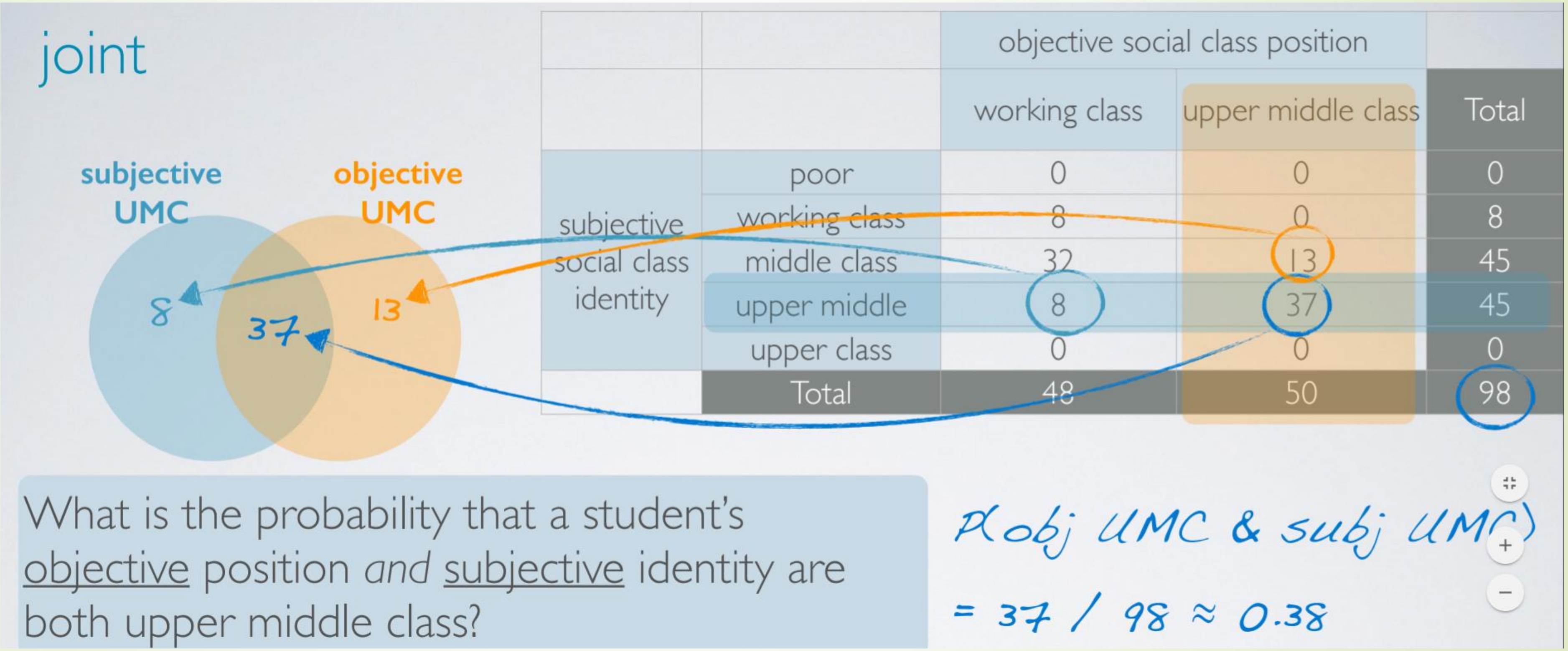
		objective social class position		
		working class	upper middle class	Total
subjective social class identity	poor	0	0	0
	working class	8	0	8
	middle class	32	13	45
	upper middle class	8	37	45
	upper class	0	0	0
	Total	48	50	98

What is the probability that a student's objective social class position is upper middle class?

$P(\text{obj } UMC)$

$= 50 / 98 \approx 0.51$

Joint Probabilities



joint

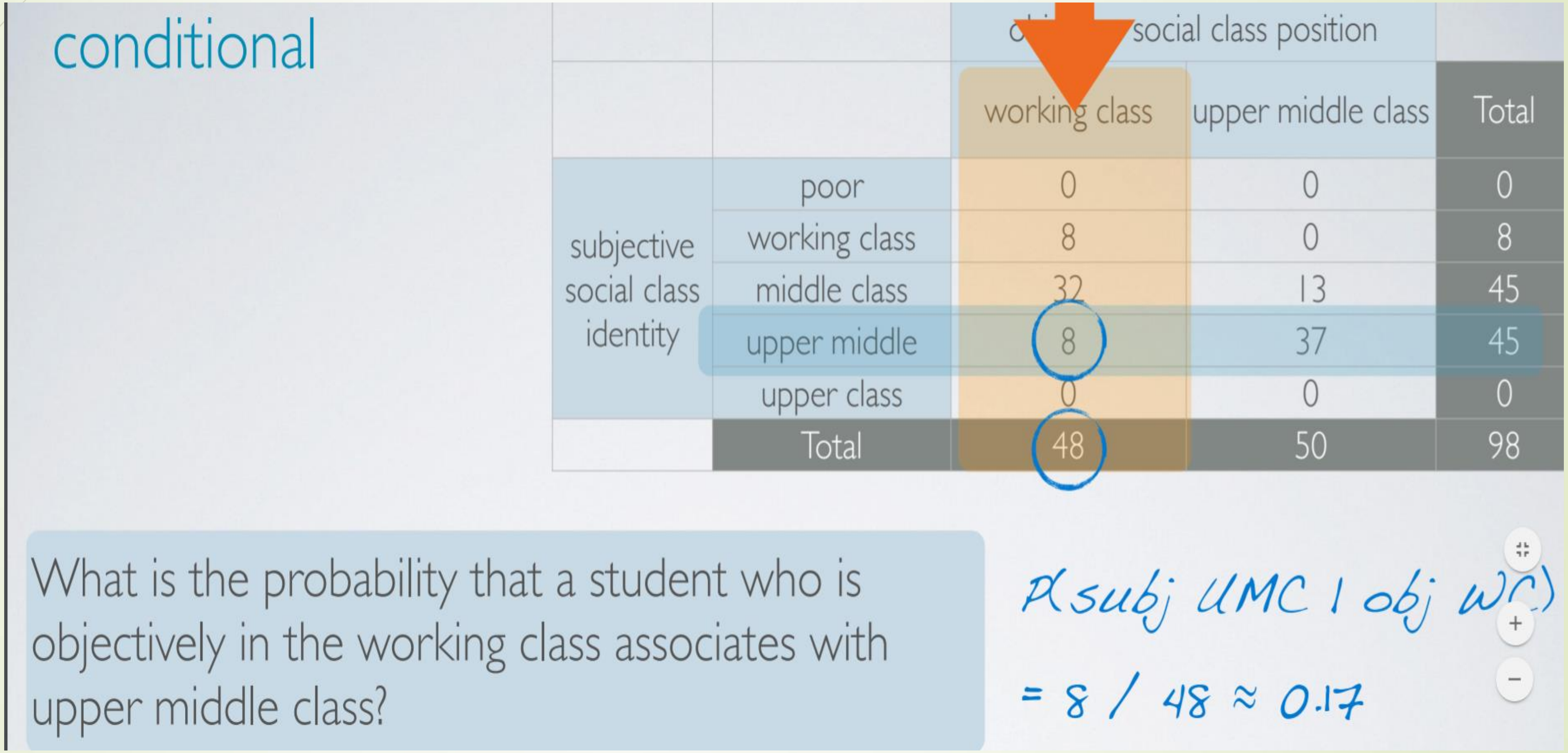
		objective social class position		
		working class	upper middle class	Total
subjective social class identity	poor	0	0	0
	working class	8	0	8
	middle class	32	13	45
	upper middle	8	37	45
	upper class	0	0	0
	Total	48	50	98

What is the probability that a student's objective position *and* subjective identity are both upper middle class?

$P(\text{obj UMC \& subj UMC})$

$= 37 / 98 \approx 0.38$

Conditional



conditional

		objective social class position		
		working class	upper middle class	Total
subjective social class identity	poor	0	0	0
	working class	8	0	8
	middle class	32	13	45
	upper middle	8	37	45
	upper class	0	0	0
	Total	48	50	98

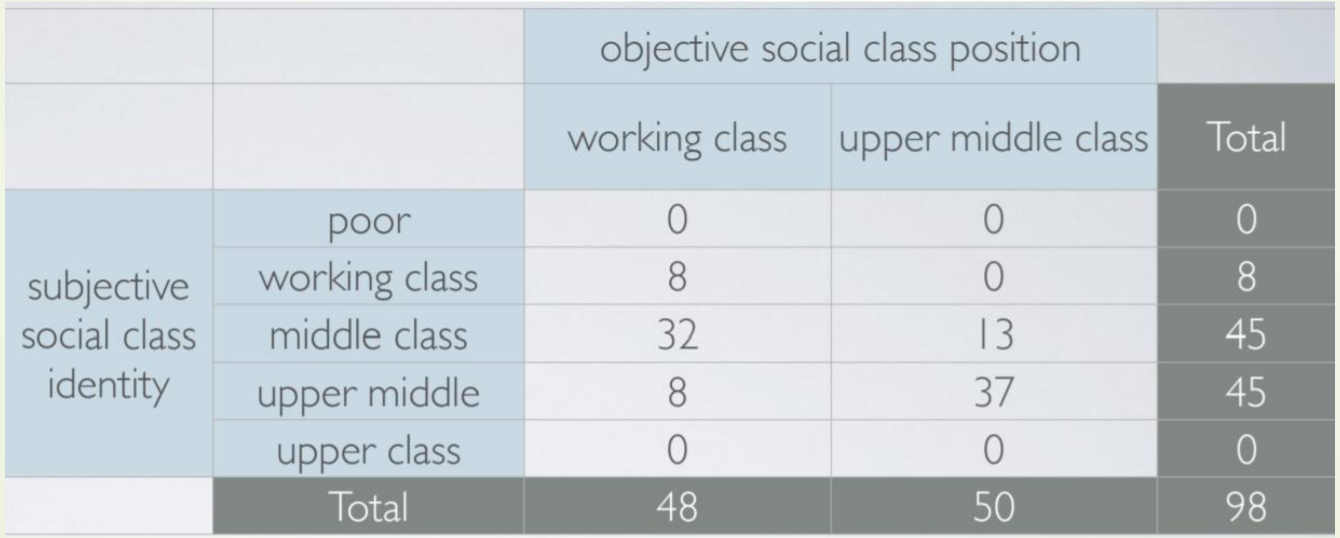
What is the probability that a student who is objectively in the working class associates with upper middle class?

$P(\text{subj UMC} \mid \text{obj WC})$

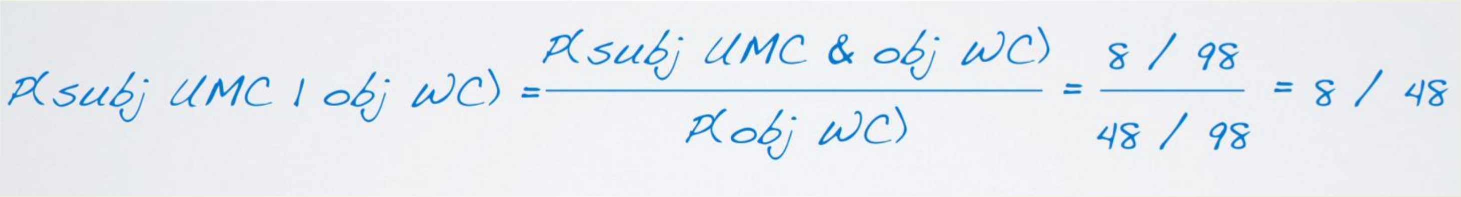
$= 8 / 48 \approx 0.17$

Conditional probabilities

- P(A | B) = P(A and B) / P(B)



		objective social class position		
		working class	upper middle class	Total
subjective social class identity	poor	0	0	0
	working class	8	0	8
	middle class	32	13	45
	upper middle	8	37	45
	upper class	0	0	0
	Total	48	50	98



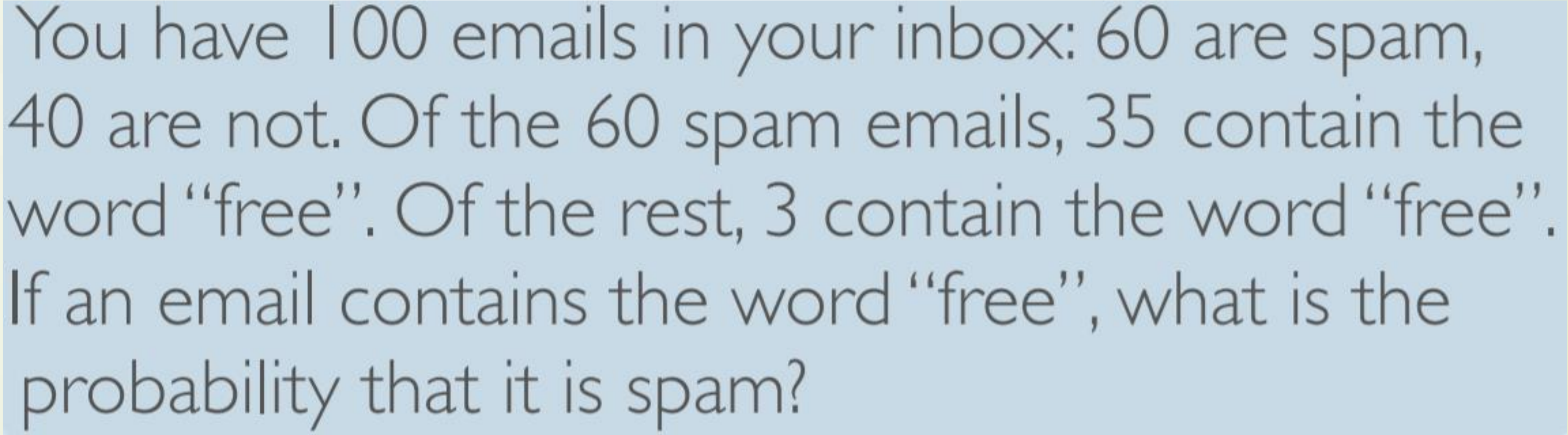
$$P(\text{subj UMC} \mid \text{obj WC}) = \frac{P(\text{subj UMC \& obj WC})}{P(\text{obj WC})} = \frac{8\,/\,98}{48\,/\,98} = 8\,/\,48$$

Problem

- The American Community Survey is an ongoing survey that provides data every year to give communities the current information they need to plan investments and services. The 2010 American Community Survey estimates that 14.6% of Americans live below the poverty line, 20.7% speak a language other than English at home, and 4.2% fall into both categories. Based on this information, what percent of Americans live below the poverty line given that they speak a language other than English at home?
- P(below PL | speak non-Eng) = P(below PL & speak non-Eng)/ P(speak non-Eng) = 0.042/ 0.207 = ≈ 0.2
- P(A | B) = P(A and B) / P(B)

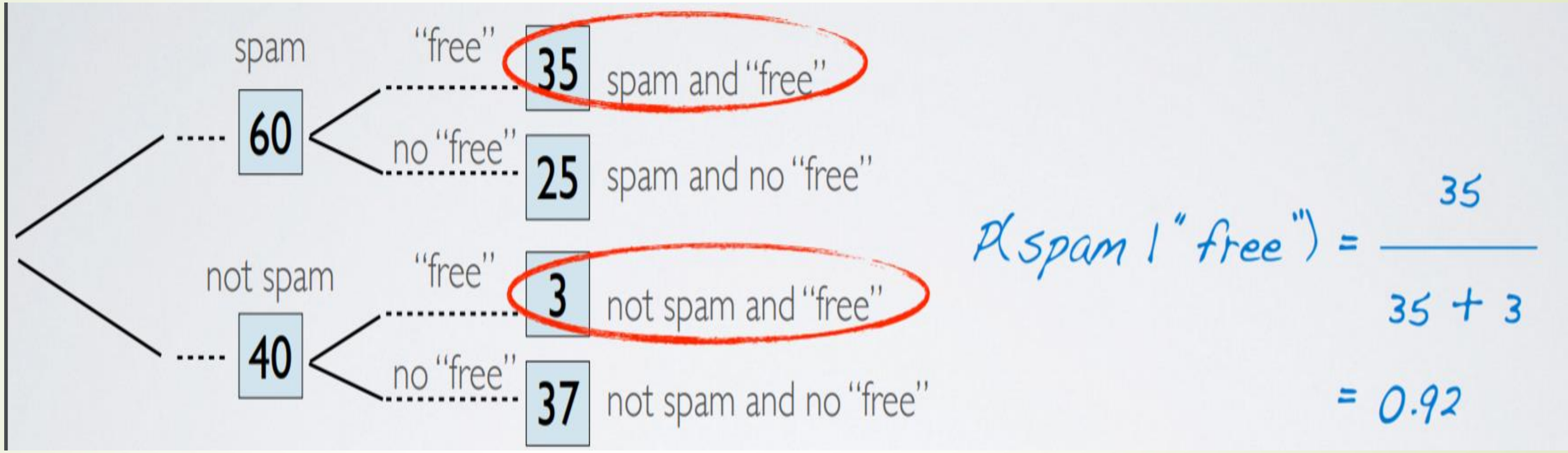
Probability Trees

- P(A | B) → P(B | A)



You have 100 emails in your inbox: 60 are spam, 40 are not. Of the 60 spam emails, 35 contain the word "free". Of the rest, 3 contain the word "free". If an email contains the word "free", what is the probability that it is spam?

Solution



spam

60

"free"

35 spam and "free"

no "free"

25 spam and no "free"

not spam

40

"free"

3 not spam and "free"

no "free"

37 not spam and no "free"

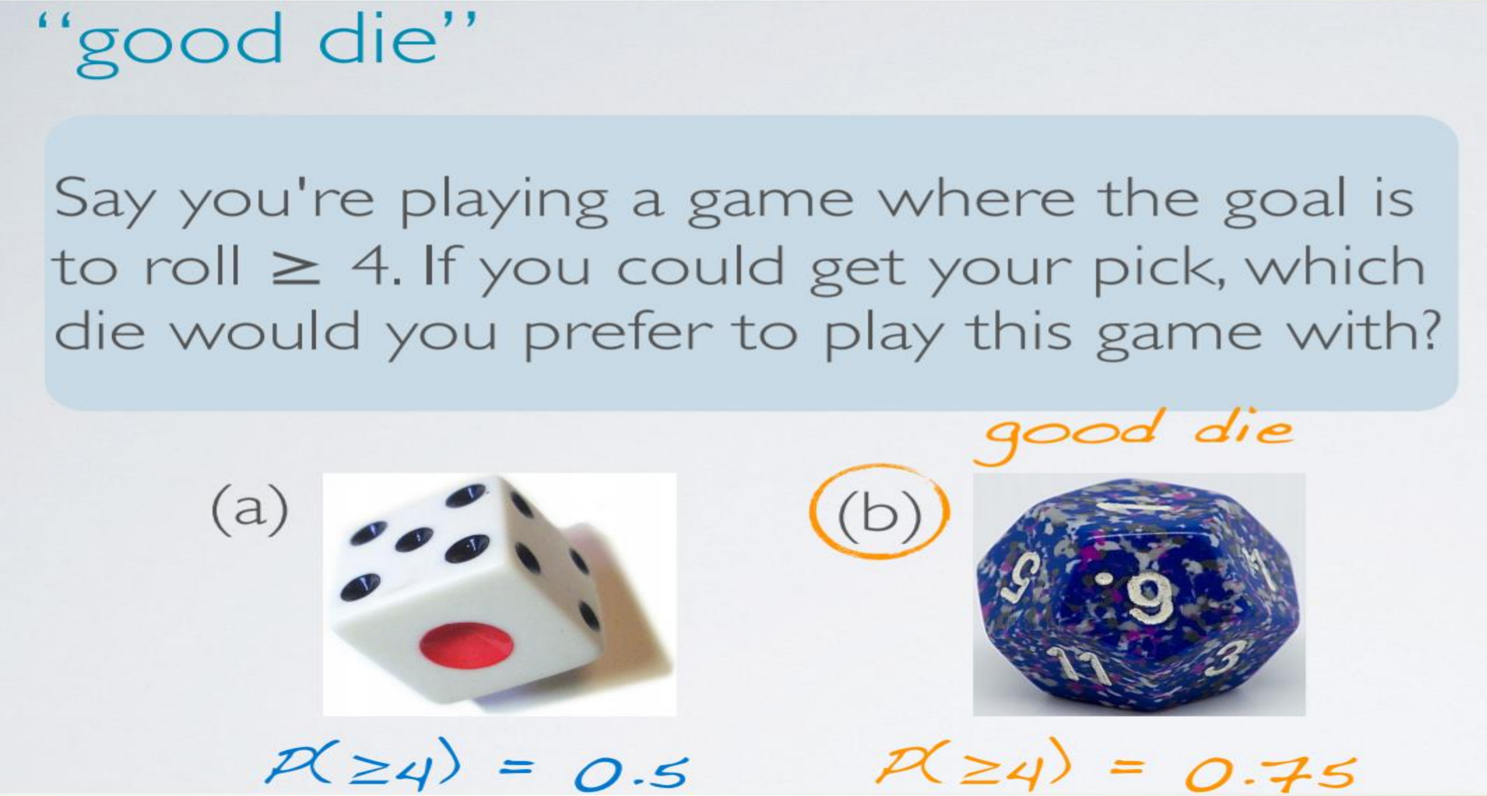
$$P(\text{spam} \mid \text{"free"}) = \frac{35}{35+3}$$

$$= 0.92$$

Bayesian inference

- Probabilities:
- What is the probability of rolling ≥4 with a 6-sided die?
- S = {1,2,3,4,5,6}
- P(≥4) = 3/6 = 1/2 = 0.5

- What is the probability of rolling ≥4 with a 12-sided die?
- S = {1,2,3,4,5,6,7,8,9,10,11,12}
- P(≥4) = 9/12 = 3/4 = 0.75



"good die"

Say you're playing a game where the goal is to roll ≥ 4. If you could get your pick, which die would you prefer to play this game with?

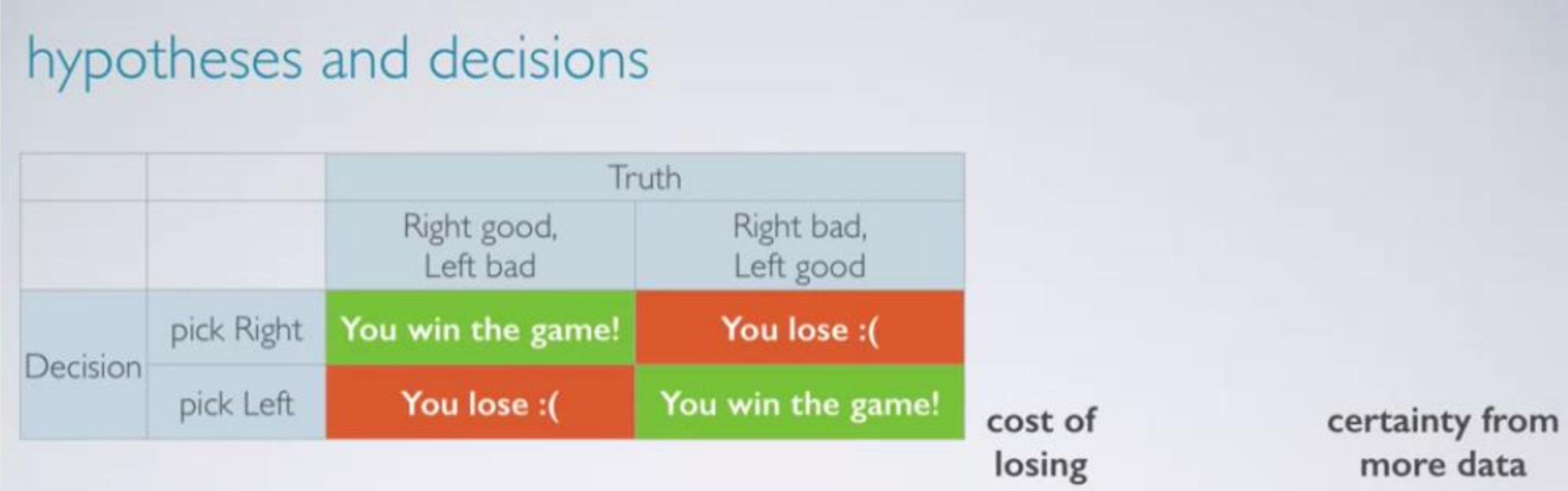
good die

(a) [✓ (b)]

$P(\geq 4) = 0.5$

$P(\geq 4) = 0.75$

Bayes Theorem



hypotheses and decisions

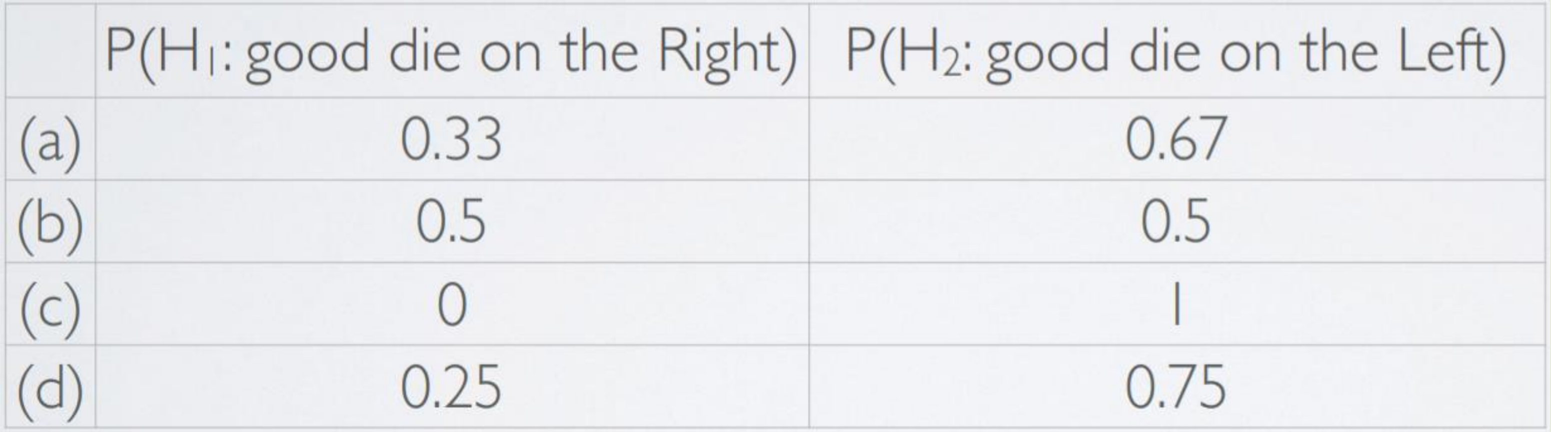
		Truth	
		Right good, Left bad	Right bad, Left good
Decision	pick Right	You win the game!	You lose :(
	pick Left	You lose :(	You win the game!

cost of losing

certainty from more data

before you collect data

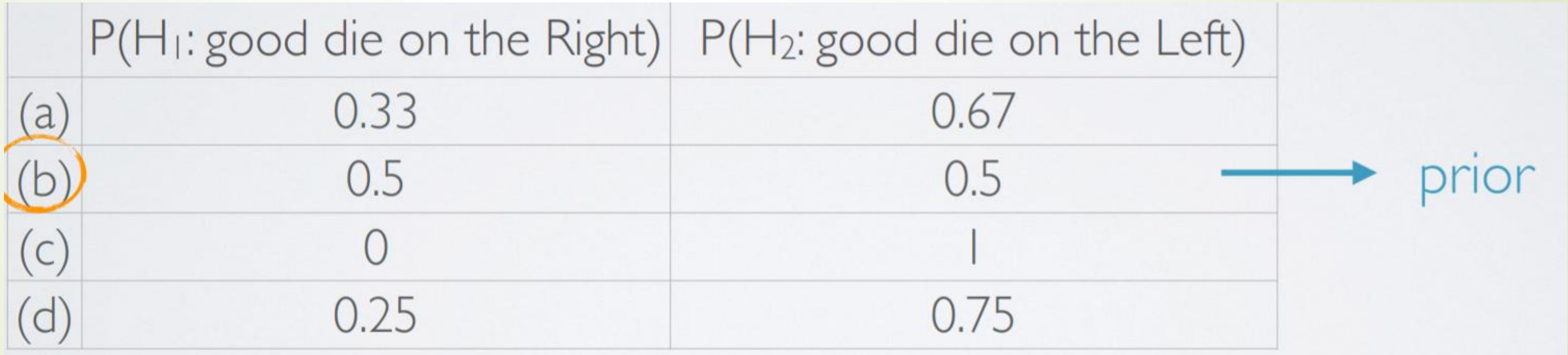
- Before we collect any data, you have no idea if I am holding the good die (12-sided) on the right hand or the left hand. Then, what are the probabilities associated with the following hypotheses?
- H1: good die on the Right (bad die on the Left)
- H2: good die on the Left (bad die on the Right)



	P(H_1: good die on the Right)	P(H_2: good die on the Left)
(a)	0.33	0.67
(b)	0.5	0.5
(c)	0	1
(d)	0.25	0.75

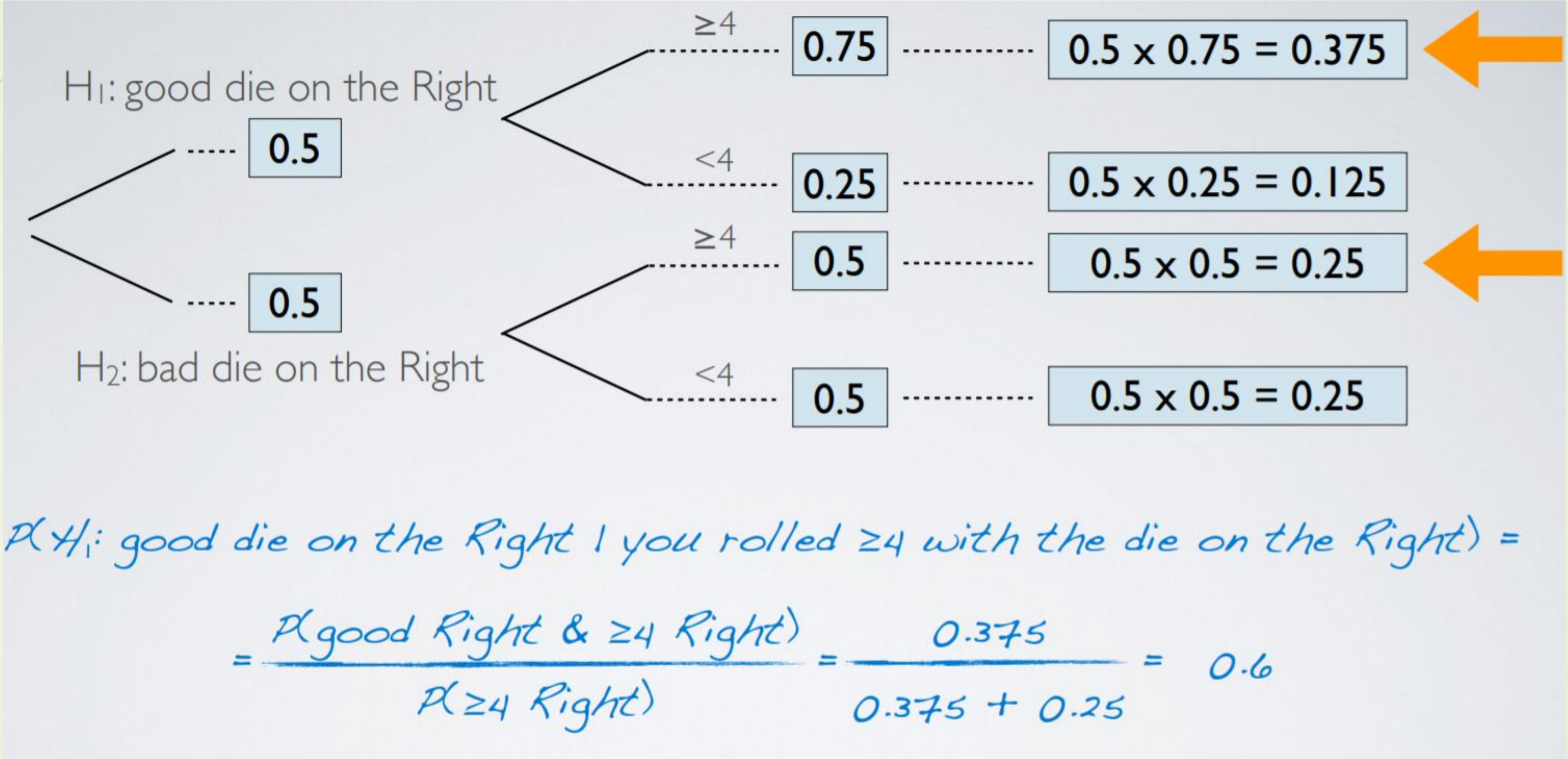
Prior

- Your assumption before you collect the data.



	P(H_1: good die on the Right)	P(H_2: good die on the Left)
(a)	0.33	0.67
[✓ (b)]	0.5	0.5
(c)	0	1
(d)	0.25	0.75

→ prior



$P(H_1$: good die on the Right | you rolled ≥ 4 with the die on the Right$) =$

$$= \frac{P(\text{good Right \& } \geq 4 \text{ Right})}{P(\geq 4 \text{ Right})} = \frac{0.375}{0.375 + 0.25} = 0.6$$

posterior

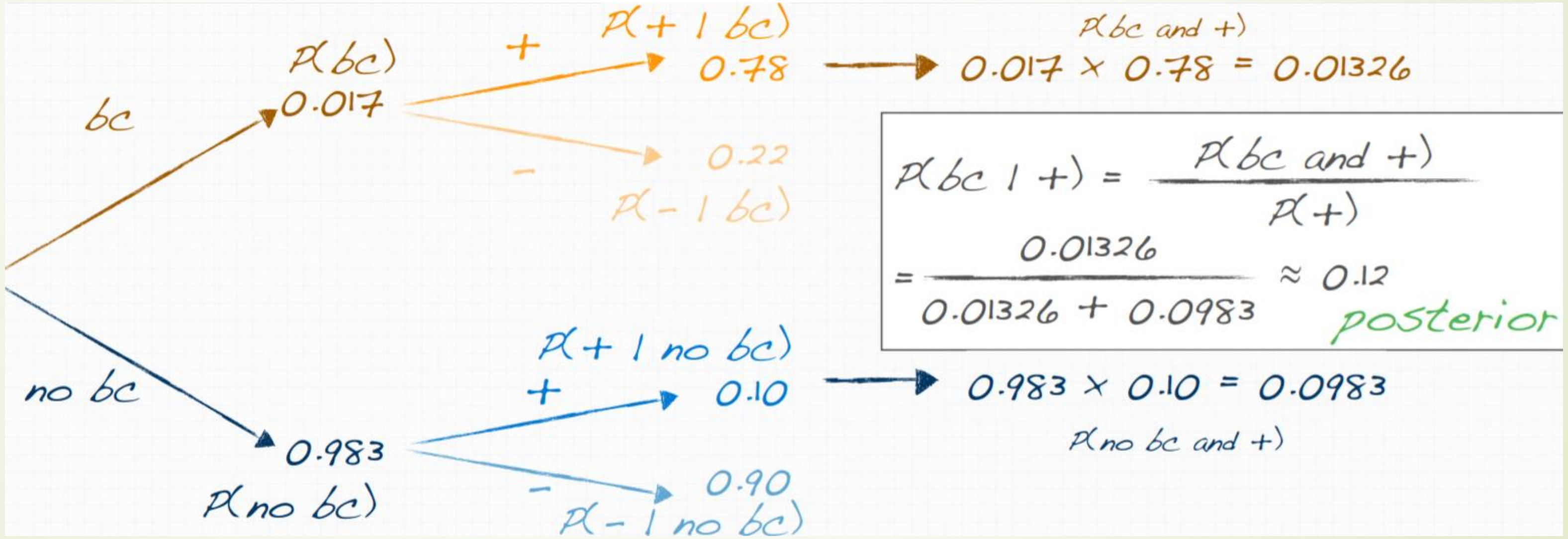
- The probability we just calculated is also called the posterior probability.
- The probability we just calculated is also called the posterior probability. P(H1: good die on the Right | you rolled ≥4 with the die on the Right)
- Posterior probability is generally defined as P(hypothesis | data).
- It tells us the probability of a hypothesis we set forth, given the data we just observed.
- It depends on both the prior probability we set and the observed data.

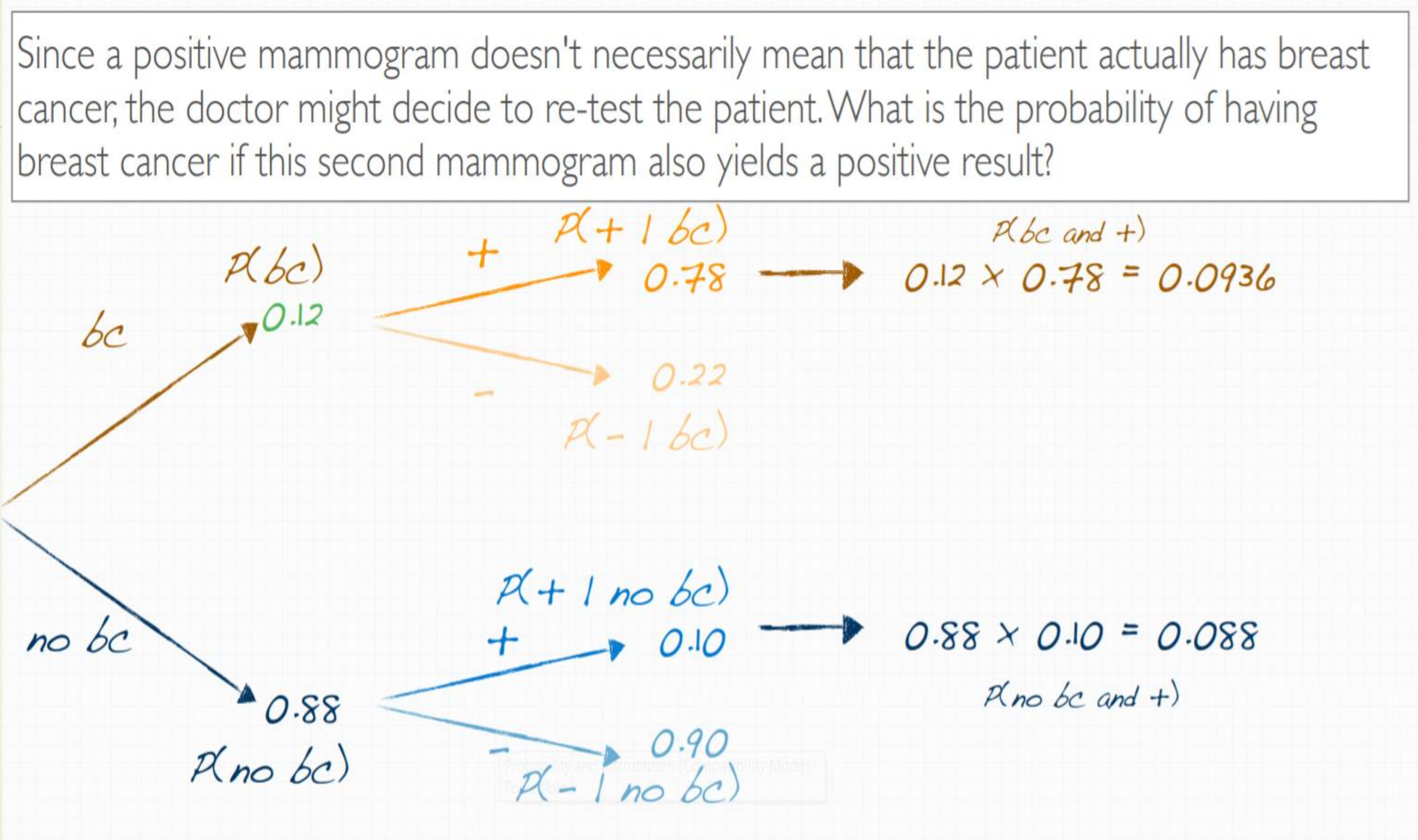
Example

- American Cancer Society estimates that about 1.7% of women have breast cancer. Susan G. Komen For The Cure Foundation states that mammography correctly identifies about 78% of women who truly have breast cancer. An article published in 2003 suggests that up to 10% of all mammograms are false positive.
- Prior to any testing and any information exchange between the patient and the doctor, what probability should a doctor assign to a female patient having breast cancer?
- P(bc) = 0.017 prior

Contd....

- When a patient goes through breast cancer screening there are two competing claims: patient has cancer and patient doesn't have cancer. If a mammogram yields a positive result, what is the probability that patient has cancer? .





Since a positive mammogram doesn't necessarily mean that the patient actually has breast cancer, the doctor might decide to re-test the patient. What is the probability of having breast cancer if this second mammogram also yields a positive result?

- Thanks