Review Session 2

API-201, 9.17.21
Sophie Hill

Logistics

- **PSet 2** is out, due Tuesday am
- **PSet 1** solutions are posted on Canvas, and grades will be released by next week
- Remember to come to **Office Hours** or use **Slack** to #askforhelp!
- Video, slides, and worksheets from the **Excel training sessions** are also available now

Logistics

Office hours schedule

API 201A: Quantitative Analysis and Empirical Methods

Edit

API-201 Schedule, Materials & Deliverables (Fall 2021)

Note - All readings are optional unless explicitly stated in a problem set.

Syllabus

Teaching team office hours

Pre-Class Exercise Solutions (updated every week)

Excel training sessions: Sign-up sheet Zoom link Slides Worksheet (basic) Worksheet (intermediate) FAQs Recording (intermediate)

Review Session: Anonymous Feedback Form

Date	Class	Topic	Assignment Due	Readings	Class & Section Handouts
Part I - UNCERTAINTY					
Unit IA: Probability					

Logistics

API-201 Office Hours

Teaching Team Office Hours

Day	Time	TT Member	Location
Sunday	2:00 - 4:00 PM	Camila de la Vega	Zoom
Sunday	4:30 - 6:30 PM	Aline Atie	Zoom
Monday	8:30 - 10:30 AM	Will Whitehurst	L-330
Monday	12:00 - 2:00 PM	Sophie Hill	Zoom
Monday	2:30 - 4:30 PM	Avery Schmidt	T-401
Monday	3:00 - 5:00 PM	Danica Yu	B-400
Monday	4:30 - 6:30 PM	Svenja Kirsch	Zoom
Monday	4:30 - 6:30 PM	Nikhil Swaminathan	L-332

Logistics

Excel training session materials

API 201A: Quantitative Analysis and Empirical Methods

Edit

API-201 Schedule, Materials & Deliverables (Fall 2021)

Note - All readings are optional unless explicitly stated in a problem set.

Syllabus

Teaching team office hours

Pre-Class Exercise Solutions (updated every week)

Excel training sessions: Sign-up sheet Zoom link Slides Worksheet (basic) Worksheet (intermediate) FAQs Recording (intermediate)

Review Session: Anonymous Feedback Form

Date	Class	Topic	Assignment Due	Readings	Class & Section Handouts
Part I - UNCERTAINTY					
Unit IA: Probability					

Logistics

Review Session slides

Part I - UNCERTAINTY					
Unit IA: Probability					
9/2	1	Thinking Probabilistically About the World	PS #0		H1
9/7	2	Probability and Conditional Probability	PCE	MMC: 4.5, 2.6 OIS: 3.1, 3.2	H2
9/9	3	Bayes' Rule	PCE	New York Times(2016) MMC: 4.5 (pp. 292-293) OIS: 3.2.8	H3 Review Session 1
9/14	4	Application: Public Pensions in Mexico	PS #1 PCE Solutions	Providing Pensions for the Poor: Targeting Cash Transfers for the Elderly in Mexico	H4

Logistics

PSet 1 solutions

Part I - UNCERTAINTY					
Unit IA: Probability					
9/2	1	Thinking Probabilistically About the World	PS #0		H1
9/7	2	Probability and Conditional Probability	PCE	MMC: 4.5, 2.6 OIS: 3.1, 3.2	H2
9/9	3	Bayes' Rule	PCE	New York Times(2016) MMC: 4.5 (pp. 292-293) OIS: 3.2.8	H3 Review Session 1
9/14	4	Application: Public Pensions in Mexico	PS #1 PCE Solutions	Providing Pensions for the Poor: Targeting Cash Transfers for the Elderly in Mexico	H4

Common confusions

Common confusions

1. Independent vs. mutually exclusive
2. Probabilities vs. counts
3. Complement of a conditional probability
4. Sensitivity vs specificity

1. Independent vs. mutually exclusive

From Handout 2:

Events A and B are **independent** if knowing that one event occurs does not change the probability we would assign to the other event.

Two events are **disjoint** (or **mutually exclusive**) if they cannot occur together.

1. Independent vs. mutually exclusive

Can we “translate” these definitions into formal probability notation?

From Handout 2:

Events A and B are **independent** if knowing that one event occurs does not change the probability we would assign to the other event.

Two events are **disjoint** (or **mutually exclusive**) if they cannot occur together.

1. Independent vs. mutually exclusive

Can we "translate" these definitions into formal probability notation?

From Handout 2:

Events A and B are **independent** if knowing that one event occurs does not change the probability we would assign to the other event.

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

Two events are **disjoint** (or **mutually exclusive**) if they cannot occur together.

$$P(A \,\&\, B) = 0$$

1. Independent vs. mutually exclusive

From Handout 2:

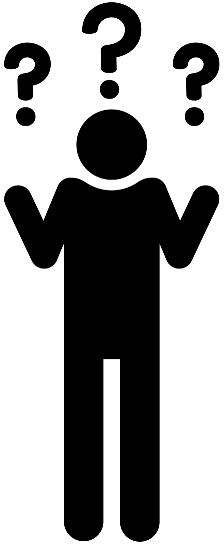
Events A and B are **independent** if knowing that one event occurs does not change the probability we would assign to the other event.

P(being in section A | tall) = P(being in section A)

Two events are **disjoint** (or **mutually exclusive**) if they cannot occur together.

P(being in section A & being in section B) = 0

2. Probabilities vs counts



Sophie, I did Q2 using counts but my PSet group did it using probabilities...

which method is right?

2. Probabilities vs counts

Here's the probability table from the mammograms example (Handout 3)

Cancer?

		Yes	No	
Test result	**Positive**	0.00675	0.099	0.10575
	Negative	0.00225	0.8919	0.89415
		0.009	0.9909	1

2. Probabilities vs counts

To convert to counts, we multiply each cell by our population size (e.g. 1,000)

Cancer?

		Yes	No	
Test result	**Positive**	0.00675 *1000	0.099 *1000	0.10575 *1000
	Negative	0.00225 *1000	0.8919 *1000	0.89415 *1000
		0.009 *1000	0.9909 *1000	1 *1000

2. Probabilities vs counts

To convert to counts, we multiply each cell by our population size (e.g. 1,000)

Cancer?

		Yes	No	
Test result	**Positive**	6.75	99	105.75
	Negative	2.25	891.9	894.15
		9	990.9	1,000

2. Probabilities vs counts

What is P(Cancer | +) ?

Cancer?

		Yes	No	
Test result	**Positive**	6.75	99	105.75
	Negative	2.25	891.9	894.15
		9	990.9	1,000

2. Probabilities vs counts

What is P(Cancer | +) ?

6.75 / 105.75 = 0.064

		Cancer? Yes	Cancer? No	
Test result	Positive	6.75	99	105.75
	Negative	2.25	891.9	894.15
		9	990.9	10,000

2. Probabilities vs counts

$$P(Cancer|+) = \frac{P(+ \ \& \ Cancer)}{P(+)}$$

probabilities

$$= \frac{0.00675}{0.10575}$$

multiply top and
bottom by 1,000

$$= \frac{0.00675 \times 1000}{0.10575 \times 1000}$$

counts

$$= \frac{6.75}{105.75}$$

same answer!

$$= 0.064$$

2. Probabilities vs counts

Using probabilities is fine

Using counts is fine (perhaps easier?)

The only thing you need to remember is that you must be **CONSISTENT**!!

Always sense-check your answer:

- Probabilities must be between 0 and 1
- Counts must be between 0 and your total population size

3. Complement of a conditional probability

$$1 - P(A|B) = ???$$

3. Complement of a conditional probability

$$1 - P(A|B) = P(\text{not } A \mid B)$$

3. Complements of conditional probability

$$1 - P(A|B) = P(\text{not } A \mid B)$$

Why?

Let's prove it...

$$1 - P(A|B) = 1 - \frac{P(A\&B)}{P(B)}$$

Multiplication rule

If you're having trouble remembering the probability rules, print out this box from **Handout 2** and have it in front of you when working on a PSet!

PROBABILITY RULES

1. The probability of any event A satisfies $0 \leq P(A) \leq 1$
2. Something must happen, so the sum of the probabilities of distinct events is 1.
3. **Addition Rule:** For any two events $A\ and\ B,$

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

*Special Case: If two events $A\ and\ B$ are **disjoint/mutually exclusive**, then,

$$P(A\ or\ B) = P(A) + P(B)$$

4. **Complement Rule**: For any event $A,$

$$P(A^C) = 1 - P(A)$$

5. **Multiplication Rule**: For two events $A\ and\ B,$

$$P(A\ and\ B) = P(A) * P(B|A)$$

*Special Case (next class): If $A\ and\ B$ are **independent** events, then

$$P(A\ and\ B) = P(A) * P(B)$$

$$1 - P(A|B) = 1 - \frac{P(A\&B)}{P(B)}$$

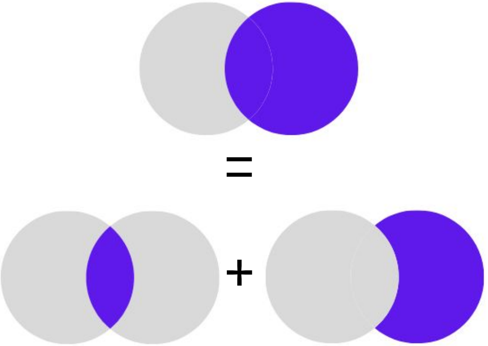
Multiplication rule

$$1 - P(A|B) = 1 - \frac{P(A\&B)}{P(B)}$$

Rearranging

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

$$1 - P(A|B) = 1 - \frac{P(A\&B)}{P(B)}$$



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

$$= \frac{[P(A\&B) + P(\neg A\&B)] - P(A\&B)}{P(B)}$$

$$
\begin{aligned}
1 - P(A|B) &= 1 - \frac{P(A\&B)}{P(B)} \\
&= \frac{P(B) - P(A\&B)}{P(B)} \\
&= \frac{[\cancel{P(A\&B)} + P(\neg A\&B)] - \cancel{P(A\&B)}}{P(B)} \\
&= \frac{P(\neg A\&B)}{P(B)}
\end{aligned}
$$

$$
\begin{aligned}
1 - P(A|B) &= 1 - \frac{P(A\&B)}{P(B)} \\
&= \frac{P(B) - P(A\&B)}{P(B)} \\
&= \frac{[P(A\&B) + P(\neg A\&B)] - P(A\&B)}{P(B)} \\
&= \frac{P(\neg A\&B)}{P(B)} \\
&= P(\neg A|B)
\end{aligned}
$$

Multiplication rule (again!)

Wow, that multiplication rule sure seems useful!!

Wow, that multiplication rule sure seems useful!!

`“Conditional = Joint / 2nd Marginal”`

Wow, that multiplication rule sure seems useful!!

"Conditional = Joint / 2nd Marginal"

$$P(A|B) = P(A\&B) / P(B)$$
$$P(B|A) = P(A\&B) / P(A)$$

4. Sensitivity vs Specificity

Sensitivity	P(+ | Disease)	Probability of *correctly* classifying those *with* the disease
Specificity	P(– | No disease)	Probability of *correctly* classifying those *without* the disease

4. Sensitivity vs Specificity

Sensitivity	P(+ | Disease)	Probability of *correctly* classifying those *with* the disease
1 - Sensitivity	P(– | Disease)	Probability of *incorrectly* classifying those *with* the disease
Specificity	P(– | No disease)	Probability of *correctly* classifying those *without* the disease
1 - Specificity	P(+ | No disease)	Probability of *incorrectly* classifying those *without* the disease

4. PPV vs NPV

Positive predictive value	P(Disease | +)	Probability that a *positive* test result is a true positive
Negative predictive value	P(No disease | –)	Probability that a *negative* test result is a true negative

4. PPV vs NPV

Positive predictive value	P(Disease | +)	Probability that a *positive* test result is a true positive
1 – PPV	P(No disease | +)	Probability that a *positive* test result is a false positive
Negative predictive value	P(No disease | –)	Probability that a *negative* test result is a true negative
1 – NPV	P(Disease | –)	Probability that a *negative* test result is a false negative

4. Sensitivity and Specificity vs PPV and NPV

Sensitivity and **specificity** are what matters to the medical testing company – they want to make the test as accurate as possible.

PPV and **NPV** are what matters to the patient – they want to know what the test results means in terms of their own health!

4. Sensitivity and Specificity vs PPV and NPV

Sensitivity and **specificity** are what matters to the medical testing company – they want to make the test as accurate as possible.

PPV and **NPV** are what matters to the patient – they want to know what the test results means in terms of their own health!

Problem Set 1, Question 2 😱

PS1, Q2: Validation

The most important thing to say about this question is:

well done everyone!

- It’s a hard question, especially this early in the semester.
- If you found it tricky, **don’t panic**. It *is* tricky!
- Bayes rule is a *core concept* in this class. It is 100% worth your time to review this question until you are comfortable with it.

PS1, Q2: Strategy

Abbott reports that this test has a specificity (P(Negative Test | No COVID)) of 98.5% and a sensitivity (P(Positive Test | COVID)) of 84.6%. Assume that the prevalence of COVID-19 (i.e., the proportion of people who are currently infected) is 1% in your population of interest.

What is the first thing we do with a question like this?

(A) Guess the answer
(B) Panic
(C) Write out all the information in formal probability notation

PS1, Q2: Strategy

Abbott reports that this test has a specificity (P(Negative Test | No COVID)) of 98.5% and a sensitivity (P(Positive Test | COVID)) of 84.6%. Assume that the prevalence of COVID-19 (i.e., the proportion of people who are currently infected) is 1% in your population of interest.

What is the first thing we do with a question like this?

(A) Guess the answer
(B) Panic
(C) Write out all the information in formal probability notation

PS1, Q2

Abbott reports that this test has a specificity (P(Negative Test | No COVID)) of 98.5% and a sensitivity (P(Positive Test | COVID)) of 84.6%. Assume that the prevalence of COVID-19 (i.e., the proportion of people who are currently infected) is 1% in your population of interest.

P(– | No COVID) = 0.985

PS1, Q2

Abbott reports that this test has a specificity (P(Negative Test | No COVID)) of 98.5% and a sensitivity (P(Positive Test | COVID)) of 84.6%. Assume that the prevalence of COVID-19 (i.e., the proportion of people who are currently infected) is 1% in your population of interest.

$P(- \mid \text{No COVID}) = 0.985$

$P(+ \mid \text{COVID}) = 0.846$

PS1, Q2

Abbott reports that this test has a specificity (P(Negative Test | No COVID)) of 98.5% and a sensitivity (P(Positive Test | COVID)) of 84.6%. Assume that the prevalence of COVID-19 (i.e., the proportion of people who are currently infected) is 1% in your population of interest.

$P(- \mid \text{No COVID}) = 0.985$

$P(+ \mid \text{COVID}) = 0.846$

$P(\text{COVID}) = 0.01$

PS1, Q2

Abbott reports that this test has a specificity (P(Negative Test | No COVID)) of 98.5% and a sensitivity (P(Positive Test | COVID)) of 84.6%. Assume that the prevalence of COVID-19 (i.e., the proportion of people who are currently infected) is 1% in your population of interest.

Let's take complements!

P(– | No COVID) = 0.985 → P(+ | No COVID) = 1 - 0.985 = 0.015

P(+ | COVID) = 0.846 → P(– | COVID) = 1 - 0.846 = 0.154

P(COVID) = 0.01 → P(No COVID) = 1 - 0.01 = 0.99

PS1, Q2

$P(- \mid \text{No COVID}) = 0.985$ → $P(+ \mid \text{No COVID}) = 1 - 0.985 = 0.015$

$P(+ \mid \text{COVID}) = 0.846$ → $P(- \mid \text{COVID}) = 1 - 0.846 = 0.154$

$P(\text{COVID}) = 0.01$ → $P(\text{No COVID}) = 1 - 0.01 = 0.99$

OK… now what is step #2?

(A) Give up
(B) Fill in a 2x2 table with all the joint probabilities
(C) Panic

PS1, Q2

$P(- \mid \text{No COVID}) = 0.985$ → $P(+ \mid \text{No COVID}) = 1 - 0.985 = 0.015$

$P(+ \mid \text{COVID}) = 0.846$ → $P(- \mid \text{COVID}) = 1 - 0.846 = 0.154$

$P(\text{COVID}) = 0.01$ → $P(\text{No COVID}) = 1 - 0.01 = 0.99$

OK… now what is step #2?

(A) Give up

(B) Fill in a 2x2 table with all the joint probabilities

(C) Panic

PS1, Q2

$P(- \mid \text{No COVID}) = 0.985$ → $P(+ \mid \text{No COVID}) = 1 - 0.985 = 0.015$

$P(+ \mid \text{COVID}) = 0.846$ → $P(- \mid \text{COVID}) = 1 - 0.846 = 0.154$

$P(\text{COVID}) = 0.01$ → $P(\text{No COVID}) = 1 - 0.01 = 0.99$

	+ test result	– test result
Covid		
No Covid		

PS1, Q2

$P(- \mid \text{No COVID}) = 0.985 \rightarrow P(+ \mid \text{No COVID}) = 1 - 0.985 = 0.015$

$P(+ \mid \text{COVID}) = 0.846 \rightarrow P(- \mid \text{COVID}) = 1 - 0.846 = 0.154$

$P(\text{COVID}) = 0.01 \rightarrow P(\text{No COVID}) = 1 - 0.01 = 0.99$

	+ test result	**– test result**
Covid	P(+ & Covid)	P(– & Covid)
No Covid	P(+ & No Covid)	P(– & No Covid)

PS1, Q2

$P(- \mid \text{No COVID}) = 0.985$ → $P(+ \mid \text{No COVID}) = 1 - 0.985 = 0.015$

$P(+ \mid \text{COVID}) = 0.846$ → $P(- \mid \text{COVID}) = 1 - 0.846 = 0.154$

$P(\text{COVID}) = 0.01$ → $P(\text{No COVID}) = 1 - 0.01 = 0.99$

But... we don't have any joint probabilities... do we??!

	+ test result	**– test result**
Covid	P(+ & Covid)	P(– & Covid)
No Covid	P(+ & No Covid)	P(– & No Covid)

PS1, Q2

$P(- \mid \text{No COVID}) = 0.985 \rightarrow P(+ \mid \text{No COVID}) = 1 - 0.985 = 0.015$

$P(+ \mid \text{COVID}) = 0.846 \rightarrow P(- \mid \text{COVID}) = 1 - 0.846 = 0.154$

$P(\text{COVID}) = 0.01 \rightarrow P(\text{No COVID}) = 1 - 0.01 = 0.99$

But… we don't have any joint probabilities… do we??! **Yes!**

	+ test result	**– test result**
Covid	P(+ & Covid)	P(– & Covid)
No Covid	P(+ & No Covid)	P(– & No Covid)

PS1, Q2

$P(- \mid \text{No COVID}) = 0.985 \rightarrow P(+ \mid \text{No COVID}) = 1 - 0.985 = 0.015$

$P(+ \mid \text{COVID}) = 0.846 \rightarrow P(- \mid \text{COVID}) = 1 - 0.846 = 0.154$

$P(\text{COVID}) = 0.01 \rightarrow P(\text{No COVID}) = 1 - 0.01 = 0.99$

	+ test result	**– test result**
Covid	P(+ | Covid) * P(Covid)	P(– | Covid) * P(Covid)
No Covid	P(+ | No Covid) * P(No Covid)	P(– | No Covid) * P(No Covid)

Wow… the multiplication rule *again*??

Say it with me now:

Wow… the multiplication rule *again*??

Say it with me now:

"Conditional = Joint / 2nd marginal"
"Joint = Conditional * 2nd marginal"

PS1, Q2

$P(- \mid \text{No COVID}) = 0.985 \rightarrow P(+ \mid \text{No COVID}) = 1 - 0.985 = 0.015$

$P(+ \mid \text{COVID}) = 0.846 \rightarrow P(- \mid \text{COVID}) = 1 - 0.846 = 0.154$

$P(\text{COVID}) = 0.01 \rightarrow P(\text{No COVID}) = 1 - 0.01 = 0.99$

	+ test result	**– test result**
Covid	P(+ | Covid) * P(Covid)	P(– | Covid) * P(Covid)
No Covid	P(+ | No Covid) * P(No Covid)	P(– | No Covid) * P(No Covid)

PS1, Q2

$P(- \mid \text{No COVID}) = 0.985$ → $P(+ \mid \text{No COVID}) = 1 - 0.985 = 0.015$

$P(+ \mid \text{COVID}) = 0.846$ → $P(- \mid \text{COVID}) = 1 - 0.846 = 0.154$

$P(\text{COVID}) = 0.01$ → $P(\text{No COVID}) = 1 - 0.01 = 0.99$

	+ test result	**– test result**
Covid	0.846 * 0.01	0.154 * 0.01
No Covid	0.015 * 0.99	0.985 * 0.99

PS1, Q2

$P(- \mid \text{No COVID}) = 0.985 \rightarrow P(+ \mid \text{No COVID}) = 1 - 0.985 = 0.015$

$P(+ \mid \text{COVID}) = 0.846 \rightarrow P(- \mid \text{COVID}) = 1 - 0.846 = 0.154$

$P(\text{COVID}) = 0.01 \rightarrow P(\text{No COVID}) = 1 - 0.01 = 0.99$

	+ test result	**– test result**	**Totals**
Covid	85	15	100
No Covid	148	9,752	9,900
Totals	233	9,767	10,000

PS1, Q2

(a) What is the *positive predictive value*, P(COVID | +)?

	+ test result	– test result	Totals
Covid	85	15	100
No Covid	148	9,752	9,900
Totals	233	9,767	10,000

PS1, Q2

$$P(COVID|+) = \frac{P(COVID\ \&\ +)}{P(+)}$$

$$= \frac{\text{TRUE positives}}{\text{ALL positives}}$$

	+ test result	– test result	Totals
Covid	85	15	100
No Covid	148	9,752	9,900
Totals	233	9,767	10,000

PS1, Q2

So P(COVID | +)

$$= (85 / 10{,}000) / (233 / 10{,}000)$$
$$= 85 / 233$$
$$= 0.36$$

	+ test result	– test result	Totals
Covid	85	15	100
No Covid	148	9,752	9,900
Totals	233	9,767	10,000

PS1, Q2(c)

Part (b) was about comparing this scenario to the mammograms example - see solutions.

Let’s do part (c) together…

(c) What is P(No Covid | –) ?

PS1, Q2(c)

Part (b) was about comparing this scenario to the mammograms example - see solutions.

Let's do part (c) together...

(c) What is P(No Covid | –) ?

$$P(No\ COVID|-) = \frac{P(No\ COVID\ \&\ -)}{P(-)}$$

$$= \frac{\text{TRUE negatives}}{\text{ALL negatives}}$$

PS1, Q2(c)

Part (b) was about comparing this scenario to the mammograms example - see solutions.

Let's do part (c) together...

$$P(No\ COVID|-) = \frac{P(No\ COVID\ \&\ -)}{P(-)}$$

(c) What is P(No Covid | –) ?

	+	–	Totals
Covid	85	15	100
No Covid	148	9,752	9,900
Totals	233	9,767	10,000

$$= \frac{\text{TRUE negatives}}{\text{ALL negatives}}$$

PS1, Q2(c)

Part (b) was about comparing this scenario to the mammograms example - see solutions.

Let's do part (c) together...

$$P(No\ COVID|-) = \frac{P(No\ COVID\ \&\ -)}{P(-)}$$

(c) What is P(No Covid | –) ?

$$= \frac{\text{TRUE negatives}}{\text{ALL negatives}}$$

	+	**–**	**Totals**
Covid	85	15	100
No Covid	148	9,752	9,900
Totals	233	9,767	10,000

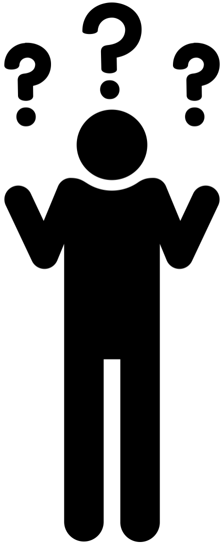
= 9752 / 9767

= 0.998 (to 3dp)

PS1, Q2(c)

But Sophie, I thought this question was about Bayes rule!

We haven’t used it...have we??



PS1, Q2(c)

(c) What is P(No Covid | –) ?

How did we get these numbers in the table?

	+	–	Totals
Covid	85	15	100
No Covid	148	9,752	9,900
Totals	233	9,767	10,000

PS1, Q2(c)

(c) What is P(No Covid | –) ?

How did we get these numbers in the table?

	+ test result	**– test result**
Covid	P(+ | Covid) * P(Covid)	P(– | Covid) * P(Covid)
No Covid	P(+ | No Covid) * P(No Covid)	P(– | No Covid) * P(No Covid)

PS1, Q2(c)

(c) What is P(No Covid | –) ?

Let's add the column totals as well...

	+ test result	**– test result**
Covid	P(+ | Covid) * P(Covid)	P(– | Covid) * P(Covid)
No Covid	P(+ | No Covid) * P(No Covid)	P(– | No Covid) * P(No Covid)
Totals	P(+ | No Covid) * P(No Covid) + P(+ | Covid) * P(Covid)	P(– | No Covid) * P(No Covid) + P(– | Covid) * P(Covid)

$$P(No\ Covid|-) = \frac{P(-|No\ Covid)P(No\ Covid)}{P(-|No\ Covid)P(No\ Covid) + P(-|Covid)P(Covid)}$$

	+ test result	**– test result**
Covid	P(+ | Covid) * P(Covid)	P(– | Covid) * P(Covid)
No Covid	P(+ | No Covid) * P(No Covid)	P(– | No Covid) * P(No Covid)
Totals	P(+ | No Covid) * P(No Covid) + P(+ | Covid) * P(Covid)	P(– | No Covid) * P(No Covid) + P(– | Covid) * P(Covid)

$$P(No\ Covid|-) = \frac{P(-|No\ Covid)P(No\ Covid)}{P(-|No\ Covid)P(No\ Covid) + P(-|Covid)P(Covid)}$$

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

It's Bayes rule!!!

	+ test result	**– test result**
Covid	P(+ | Covid) * P(Covid)	P(– | Covid) * P(Covid)
No Covid	P(+ | No Covid) * P(No Covid)	P(– | No Covid) * P(No Covid)
Totals	P(+ | No Covid) * P(No Covid) + P(+ | Covid) * P(Covid)	P(– | No Covid) * P(No Covid) + P(– | Covid) * P(Covid)

PS1, Q2

See the solutions for the rest of Q2

Right now, let's move on to a practice problem that is designed to help you with the concept of "updating your priors" as in Q1 on PS2...

Updating your priors

Bayesian updating

P(Cancer) ***Mammogram*** **P(Cancer | Mammogram result)**

From Handout 3, Section 4

Bayesian updating

In reality, we don't just do this once.

Bayesian updating is an *iterative* process!

P(Cancer)

Mammogram

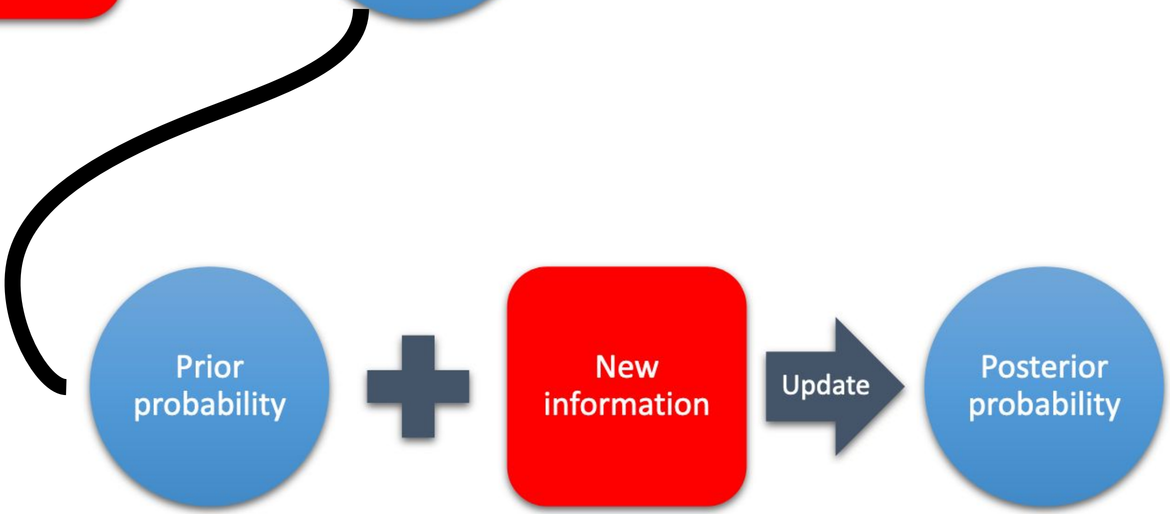
P(Cancer | Mammogram result)

From Handout 3, Section 4

Bayesian updating

When we incorporate new information using Bayes' rule, we get a "posterior probability"...

...this then becomes our new prior when we want to include another piece of information!



Practice problem

Harvard psychologist Amy Cuddy features in one of the most-viewed TED talks of all time, with over *60 million views*.

She describes her research on "**power posing**" – the idea that adopting a confident posture for 2 minutes can have significant effects on self-confidence and hormone levels.

In recent years, there has been a lively scholarly debate about whether power posing is *real* - does it actually have these effects, or was the initial study a fluke?

Let’s explore this from a Bayesian perspective!

Practice problem: Replication, Replication, Replication

How does this scenario relate to the more familiar case of medical testing?

Let's write:

- **power pose** to represent the case where power posing is real
- **¬power pose** for the case where power posing is not real
- **+** for a study on power posing that finds a significant effect
- **–** for a study that finds no significant effect

Practice problem: Replication, Replication, Replication

Before seeing any studies, my prior that power posing is real, P(power pose), is pretty low. (It's weird, right?)

I'm going to say **P(power pose) = 0.1**

The probability that a study finds a significant effect given that power posing is real, **P(+ | power pose) = 0.8**

The probability that a study finds no significant effect given that power posing isn't real, **P(– | ¬power pose) = 0.4**

Practice problem: Replication, Replication, Replication

What happens to my prior, P(power pose), after Cuddy et al.'s first study comes out?

$$P(power\ pose|+)$$

$$= \frac{P(+|power\ pose)P(power\ pose)}{P(+|power\ pose)P(power\ pose) + P(+|\neg power\ pose)P(\neg power\ pose)}$$

$$= \frac{0.8 \cdot 0.1}{0.8 \cdot 0.1 + (1 - 0.4) \cdot 0.9}$$

$$= 0.13$$

Practice problem: Replication, Replication, Replication

Before I read Cuddy et al.'s first study, I guess there was about a **10% chance** that power posing is real.

After I read Cuddy et al.'s first study, I have updated my prior *a little*: I now think there's a **13% chance** that power posing is real.

What happens if another study on power posing is published?

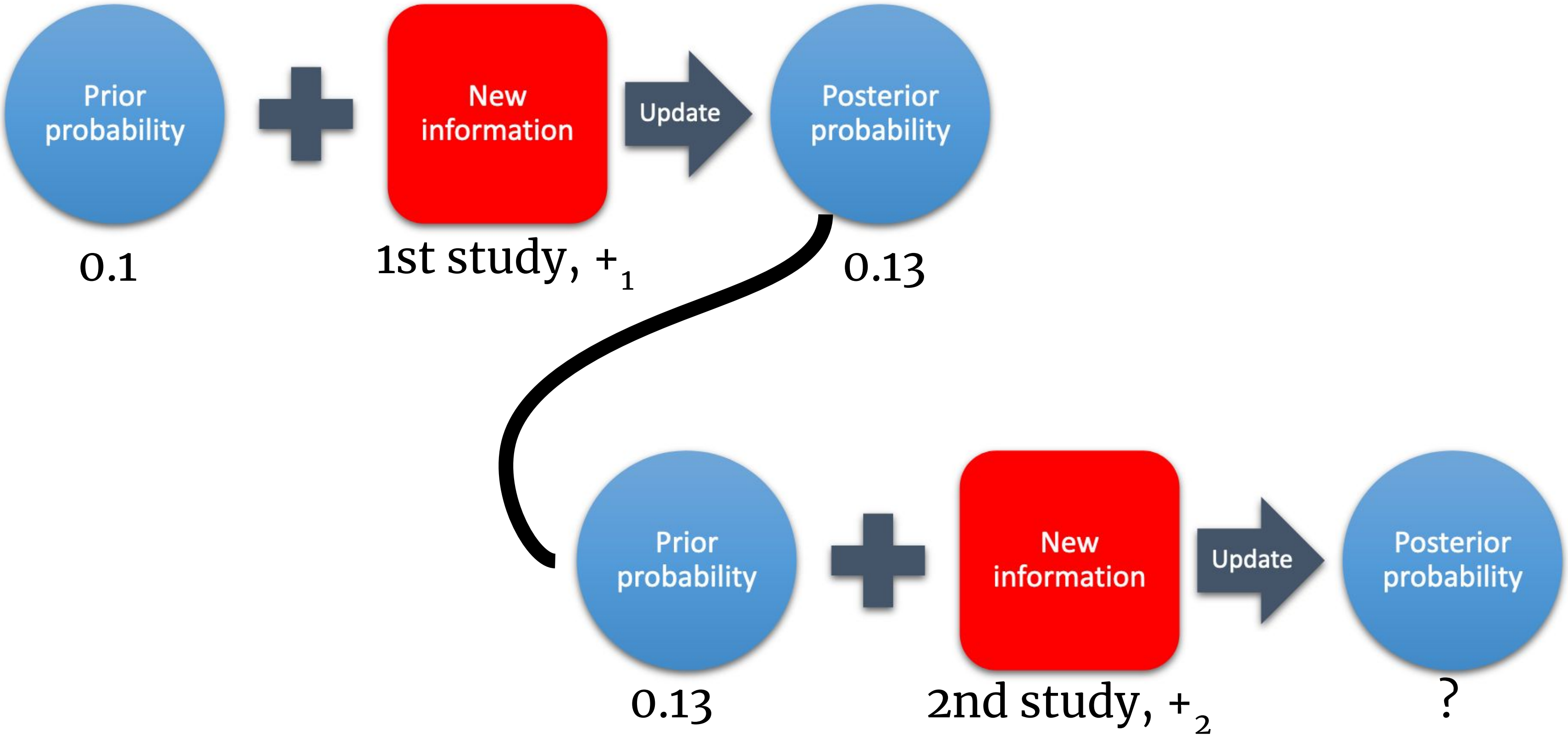
Practice problem: Replication, Replication, Replication

Suppose the 2nd study *also* finds a significant effect of power posing.

Now we want to find P(power pose | $+_2$), the probability that power posing is real given that the 1st study was successfully replicated.

To do this, we can use Bayes rule in exactly the same way as before, except our prior P(power pose) is updated to reflect the result of the 1st study…

Practice problem: Replication, Replication, Replication



Practice problem: Replication, Replication, Replication

$$P(power\ pose|+_2)$$

$$= \frac{P(+_2|power\ pose)P(power\ pose)'}{P(+_2|power\ pose)P(power\ pose)' + P(+_2|\neg power\ pose)P(\neg power\ pose)'}$$

$$= \frac{0.8 \cdot 0.13}{0.8 \cdot 0.13 + (1 - 0.4) \cdot 0.87}$$

Note: I have written P(power pose)’ to denote our updated prior, which is P(power pose | $+_1$).

Everything else in the formula is the same.

Practice problem: Replication, Replication, Replication

What if the 2nd study found *no effect* of power posing?

Practice problem: Replication, Replication, Replication

What if the 2nd study found *no effect* of power posing?

$$P(power\ pose|-_2)$$

$$= \frac{P(-_2|power\ pose)P(power\ pose)}{P(-_2|power\ pose)P(power\ pose) + P(-_2|\neg power\ pose)P(\neg power\ pose)}$$

$$= \frac{(1-0.8)\cdot 0.13}{(1-0.8)\cdot 0.13 + 0.4\cdot 0.87}$$

$$= 0.07$$

Practice problem: Replication, Replication, Replication

Before any studies, I thought that P(power pose) = 0.1

After one study that found an effect, and one that didn’t, I have actually downgraded the probability to just 0.07 = 7% !

Why?

Practice problem: Replication, Replication, Replication

Before any studies, I thought that P(power pose) = 0.1

After one study that found an effect, and one that didn't, I have actually downgraded the probability to just 0.07 = 7% !

Why?

It comes back to my assumptions:

$P(+ \mid \text{power pose}) = 0.8$

$P(- \mid \neg\text{power pose}) = 0.4$

Practice problem: Replication, Replication, Replication

Before any studies, I thought that P(power pose) = 0.1

After one study that found an effect, and one that didn't, I have actually downgraded the probability to just 0.07 = 7% !

Why?

It comes back to my assumptions:

P(+ | power pose) = 0.8

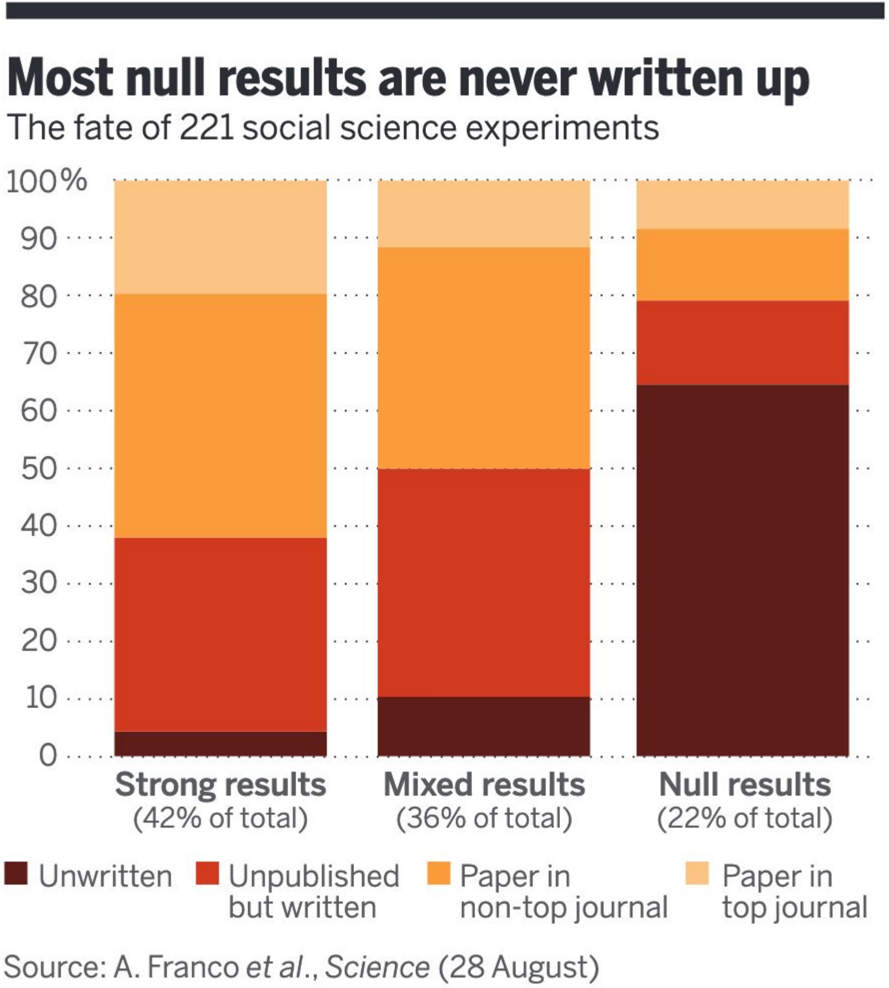
P(– | ¬power pose) = 0.4

Why did I make this probability so low?

The File Drawer Problem

Null results – i.e. studies that fail to find a predicted effect – are much harder to publish in academic journals.

This is a BIG problem and it affects the way we interpret the studies that do get published.



Power posing *might* have some temporary effects on self-confidence... but the idea that it changes your hormone levels has failed to replicate.

Conclusion: Bayes rule can even help us to understand the scientific replication crisis! Cool!!

Great job everyone!!