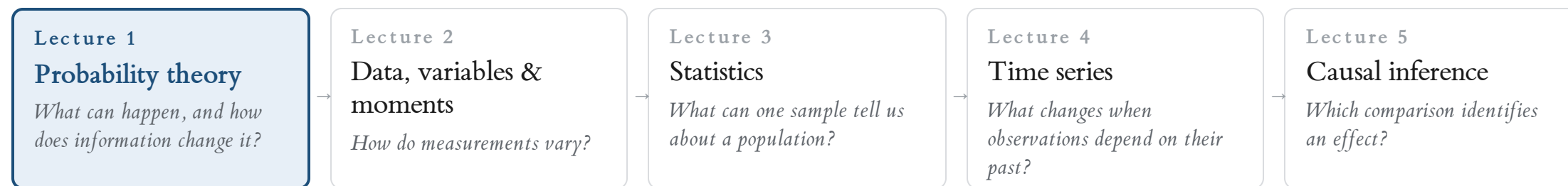
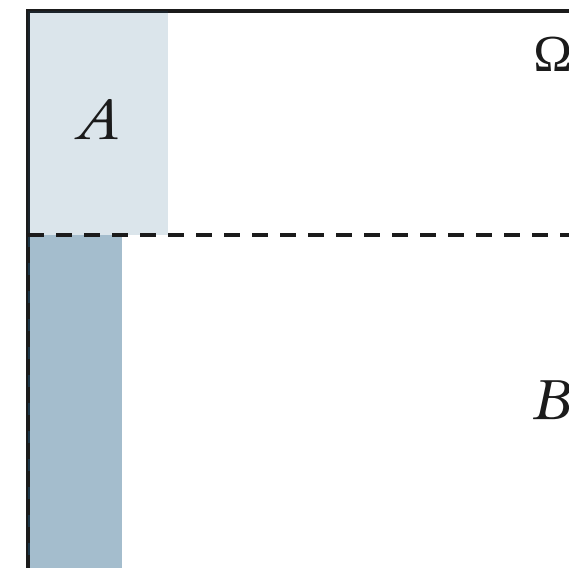


Probability theory

What can happen, and how should information change what we believe?

1 Describe → 2 Condition → 3 Learn



OPENING QUESTION

A cyclist tests positive. How likely is it that the cyclist uses EPO?

A hypothetical cycling population and an imperfect test:

15%

of cyclists use EPO

90%

of users test positive

12%

of non-users also test positive

YOUR ESTIMATE

Vote now and keep your answer. Part 3 calculates it.

A

$\approx 90\%$

B

$\approx 57\%$

C

$\approx 15\%$

WHY START HERE

Economists read signals backwards all the time: a test result, a credit score, a qualification. Today builds the rules that make that reading exact.

LECTURE 1 · THE ROUTE

Describe uncertainty, condition on information, learn from evidence

PART 1

Describe: events and their probabilities

What exactly could happen, and how do we give events probabilities that are consistent with one another?

sample space event axioms

union rule

Notes 1.1

PART 2

Condition: how information changes a probability

When we learn that one event has happened, how should the probability of another change?

conditional probability product rule

total probability

Notes 1.2

PART 3

Learn: independence and Bayes' rule

When does information leave a probability unchanged, and how should evidence change what we believe?

independence Bayes' rule prior

posterior base rate

Notes 1.3

BY THE END OF THIS LECTURE YOU CAN

- ✓ Write the sample space and events for an experiment, and derive the complement and union rules from the axioms.
- ✓ Calculate a conditional probability as a share within the information event, keeping track of which way it runs.
- ✓ Combine group rates by total probability, and check whether two events are independent.
- ✓ Update a prior with Bayes' rule, and explain why a positive EPO test means about 57%, not 90%.

PART 1 OF 3

Describe: events and their probabilities

What exactly could happen, and how do we give events probabilities that are consistent with one another?

KEY TERMS

sample space

event

axioms

union rule

[Notes 1.1 ↗](#)

WHY IT MATTERS

Every economic question about risk, from becoming unemployed to a loan default, starts by stating exactly which outcome is uncertain. The axioms keep the probabilities we attach consistent with one another, whatever the weights.

IN THIS PART

1.1.1–3 Outcomes, events and set operations

1.1.4 The axioms and the complement rule

1.1.5 The union rule

▶ Animation: where each term of the union rule comes from

1.1.1-3 | OUTCOMES AND EVENTS

Events are sets of outcomes, so “not”, “and”, “or” are set operations

DEFINITION

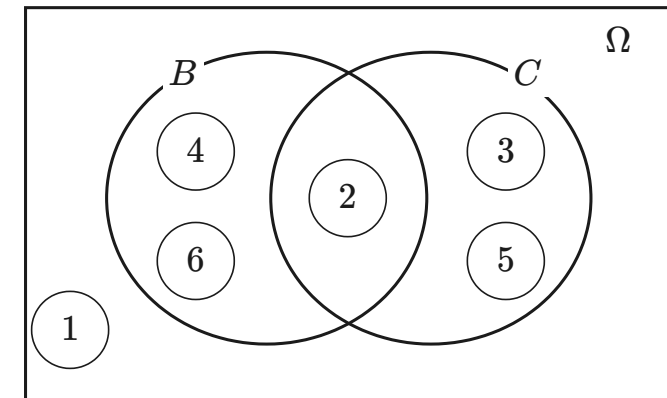
Sample space and events

An **outcome** ω is one possible result; the **sample space** Ω collects them all. An **event** $A \subseteq \Omega$ is the set of outcomes that make a statement true.

Statement	Event	One die: B even, C prime
not B	B^c	$\{1, 3, 5\}$
B and C	$B \cap C$	$\{2\}$
B or C	$B \cup C$	$\{2, 3, 4, 5, 6\}$

“Or” is inclusive. Events are **disjoint** when $A \cap B = \emptyset$.

A picture of “and”, “or” and “not”



1.1.4 | THE AXIOMS

Three axioms fix how probabilities combine, not what they are

DEFINITION

The axioms of probability

1. $P(A) \geq 0$ for every event: no negative mass.
2. $P(\Omega) = 1$: the total is fixed.
3. Disjoint events add: $P(\bigcup_i A_i) = \sum_i P(A_i)$ for pairwise disjoint A_i .

WORKED EXAMPLE *a fair and a loaded die*

Fair: $P(\text{even}) = 3 \times \frac{1}{6} = \frac{1}{2}$. Loaded, 0.1 on faces 1–5 and 0.5 on face 6: $P(\text{even}) = 0.1 + 0.1 + 0.5 = 0.7$. Both satisfy the axioms.

KEY FORMULA

RESULT

(1)

Complement rule

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

IN WORDS A and A^c are disjoint and fill Ω , so their probabilities add to 1.

COMMON MISTAKE

Probability zero is not impossibility: one exact bulb lifetime can have probability 0 without being impossible.

1.1.5 | ADDING PROBABILITIES

Adding two events counts their overlap twice

KEY FORMULA

RESULT

(2)

Union rule

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

IN WORDS add both, then remove the overlap counted twice.

NEEDS any two events; when they are disjoint the last term is zero.

CHECK

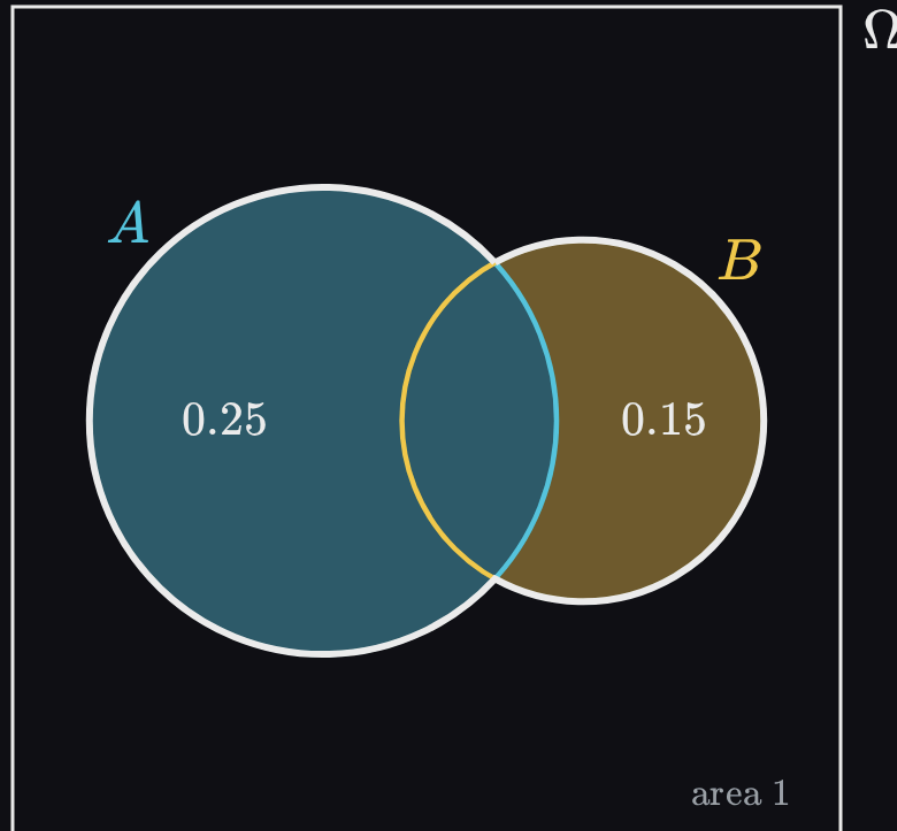
Fair die, B even, C prime. What is $P(B \cup C)$?

ANSWER

$\frac{3}{6} + \frac{3}{6} - \frac{1}{6} = \frac{5}{6}$. The plain sum, 1, counts the face 2 twice.

Disjointness lets you add; overlap must be subtracted once.

Where does each term of $P(A \cup B)$ come from?



$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$0.35 = 0.25 + 0.15 - 0.05$$

PART 2 OF 3

Condition: how information changes a probability

When we learn that one event has happened, how should the probability of another change?

KEY TERMS

conditional probability

product rule

total probability

[Notes 1.2 ↗](#)

SO FAR a model $(\Omega, \mathcal{F}, P)$, three axioms, and the complement and union rules.

WHY IT MATTERS

Economic information usually arrives as rates within groups: defect rates per production line, unemployment among graduates, default among borrowers with a given score. Answering the question asked needs the right reference group, and recombining groups needs their sizes as weights.

IN THIS PART

- 1.2.1 Two children: information and the reference group
- 1.2.2 Conditional probability: restrict, then rescale
- ▶ Animation: why conditioning divides by the probability of the information
- 1.2.3 Reversing the condition
- 1.2.4 Total probability: a weighted average of group rates
- ▶ Animation: total probability as a weighted average

1.2.1 | INFORMATION CHANGES THE QUESTION

Learning something changes the group in which we calculate

Two children, each a boy or a girl with probability $\frac{1}{2}$, independently, older child first:

$$\Omega = \{(b, b), (b, g), (g, b), (g, g)\}$$

Each family has probability $\frac{1}{4}$. Let A be “two boys” and B “at least one boy”.

PREDICT FIRST

We learn that B is true. Are two boys now more or less likely than $\frac{1}{4}$?

ANSWER

More: (g, g) is ruled out, so three equally likely families remain and $P(A | B) = \frac{1}{3}$.

TAKEAWAY No family changes. What changes is the group in which we calculate the share.

1.2.2 | THE DEFINITION

Conditional probability: restrict to B , then rescale B to one

KEY FORMULA DEFINITION

(3)

Conditional probability

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

IN WORDS the share of B that also lies in A : **conditional probability**.

NEEDS $P(B) > 0$.

1. Restrict to B : only $A \cap B$ survives.
2. Rescale by $1/P(B)$: the new total is 1.

KEY FORMULA IDENTITY

(4)

Product rule

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

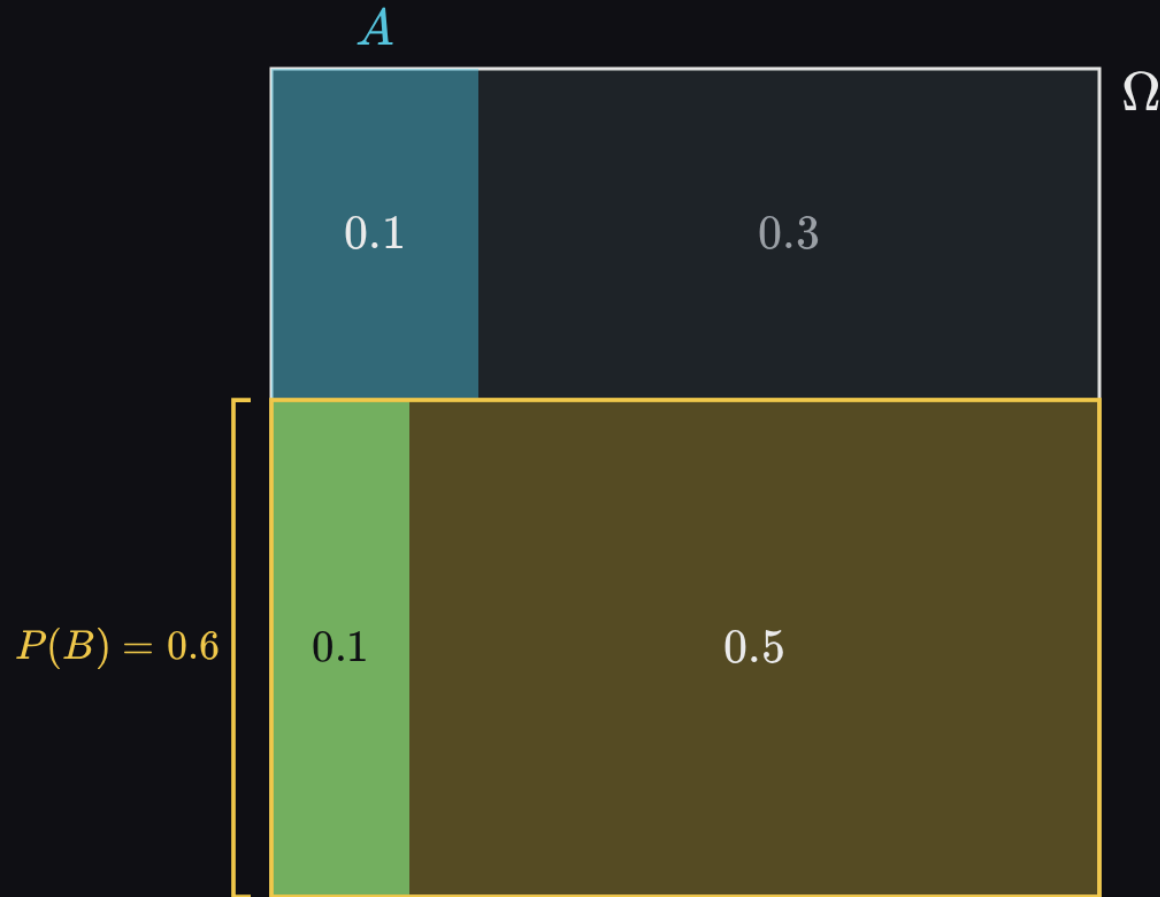
IN WORDS the definition turned around: a share of a share.

WORKED EXAMPLE

$P(B) = 0.6$ and $P(A \cap B) = 0.1$, so

$$P(A \mid B) = \frac{0.1}{0.6} = \frac{1}{6}, \quad P(A^c \mid B) = \frac{5}{6}.$$

Conditioning on B : why divide by $P(B)$?



$$P(A) = 0.2, \quad P(B) = 0.6, \quad P(A \cap B) = 0.1$$

Learn that B happened:
everything outside B is ruled out

B becomes the whole: each part $\times 1/0.6$

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{0.1}{0.6} = \frac{1}{6}$$

Shrink back: multiply by $P(B)$

$$P(A \cap B) = P(A \mid B) P(B) = 1/6 \times 0.6 = 0.1$$

1.2.3 | DIRECTION MATTERS

Reversing the condition changes the denominator

100 hypothetical cases

	B	B^c	Total
A	10	10	20
A^c	50	30	80
Total	60	40	100

Within column B : $P(A | B) = 10/60 = \frac{1}{6}$ Within row A : $P(B | A) = 10/20 = \frac{1}{2}$

COMMON MISTAKE

Treating $P(A | B)$ and $P(B | A)$ as the same number.They share a numerator, $P(A \cap B)$, but divide by different groups.

READING CLAIMS

“Most students are Democrats” does not say that most Democrats are students. Read the condition first: it names the group you divide by.

1.2.4 | RECOVERING THE WHOLE

Total probability weights each group's rate by the group's size

KEY FORMULA **RESULT**

(5)

Law of total probability

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

IN WORDS each group's rate times its share, added over the groups.

NEEDS groups B_i that cover Ω without overlap (a **partition**), each $P(B_i) > 0$.

CHECK

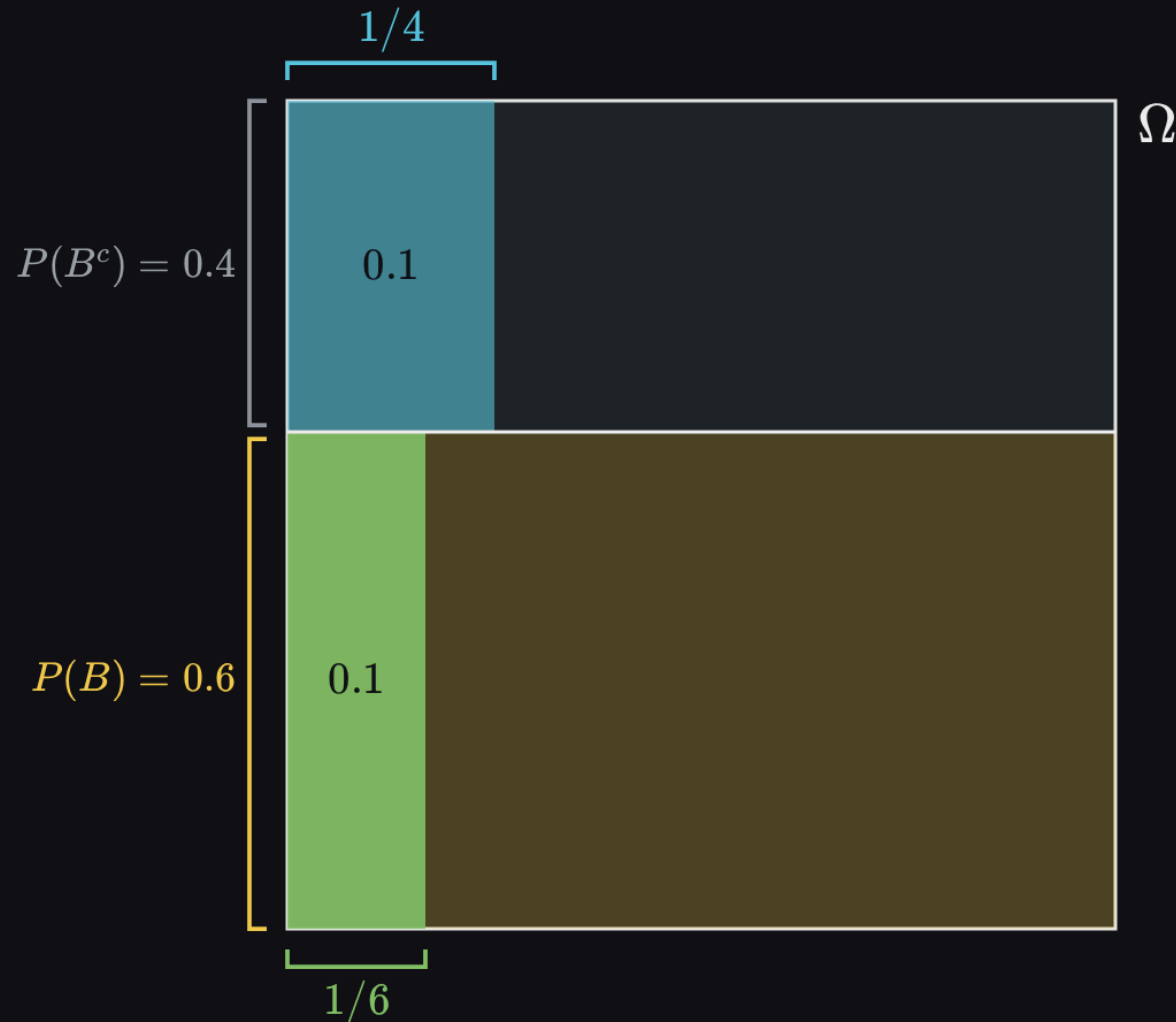
Line 1 makes 40% of items with defect rate 0.05; line 2 makes 60% with rate 0.10. What share of all items is defective?

ANSWER

$0.05(0.4) + 0.10(0.6) = 0.02 + 0.06 = 0.08$. The plain average, 0.075, would treat the lines as equal in size.

WORKED EXAMPLE *the 100 cases*

$$P(A) = \frac{1}{6}(0.6) + \frac{1}{4}(0.4) = 0.1 + 0.1 = 0.2$$

Recovering $P(A)$ from two groups

A partition: every case is in B or B^c

Rates: $P(A | B) = 1/6$, $P(A | B^c) = 1/4$

$$P(A) = P(A | B) P(B) + P(A | B^c) P(B^c)$$

$$= 1/6 \times 0.6 + 1/4 \times 0.4$$

$$= 0.1 + 0.1 = 0.2$$

If the two groups were equal in size:

$$(1/6 + 1/4)/2 \approx 0.208, \text{ not } 0.2$$

The group sizes are the weights.

PART 3 OF 3

Learn: independence and Bayes' rule

When does information leave a probability unchanged, and how should evidence change what we believe?

KEY TERMS

independence

Bayes' rule

prior

posterior

base rate

[Notes 1.3 ↗](#)

SO FAR conditioning restricts to the information event and rescales it; total probability weights group rates by group sizes.

WHY IT MATTERS

Decisions rest on imperfect signals read backwards: a positive drug test, a credit score, rare evidence in court. Bayes' rule makes the reversal exact, and independence says when information is irrelevant and when probabilities may be multiplied.

IN THIS PART

- 1.3.1 Independence: an unchanged share
- ▶ Animation: independence as an unchanged share
- 1.3.2 Bayes' rule: one joint probability, two directions
- 1.3.3 The EPO test, step by step
- ▶ Animation: every number in the posterior
- 1.3.4–5 Base rates and the Sally Clark case

1.3.1 | INDEPENDENCE

Independent events: learning B leaves the probability of A unchanged

KEY FORMULA DEFINITION

(6)

Independence

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

IN WORDS equivalently $P(A | B) = P(A)$: the share of A inside B equals its share overall.

NEEDS the equivalent form needs $P(B) > 0$.

Independence is a fact about overlap, checked by comparing probabilities.

WORKED EXAMPLE

$P(A) = 0.30$ and $P(B) = 0.40$. Independence needs $P(A \cap B) = 0.30 \times 0.40 = 0.12$, and then

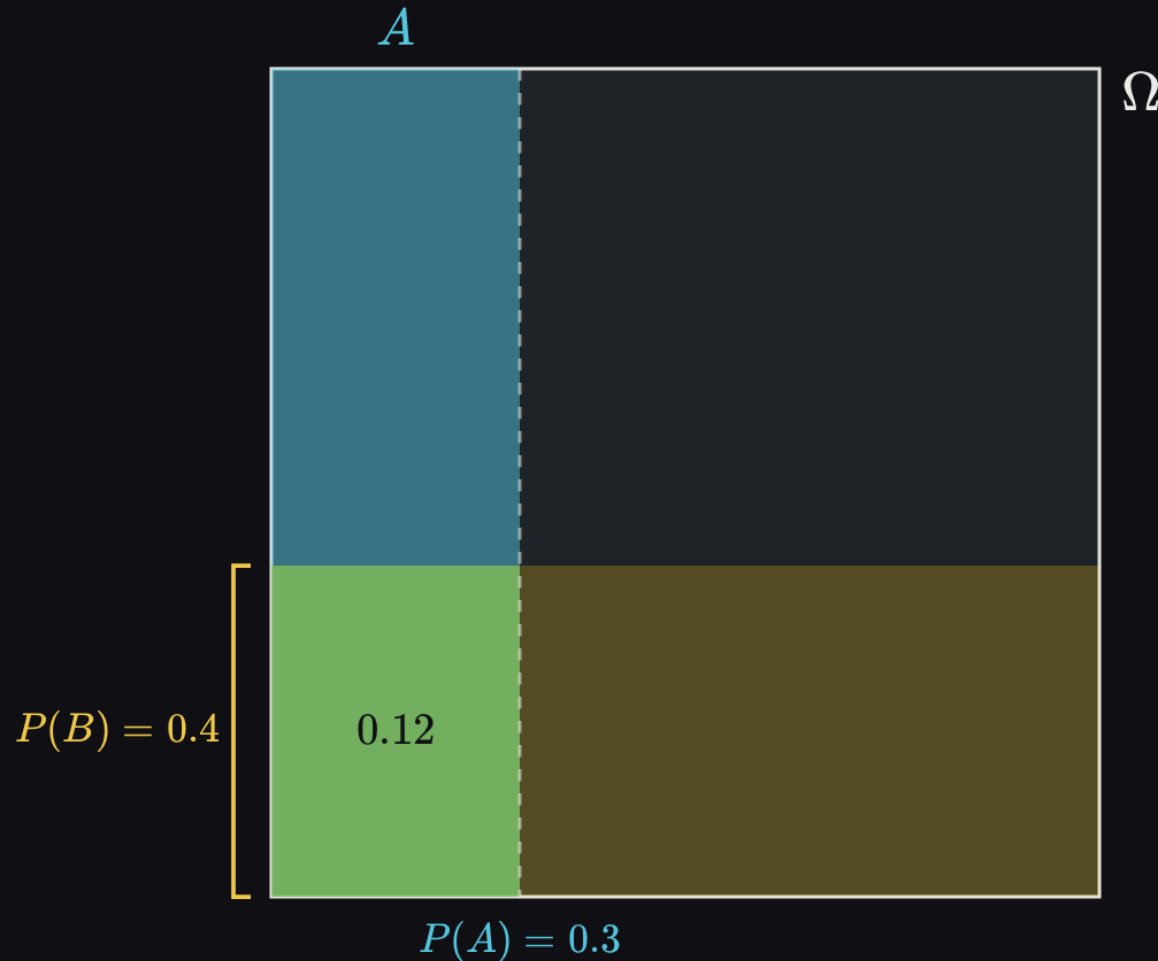
$$P(A | B) = \frac{0.12}{0.40} = 0.30 = P(A).$$

COMMON MISTAKE

Disjoint is not independent. If $A \cap B = \emptyset$ and both have positive probability, learning B rules A out:

$$P(A | B) = 0 \neq P(A).$$

Independence: does learning B change the chance of A ?



$$P(A) = 0.3, \quad P(B) = 0.4$$

Learn B : $P(A | B) = 0.12/0.4 = 0.3$
the same as $P(A)$: B tells us nothing

If more of A lay inside B :

$$P(A | B) = 0.5 > 0.3: A \text{ more likely}$$

No overlap: $P(A | B) = 0$: A ruled out

Independent only when the shares match:

$$P(A \cap B) = P(A) P(B) = 0.3 \times 0.4 = 0.12$$

1.3.2 | REVERSING A CONDITION

Bayes' rule rebuilds the reverse condition from the known one

KEY FORMULA

RESULT

(7)

Bayes' rule

$$P(H \mid E) = \frac{P(E \mid H) P(H)}{P(E)}$$

IN WORDS write $P(H \cap E)$ both ways with (4), equate, divide by $P(E)$.

NEEDS $P(E) > 0$; the denominator comes from (5).

DEFINITION

Prior, likelihood, posterior

Prior $P(H)$: before the evidence.

Likelihood $P(E \mid H)$: how readily H produces E .

Posterior $P(H \mid E)$: after the evidence.

Evidence favours H only if it is more likely under H than under the alternatives.

1.3.3 | WORKED EXAMPLE

A positive test: count both routes to a positive result

D : uses EPO, $P(D) = 0.15$. $+$: positive test, with $P(+ | D) = 0.90$ and $P(+ | D^c) = 0.12$.

WORKED EXAMPLE

1. Users who test positive: $0.90 \times 0.15 = 0.135$
2. Non-users who test positive: $0.12 \times 0.85 = 0.102$
3. All positives, by (5): $0.135 + 0.102 = 0.237$
4. By (7): $P(D | +) = 0.135/0.237 \approx 0.57$

In 1,000 hypothetical cyclists

	+	−	Total
Users	135	15	150
Non-users	102	748	850
Total	237	763	1,000

Row of users: $135/150 = 0.90$.

Column of positives: $135/237 \approx 0.57$.

TAKEAWAY 0.90 starts with users; 0.57 starts with positive tests.

A positive test: where does $P(D \mid +) \approx 0.57$ come from?

D	D^c	Ω
237 positive tests		
135	102	

D : uses EPO $+$: positive test, **outlined**

150 = 15% of 1,000

135 = 90% of 150

102 = 12% of 850

$$P(D \mid +) = \frac{135}{135 + 102} \approx 0.57$$

not the same as $P(+ \mid D) = 135/150 = 0.90$

The same 135, compared with a different group.

1.3.4–5 | EVIDENCE AND ASSUMPTIONS

Rare evidence is not a small posterior: base rates and assumptions decide

DEFINITION**Base rate**

The prior share of a condition in the population. With only 1% users, the same test gives

$$\frac{0.01(0.90)}{0.01(0.90) + 0.99(0.12)} \approx 7\%.$$

WORKED EXAMPLE *the Sally Clark case*

An argument squared a single-death frequency of sudden infant death: $(1/8,500)^2 = 1/72,250,000$.

TWO SEPARATE ERRORS

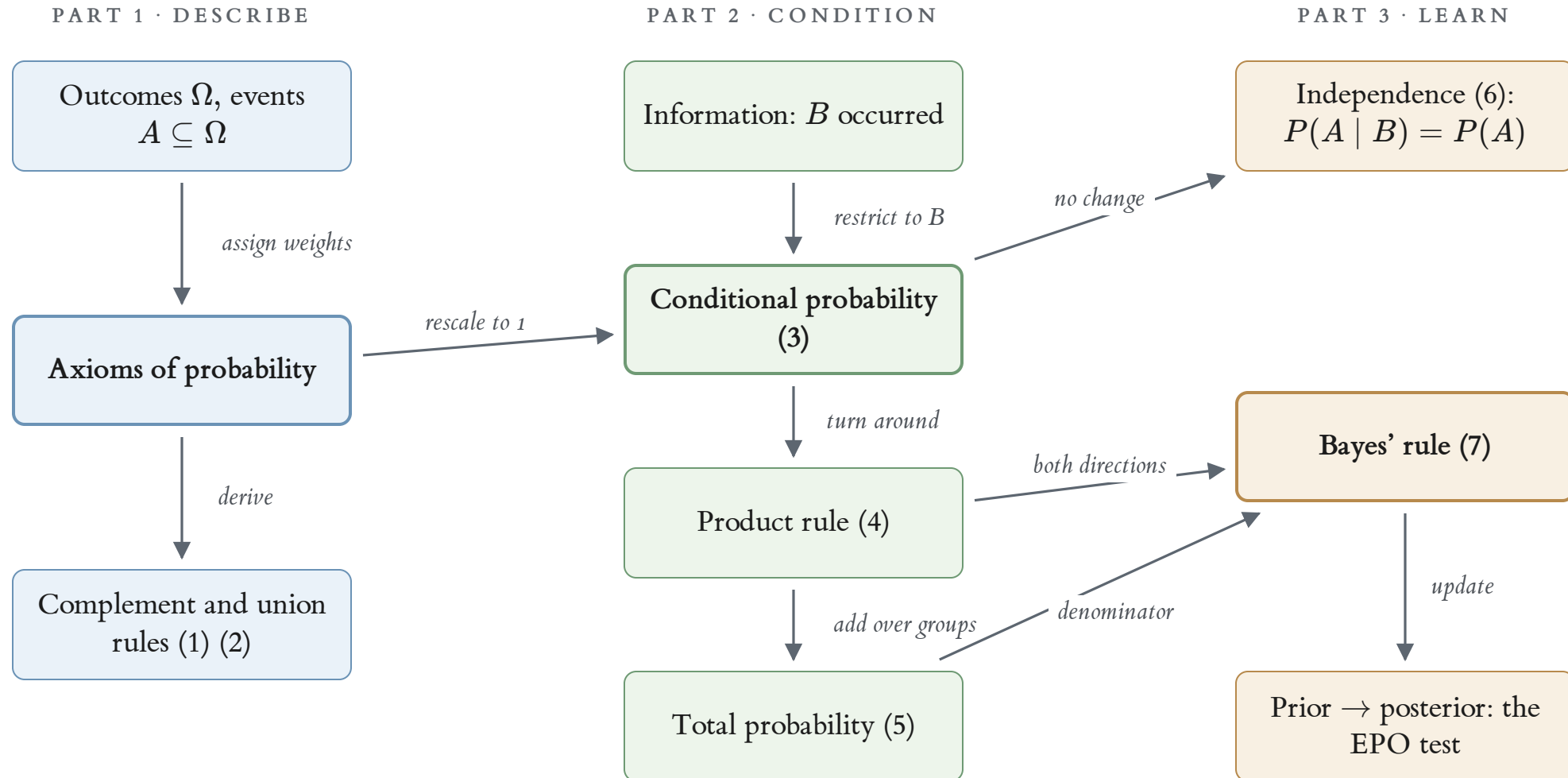
1. **Unsupported independence.** Squaring needs $P(S_2 | S_1) = P(S_2)$; a shared family risk makes it larger.
2. **The prosecutor's fallacy.** A small $P(E | H)$ is not a small $P(H | E)$: rival explanations must be compared, with their priors.

WHY ECONOMISTS CARE

The same logic governs screening, fraud detection and credit scoring: a signal's meaning depends on how common each type is.

REVIEW · HOW THE IDEAS CONNECT

One model, three moves: add, condition, reverse



REVIEW · SUMMARY SHEET

What to take away: the terms and the seven formulas

KEY TERMS

1 DESCRIBE

Outcome ω : one possible result.

Sample space Ω : the set of all outcomes the model considers.

Event $A \subseteq \Omega$: the outcomes that make a statement true.

Disjoint $A \cap B = \emptyset$: no outcome in both.

The axioms of probability nonnegativity, $P(\Omega) = 1$, and disjoint events add.

2 CONDITION

Conditional probability $P(A | B)$: the probability of A within the information event B .

Partition disjoint groups B_i that together cover Ω .

3 LEARN

Independence $P(A \cap B) = P(A)P(B)$: learning one event leaves the other's probability unchanged.

Prior $P(H)$: belief in a hypothesis before the evidence.

Likelihood $P(E | H)$: how readily H produces the evidence.

Posterior $P(H | E)$: belief after the evidence.

Base rate the prior share of a condition in the population tested.

KEY FORMULAS

1 DESCRIBE

(1) **Complement rule** RESULT $P(A^c) = 1 - P(A)$

(2) **Union rule** RESULT $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
any two events

2 CONDITION

(3) **Conditional probability** DEFINITION $P(A | B) = P(A \cap B) / P(B)$
 $P(B) > 0$.

(4) **Product rule** IDENTITY $P(A \cap B) = P(A | B) P(B)$

(5) **Law of total probability** RESULT $P(A) = \sum_i P(A | B_i) P(B_i)$
a partition $B_1, B_2, \dots$ with each $P(B_i) > 0$

3 LEARN

(6) **Independence** DEFINITION $P(A \cap B) = P(A) P(B)$
the definition; equivalently $P(A | B) = P(A)$ if $P(B) > 0$

(7) **Bayes' rule** RESULT $P(H | E) = P(E | H) P(H) / P(E)$
 $P(E) > 0$, with $P(E)$ from (5)

REVIEW · BACK TO THE OPENING QUESTION

The cyclist: about 57%, not 90%

RESULT *the opening question*

0.90 is $P(+ \mid D)$: it starts with users. We wanted the reverse, which starts with positive tests:

$$P(D \mid +) = \frac{0.135}{0.237} = \frac{45}{79} \approx 0.57$$

The test moved the probability from 15% to about 57%: strong evidence, far from certainty.

REMEMBER

1. Name the outcomes and the event before calculating.
2. Every conditional names a group: divide by its size.
3. Disjointness lets you add; independence lets you multiply.

BEFORE NEXT TIME

- Read notes 1.1–1.3 and the formula reference.
- Exercises 1.1.6, 1.2.5 and 1.3.6, with worked solutions.
- Watch the four animations not shown today.

NEXT · LECTURE 2

Data, variables & moments. From events to measurements: random variables, distributions, expectation and variance.