Probability and Statistics

Dhruva V. Raman

What's the point of data science/AI?

My feeling

1. Summarising data in an understandable way

It's a feeling

Summarising data

EG rate of
heart attacks

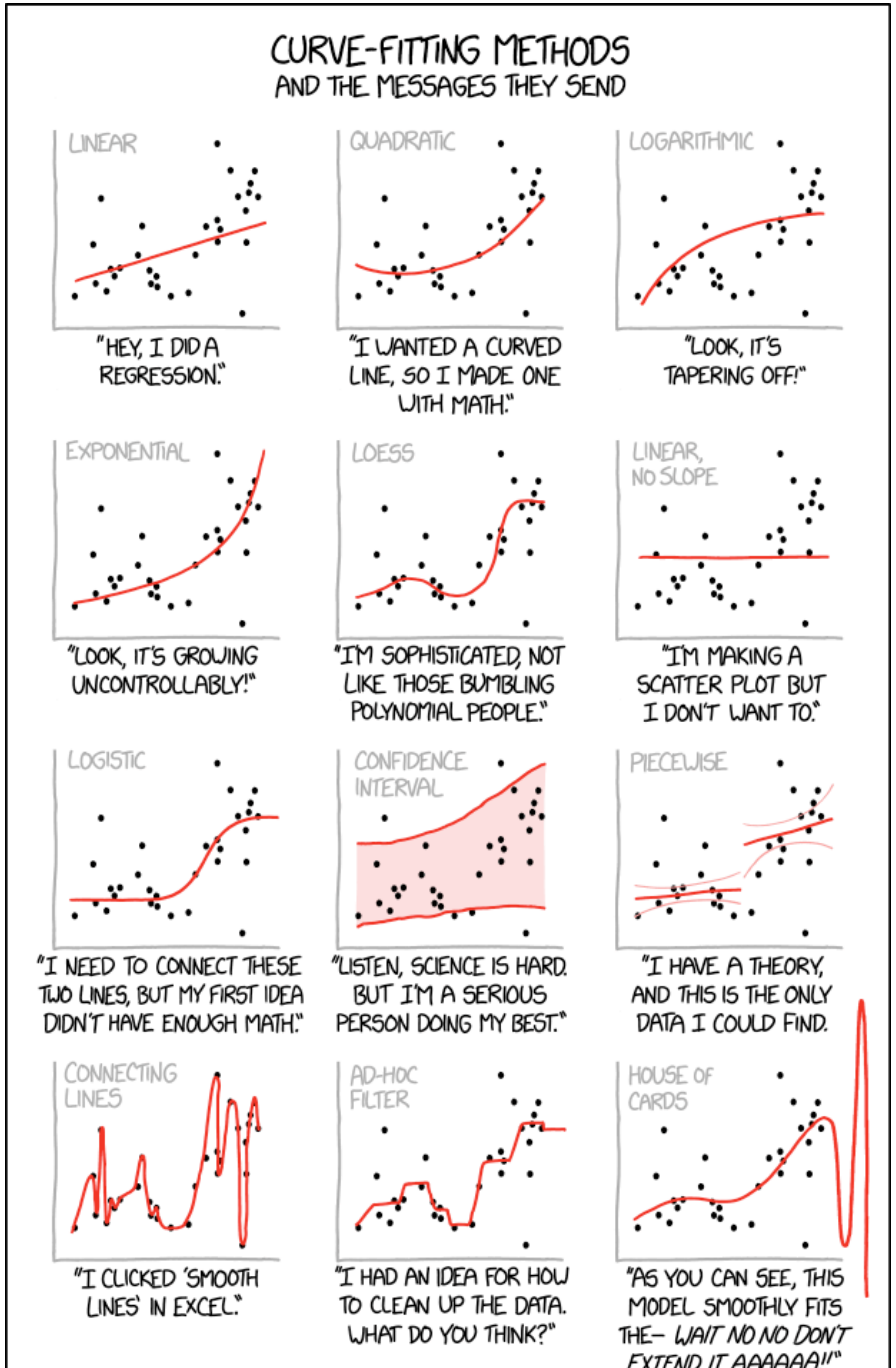
EG air pollution

EG dots are cities

Summarising data

Summaries can be wrong

Convoluted summaries
can be useless



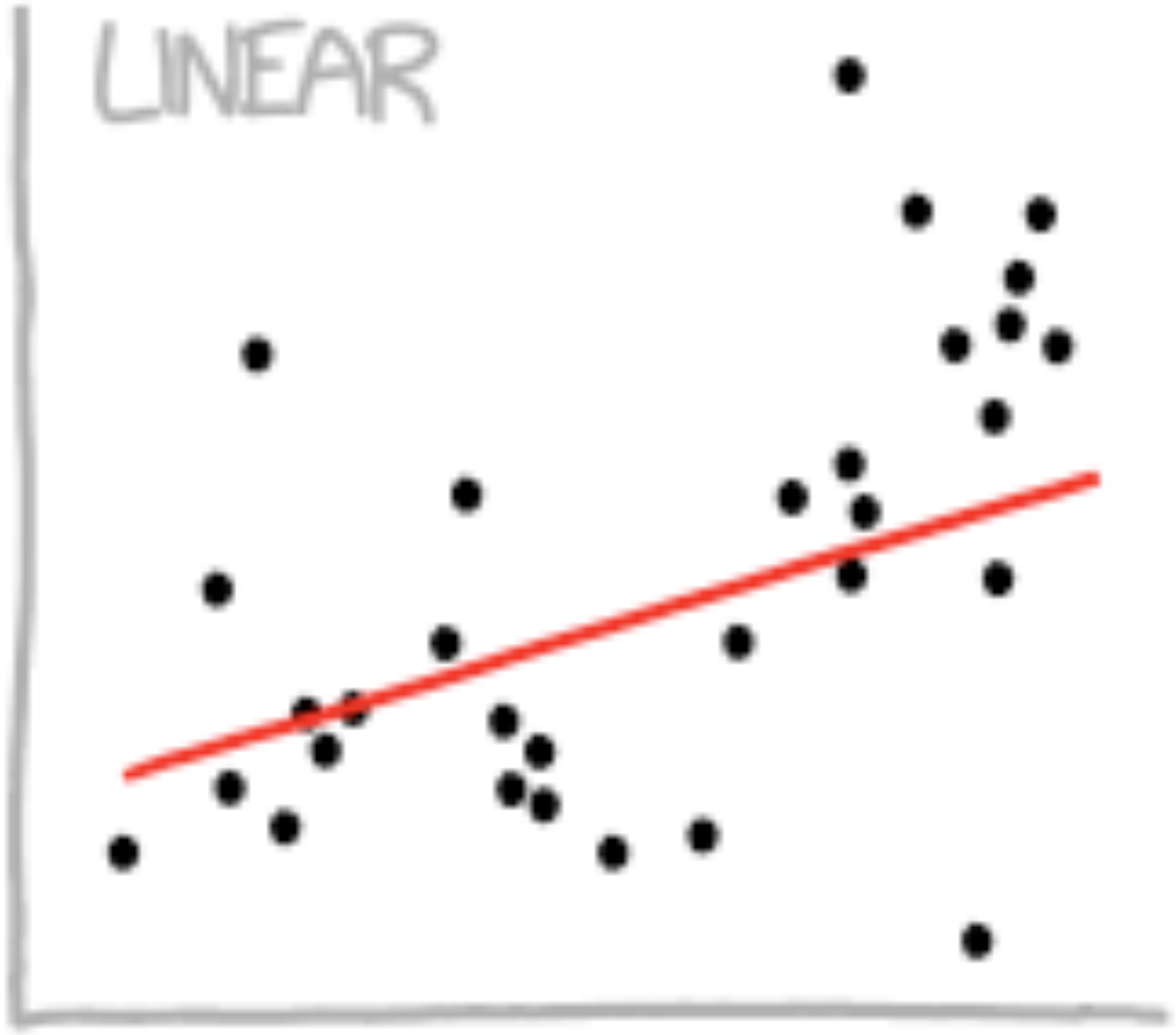
What's the point of data science/AI?

My feeling

1. Summarising data in a conceptually tractable way

2. Using these summaries to make (actionable) predictions

rate of heart attacks



air pollution

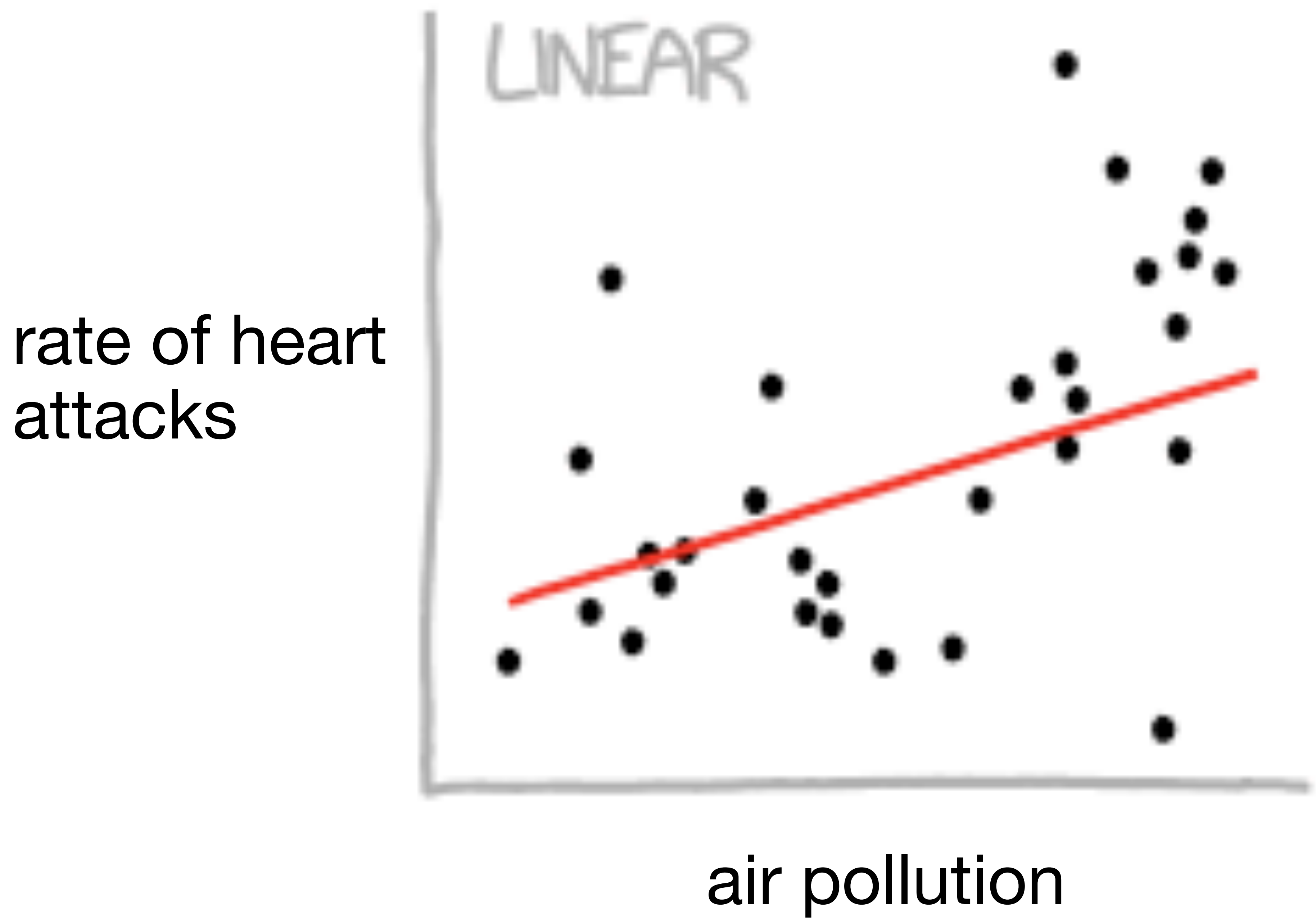
Congestion charge -> Less pollution -> Save NHS £50 million

Key Question

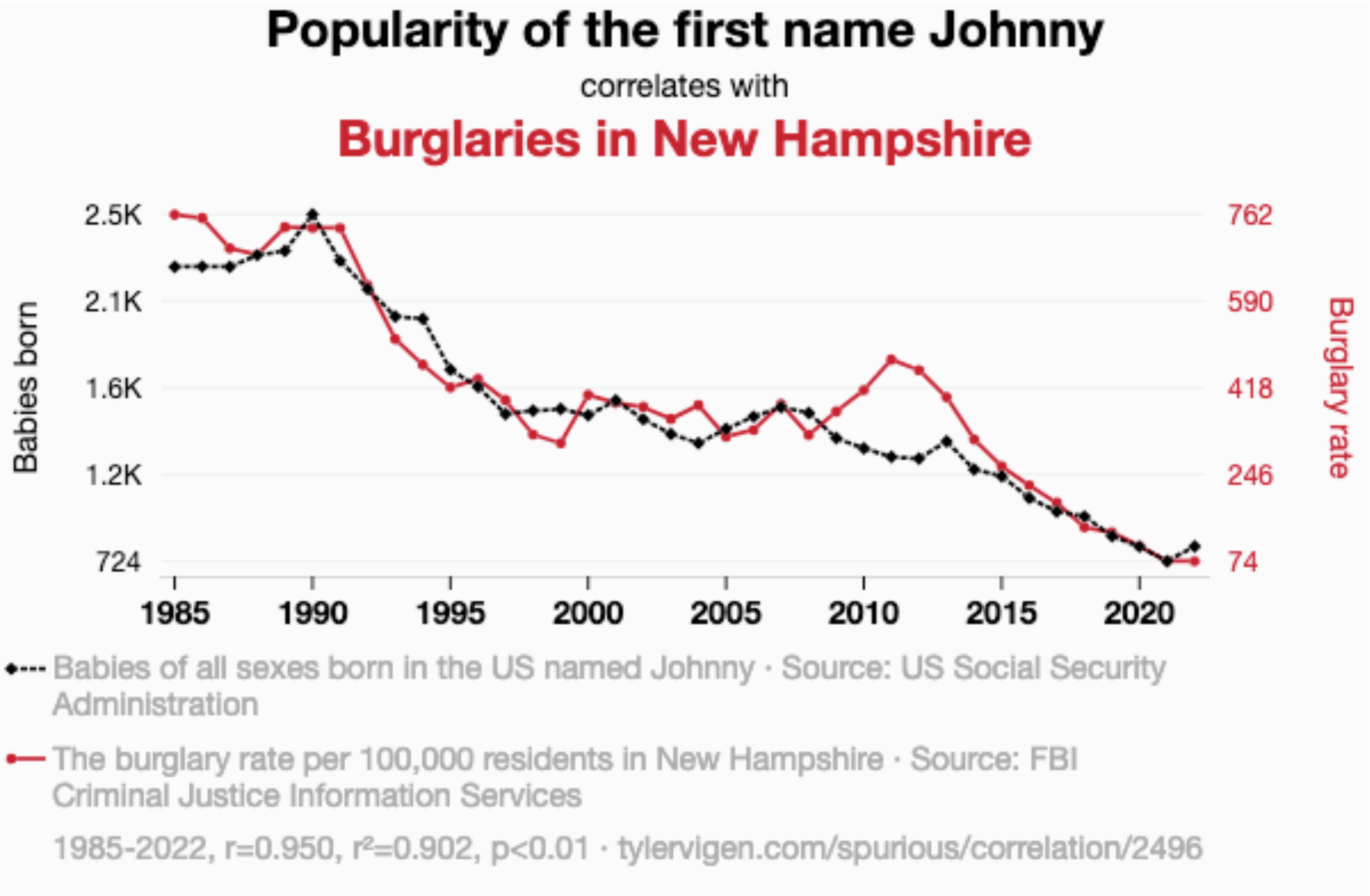
How confident am I on my summaries/predictions?

1. Summarising data in a conceptually tractable way

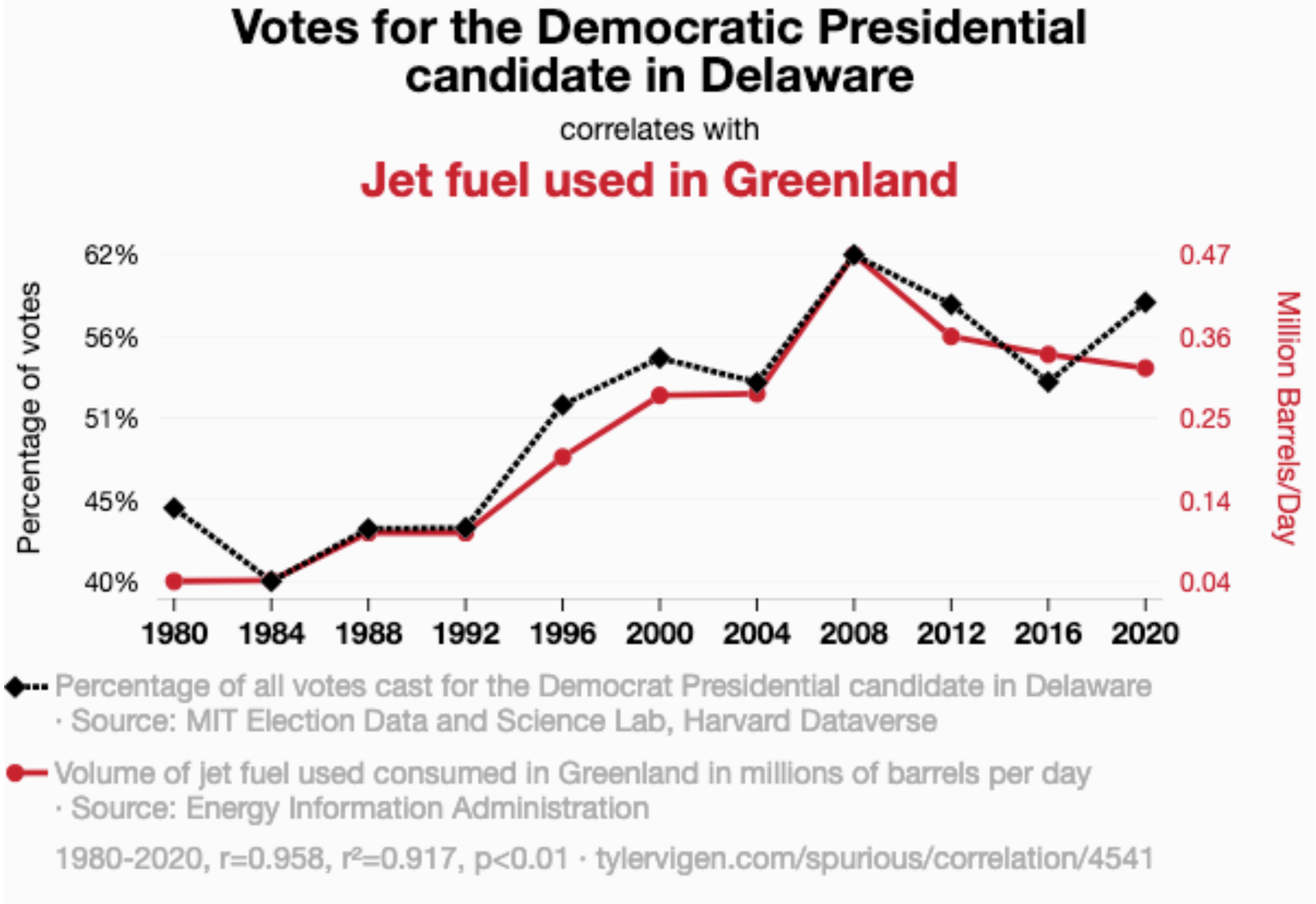
2. Using these summaries to make (actionable) predictions



Congestion charge -> Less pollution
-> Save NHS £50 million



Source: https://tylervigen.com/spurious/correlation/



Source: https://tylervigen.com/spurious/correlation/

Data scientists are **not** domain experts

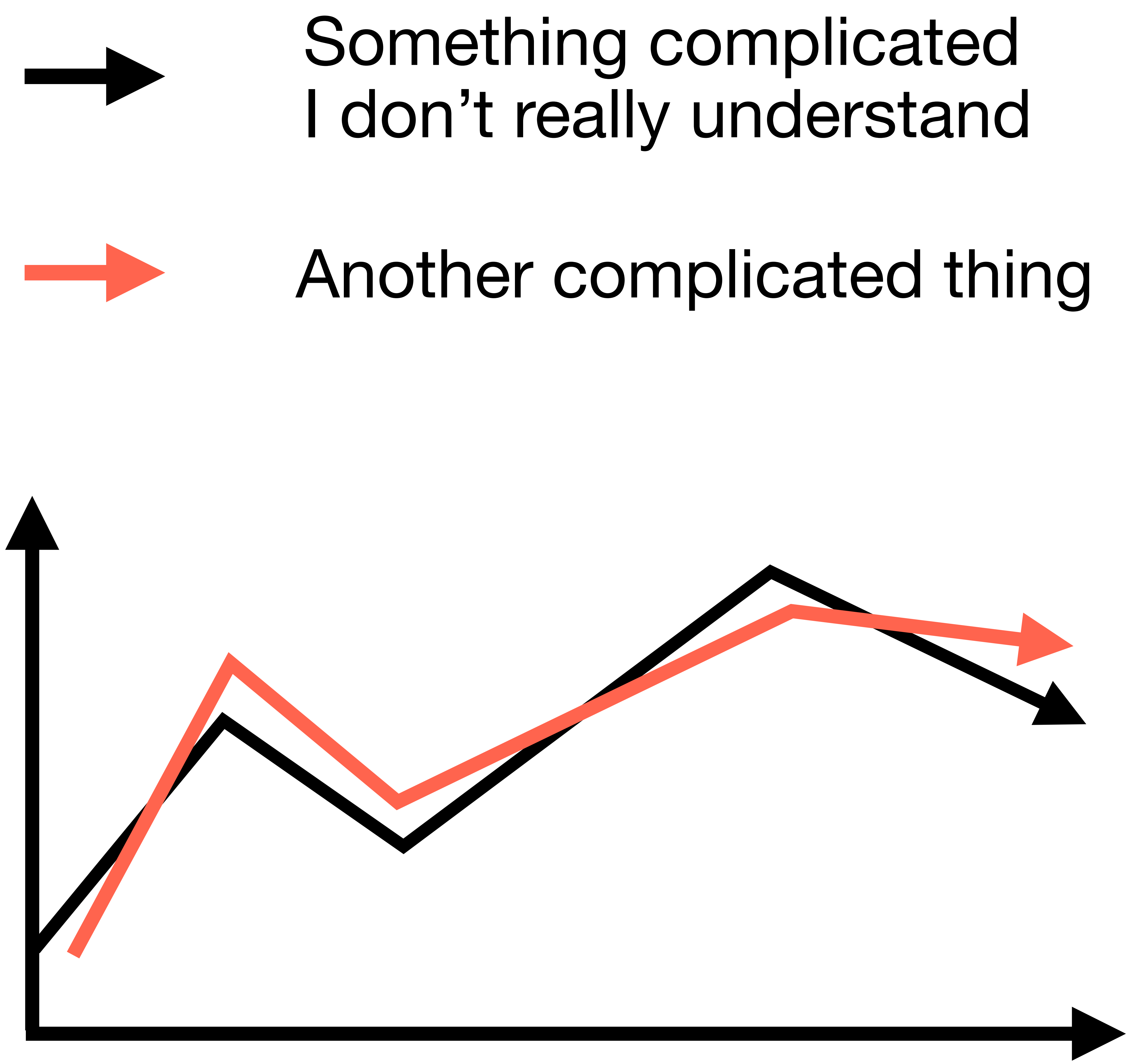
(Unfortunately)

Meaningful relationship?

Is the correlation 'silly'? I don't know!

Statistics gives an answer

Sometimes it's even correct!



Experiments

A process with
uncertain outcomes

Data scientist
Measure population’s credit card scores

Climate scientist
Measure tomorrow’s weather

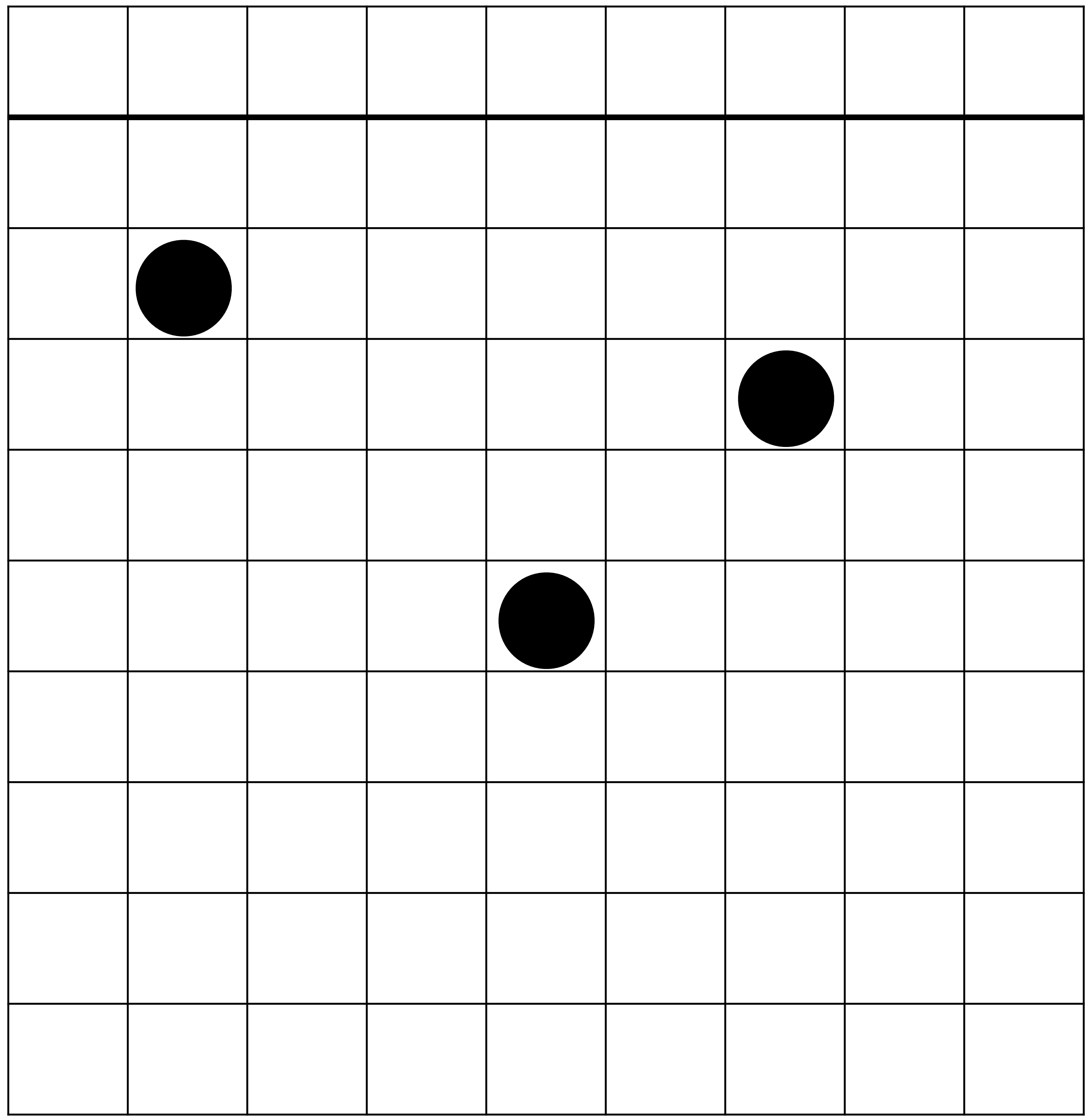
and they all need…

Mathematical framework to deal with
characterising and quantifying
uncertain events

(Probability theory)

Lecture theatre next week

● = Empty seat



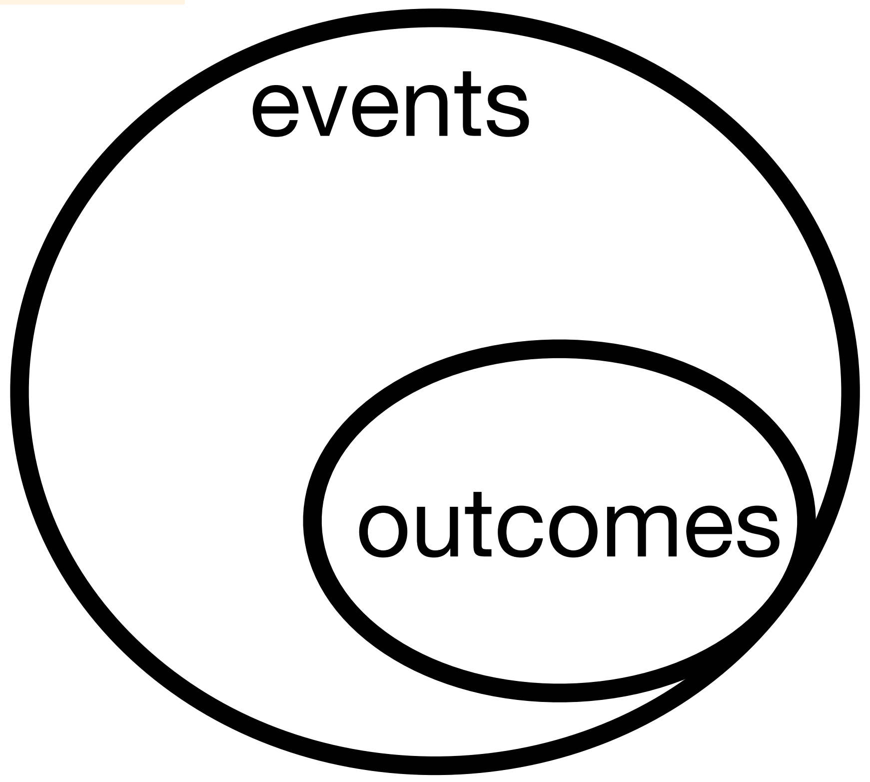
?

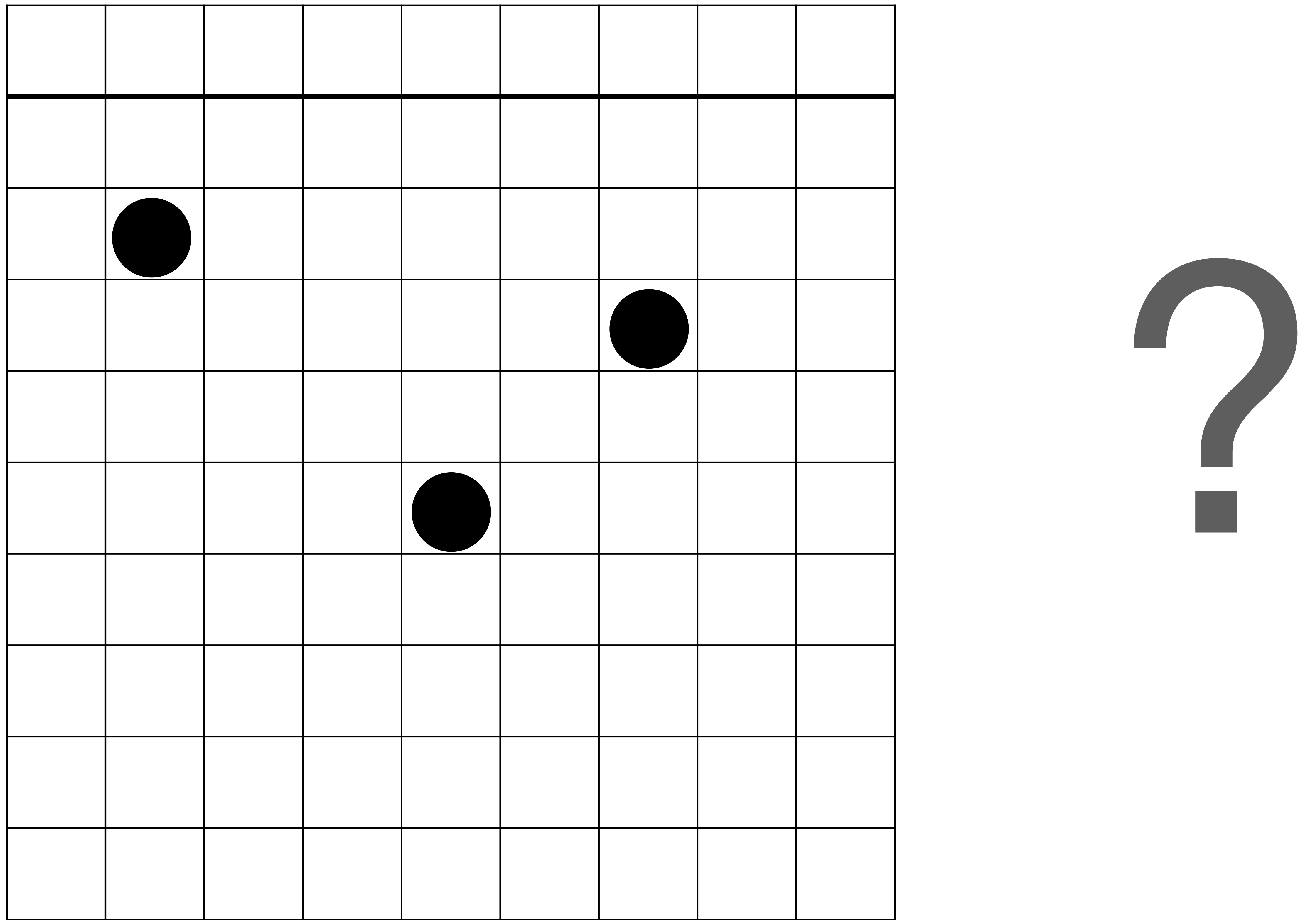
Lecture theatre next week

Formal terminology

Each seating combination is an outcome

Combinations of outcomes are events



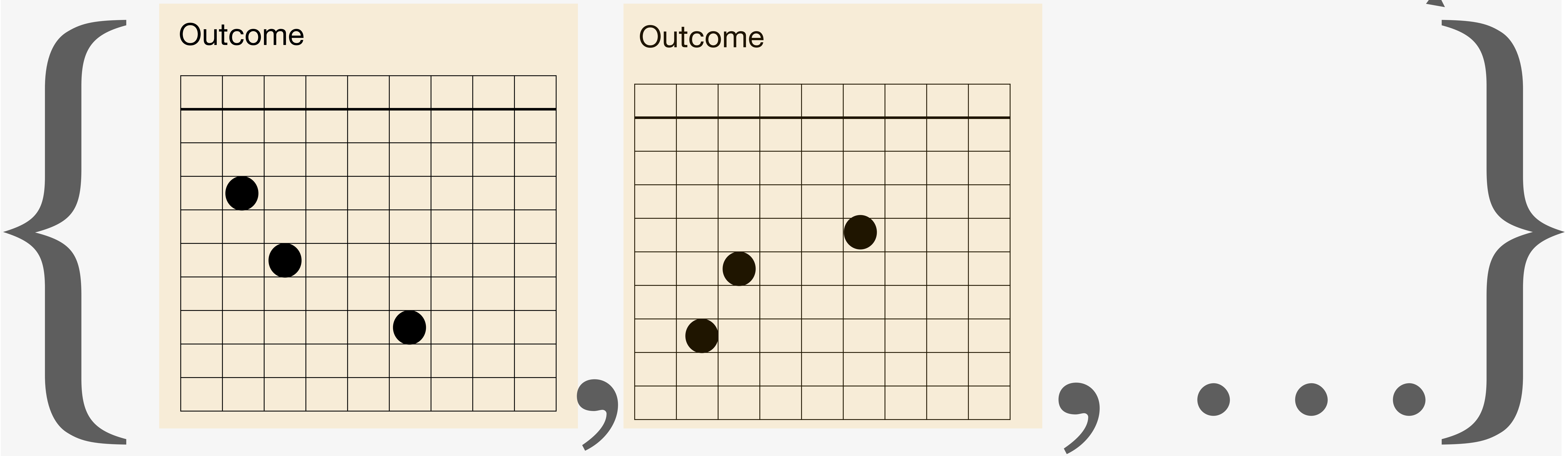


Events are sets of outcomes

i.e. collections

Curly braces denote sets in maths notation

Event: the back row is completely filled



Events have a set algebra

Algebra

Rules for elements to interact with each other

…and make babies!

Numbers

$$x + y = z$$
$$x - y = z$$
$$x * y = z$$

Also a number!

"plus, minus, times"

Events have a set algebra

Algebra

Rules for elements to interact with each other

…and make babies!

Events

$X \cup Y = Z$	**Union (or)**
$X \cap Y = Z$	**Intersection (and)**
$X \backslash Y = Z$	**Complement (not/without)**

Questions for the audience

Event X: the back row is completely filled

Event Y1: the front row is completely filled

Event Y2: the back row is partially filled

Express (+ simplify) as English sentences:

$X \cap Y_1$ $\quad$ $X \cap Y_2$

$X \cup Y_1$ $\quad$ $X \cup Y_2$

$X \backslash Y_1$ $\quad$ $X \backslash Y_2$

$\cup$ = 'or'

$\cap$ = `and'

$\backslash$ = `without'

Questions for the audience

Event X: the back row is completely filled

Event Y1: the front row is completely filled

Event Y2: the back row is partially filled

Express (+ simplify) as English sentences:

$X \cap Y_1$ $\quad$ $X \cap Y_2 = X$

$X \cup Y_1$ $\quad$ $X \cup Y_2 = X$

$X \backslash Y_1$ $\quad$ $X \backslash Y_2 = \varnothing$

The empty set. Learn this notation!

Each event has a probability

Event: the back row is completely filled

↓

Probability 0.3

Probability function

$$\mathbb{P} : \text{events} \to [0,1]$$

$$0 \leq \mathbb{P}(x) \leq 1$$

What does it mean to have a probability?

Probability is a way of expressing
partial knowledge of an event

With enough knowledge and insight, could I
predict tomorrow's weather without uncertainty?

✗ Correct probability

✓ Best probability given my knowledge of the world ← Subjective!

Probability Space

is three things:

$$(\Omega, \mathscr{F}, \mathbb{P})$$

Sample space:
the set containing all possible outcomes

Event space:
the set containing all possible events (combinations of outcomes)

Probability function:
Assigns a probability to every single event

Probability Space

for next week's lecture seating:

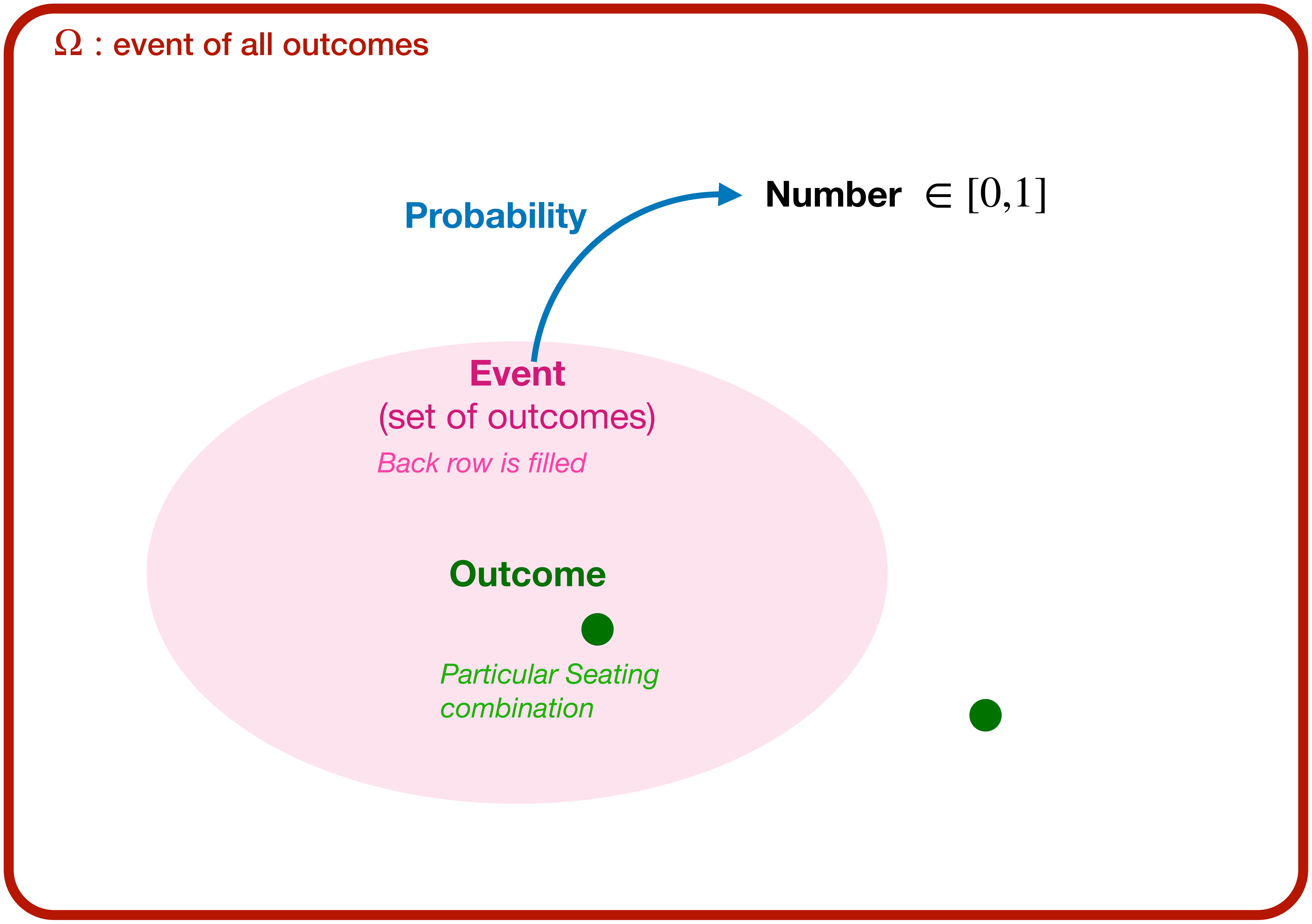
$$(\Omega, \mathscr{F}, \mathbb{P})$$

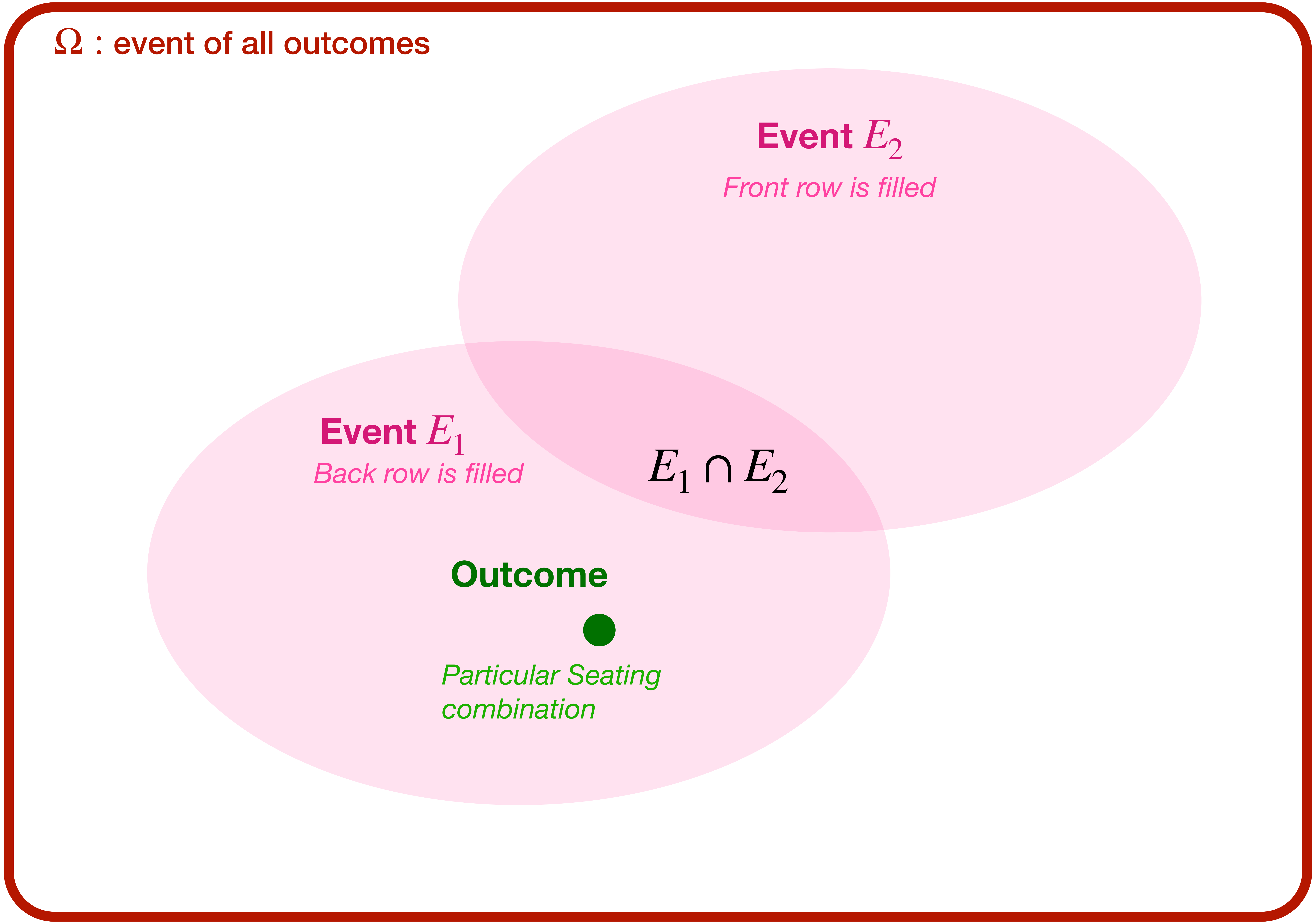
Sample space:
Every possible seating combination

Probability function:
Assigns a probability to every single event

Event space:
Sets of seating combinations

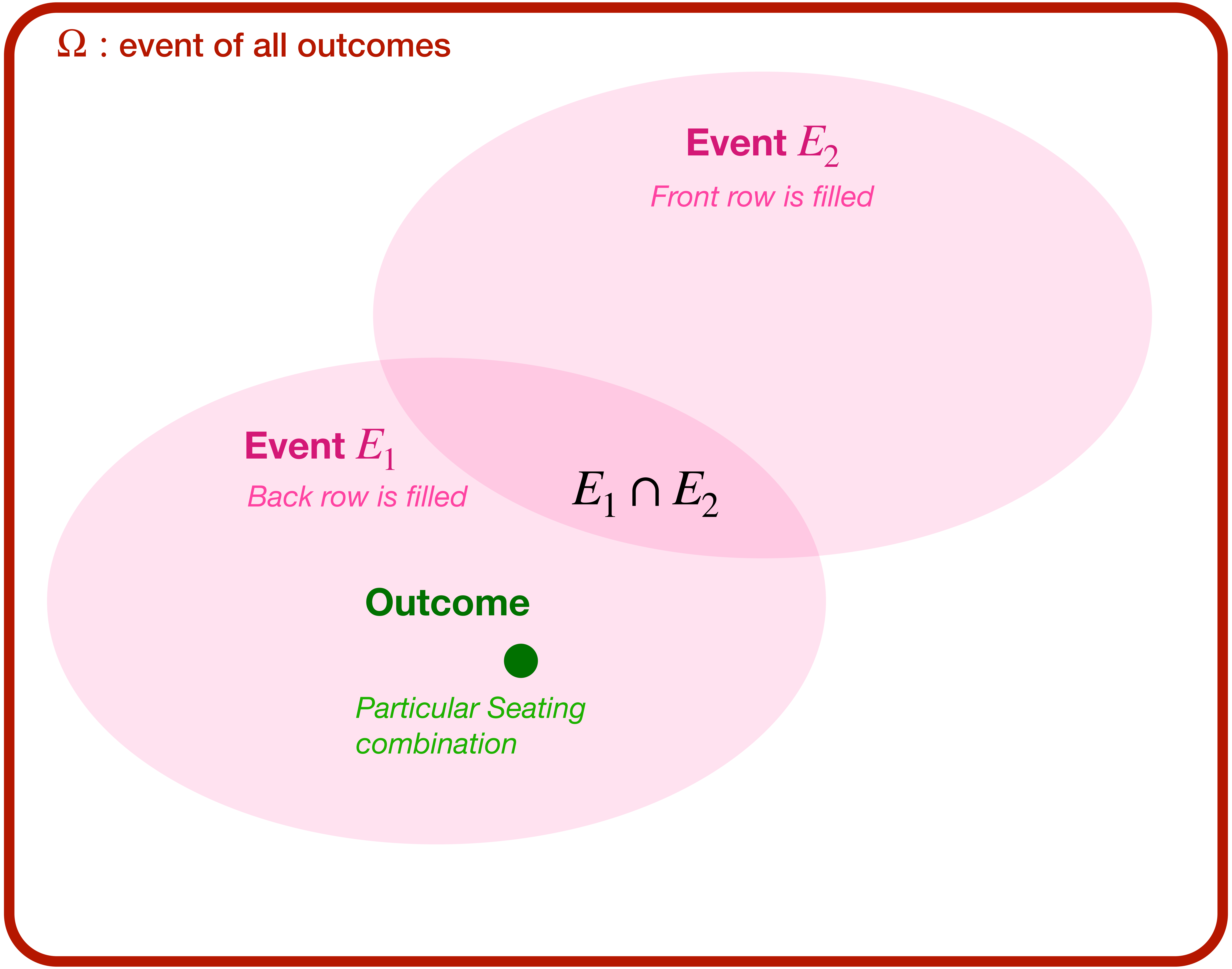
Example set:
All seating combinations where the back row is filled


Ω : event of all outcomes
Probability
Number ∈ [0,1]
Event
(set of outcomes)
Back row is filled
Outcome
Particular Seating
combination



$\cap$: and

$\cup$: or



$$\mathbb{P}[E_1 \cup E_2] =$$
$$\mathbb{P}[E_1] + \mathbb{P}[E_2] - \mathbb{P}[E_1 \cap E_2]$$

P(back OR front row filled)

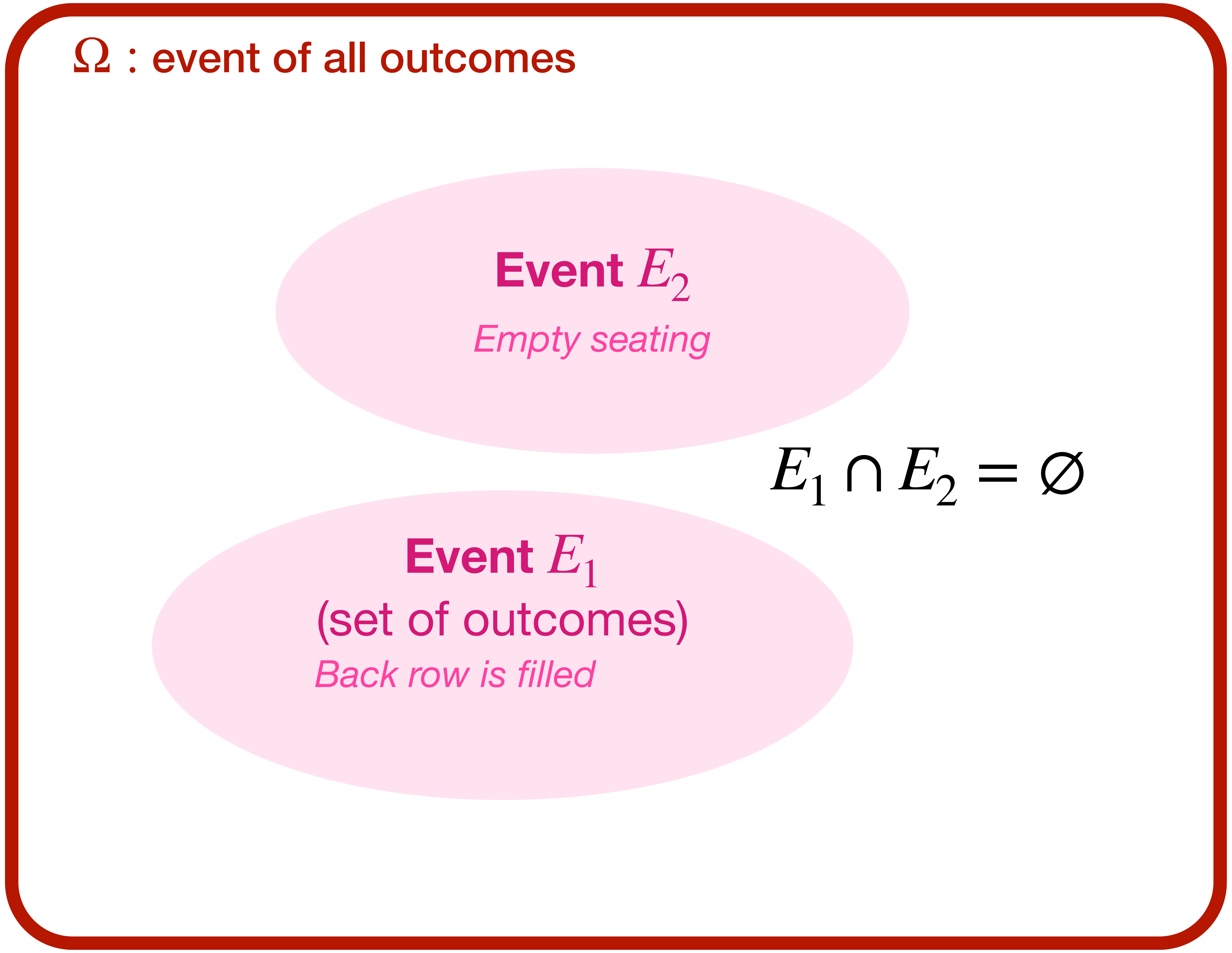
=

P(back row filled) +

P(front row filled) -

P(both filled)

Mutually exclusive events



$$\mathbb{P}[E_1 \cup E_2] =$$

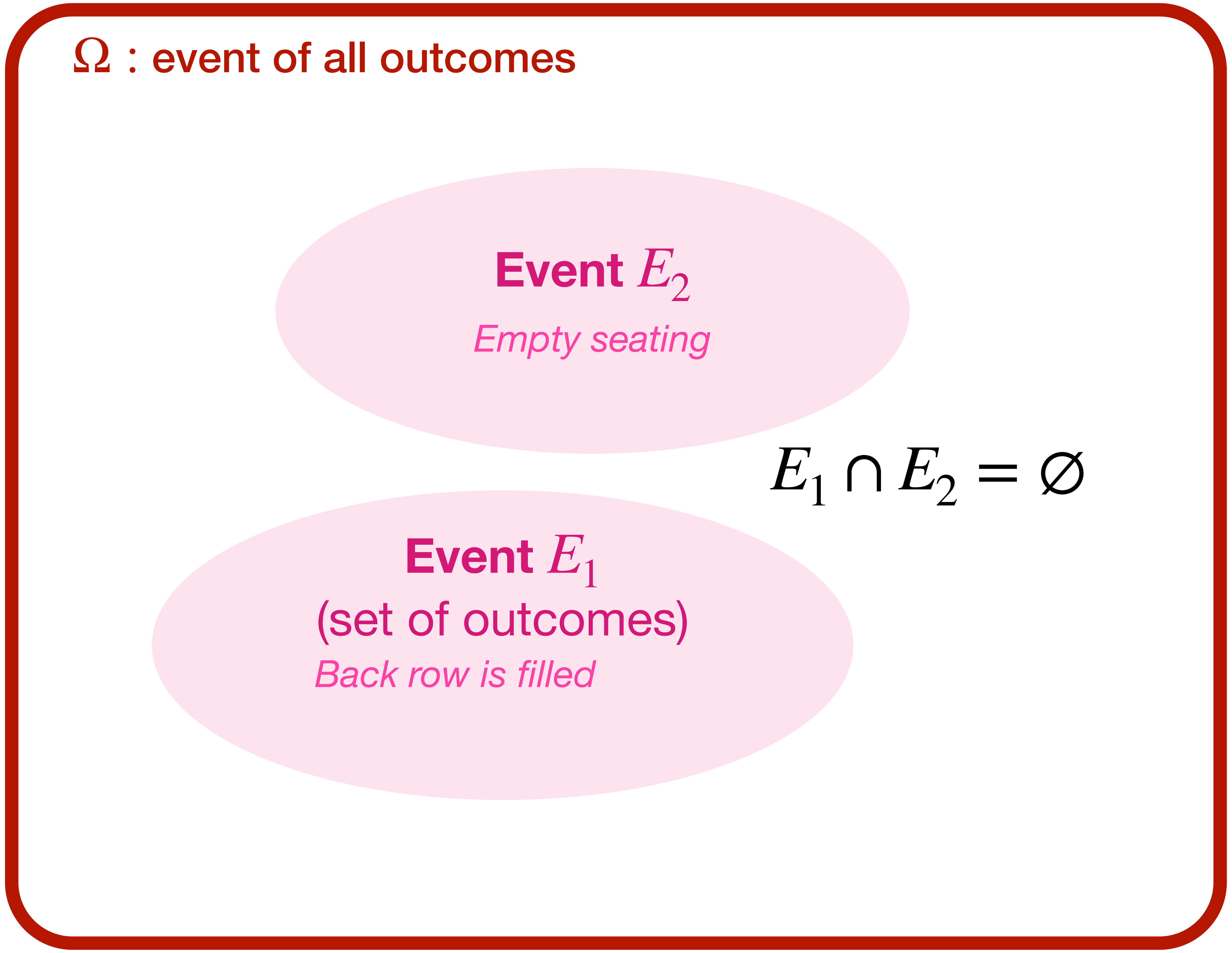
$$\mathbb{P}[E_1] + \mathbb{P}[E_2] - \cancel{\mathbb{P}[E_1 \cap E_2]}$$

Condition for mutually exclusive events:

$$\mathbb{P}[E_1 \cap E_2] = 0$$

(They can't both happen at the same time)

Mutually exclusive events



$$\mathbb{P}[E_1 \cup E_2] =$$

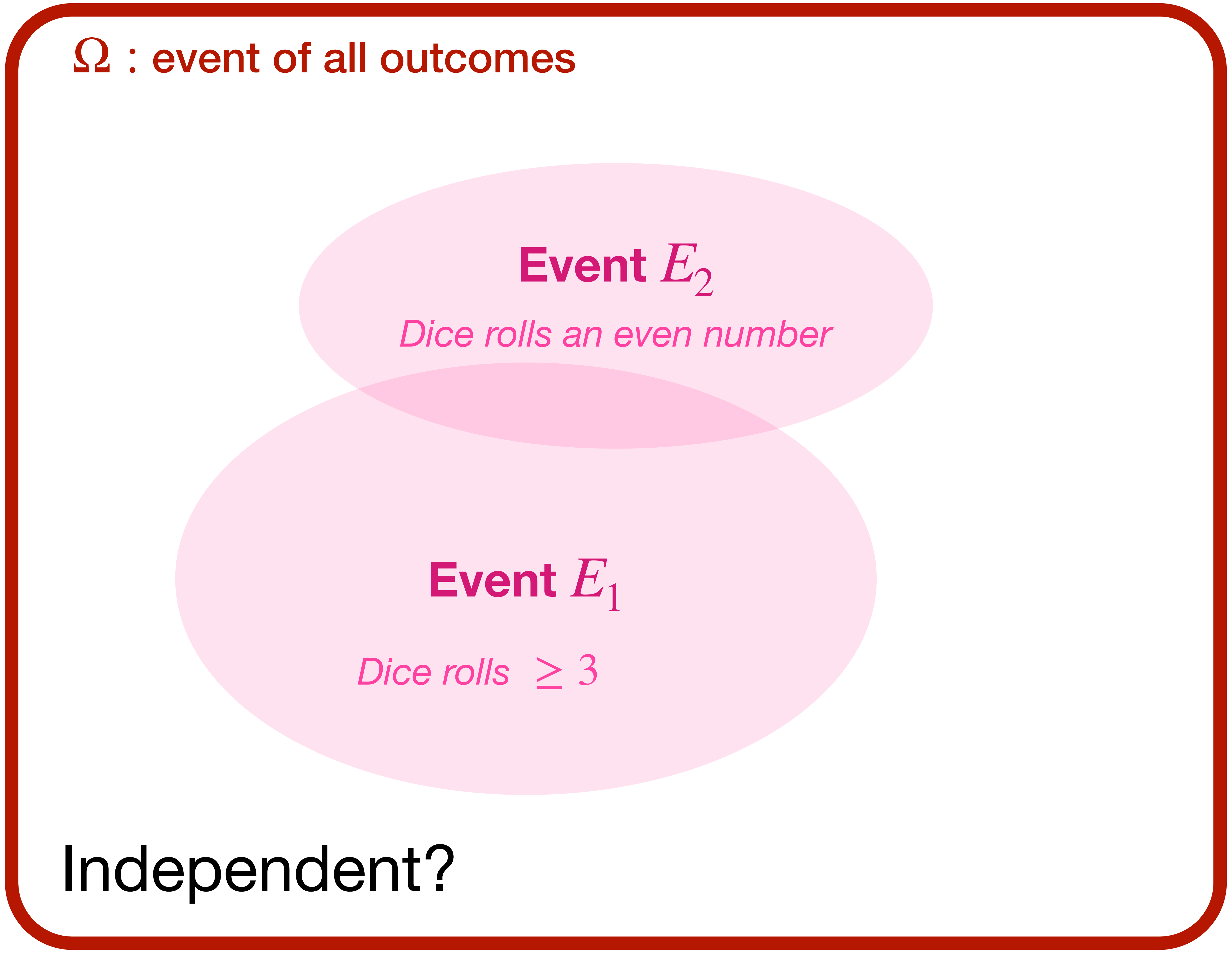
$$\mathbb{P}[E_1] + \mathbb{P}[E_2] - \mathbb{P}[\cancel{E_1 \cap E_2}]$$

Equivalent condition for mutually exclusive events

$$\mathbb{P}[E_2 \,|\, E_1] = \mathbb{P}[E_1 \,|\, E_2] = 0$$

"E_2 given E_1"
"Conditional probability"

Independent events are not mutually exclusive



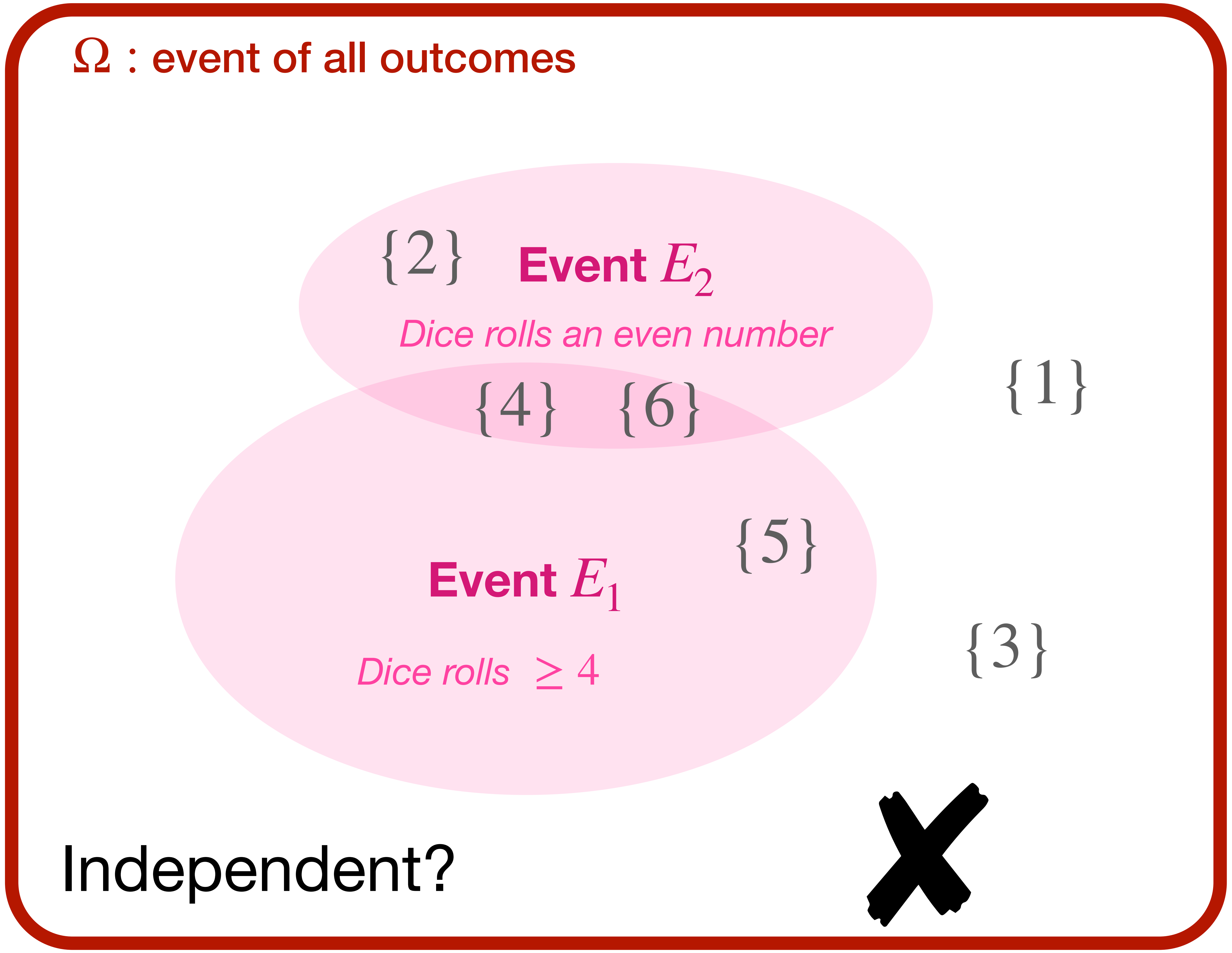
Independent events definition:

Knowledge of one outcome doesn't change probability of another

$$\mathbb{P}[E_2 \,|\, E_1] = \mathbb{P}[E_2]$$

$$\mathbb{P}[E_1 \,|\, E_2] = \mathbb{P}[E_1]$$

Independent events are not mutually exclusive



Independent events definition:

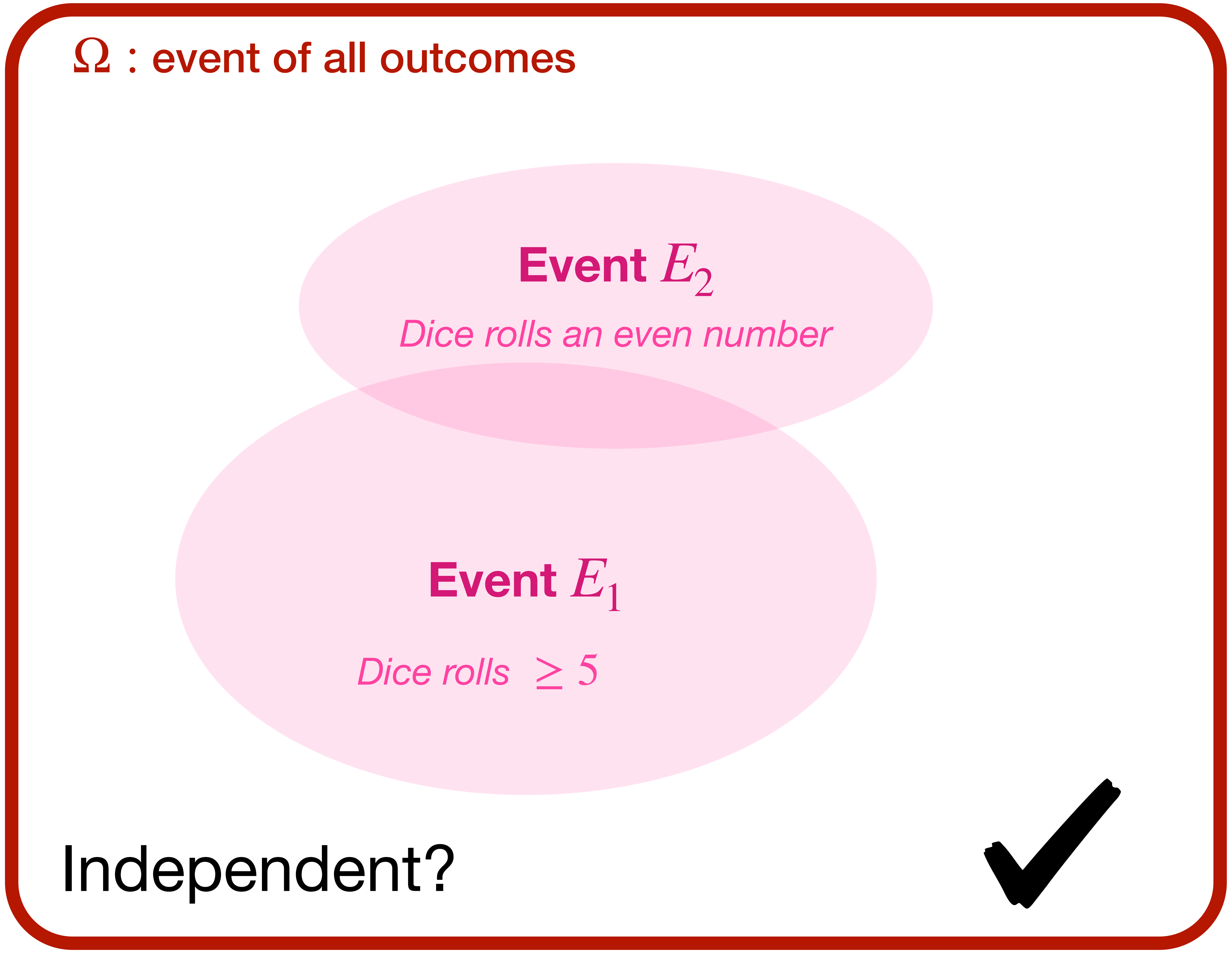
Knowledge of one outcome doesn't change probability of another

$$\mathbb{P}[E_2 \mid E_1] = \mathbb{P}[E_2] = 0.5$$

$$\mathbb{P}[E_1 \mid E_2] = \frac{2}{3}$$

$$\neq \mathbb{P}[E_1] = 0.5$$

Independent events are not mutually exclusive



Independent events definition:

Knowledge of one outcome doesn't change probability of another

$$\mathbb{P}[E_2 \mid E_1] = \mathbb{P}[E_2] = 0.5$$

$$\mathbb{P}[E_1 \mid E_2] = \mathbb{P}[E_1] = \frac{1}{3}$$

Independent?
Mutually exclusive?

$\subset$ = `subset'

E^c = `not (complement) E'

Experiment: take two balls from bag

E_1: First ball is blue

E_2: Second ball is green

E_3: First ball is big

Questions:

$E_3 \subset E_1$?

$E_2 \subset E_1$?

$\mathbb{P}[E_2 \,|\, E_1] = ?$

$\mathbb{P}[E_3 \,|\, E_1] = ?$

$E_1^c \cap E_3$?

$\mathbb{P}[E_3 \cup E_1]$?

Independent?
Mutually exclusive?

$\subset$ = `subset'

E^c = `not (complement) E'

Experiment: take two balls from bag

E_1: First ball is blue

E_2: Second ball is green

E_3: First ball is big

Questions:

$E_3 \subset E_1$

$\mathbb{P}[E_2 \,|\, E_1] = 0.5$

$E_1^c \cap E_3 = \varnothing$

$E_2 \subset E_1$

$\mathbb{P}[E_3 \,|\, E_1] = \dfrac{1}{3}$

$\mathbb{P}[E_3 \cup E_1] = \dfrac{3}{5}$

Random variables

are quantitative questions
about the experiment

Random variables

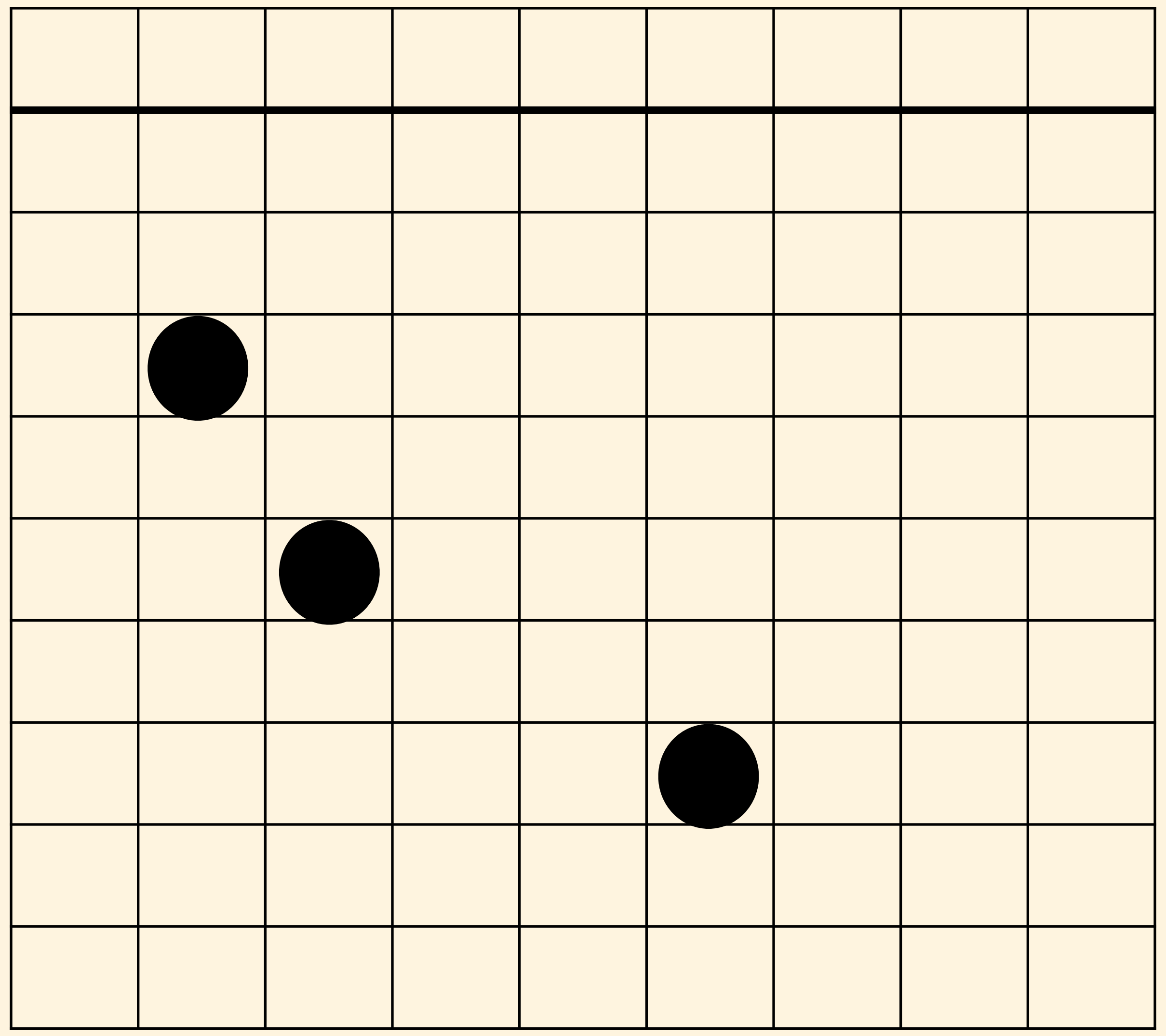
are quantitative questions
about the experiment

are functions that map from
outcomes to numbers

Random variables example

What was the number of unfilled seats? $X(\omega)$

Outcome $\omega \in \Omega$



$\longrightarrow$

Number $X(\omega)$

87

(Or any integer between
0 and *number of seats*)

Random variable taking value is an event

$$\left\{\begin{array}{c}\text{Outcome}\\ \text{Outcome}\\ \text{Outcome}\end{array}\right\} \longrightarrow \text{Random variable takes particular value} \longleftrightarrow \text{Probability}$$

Eg $\{$ outcomes where number of unfilled seats $= 87\}$

Notation

Probability of four unfilled seats?

… is an event!

Set of all outcomes

Random variable output from outcome

Correct expression

$$\mathbb{P}[\{\omega \in \Omega : X(\omega) = 4\}]$$

Curly brackets means set

Outcome

Belonging to

Such that

Shorthand

$$\mathbb{P}[X = 4]$$

Flipping a coin

H → 8

T → 0

Random variable is not ‘random’. The outcome is.

RV is the deterministic mapping from ‘random’ outcomes to numbers

Support of a random variable

Set of plausible values a random variable can take

H $\rightarrow$ 8

T $\rightarrow$ 0

$\text{supp}(X) = \{0, 8\}$

Support of a random variable

X is the number of unfilled seats

Smallest set S such that
$\mathbb{P}[X \in S] = 1$

With probability one:

$0 \leq X \leq$ number of seats
$X \in \mathbb{Z}$

$\Rightarrow$

$\text{supp}(X) = \{0, 1, \ldots \text{number of seats}\}$

Support of a random variable

Y is whether the back row is filled

What’s the support of Y?

Support of a random variable

Z is height of 2nd person in 3rd row

What’s the support of Z?

Support of a random variable

Z is height of 2nd person in 3rd row

What's the support of Z?

$$\text{supp}(Z) = (l, u)$$

l : height of shortest person on course

h : height of tallest person on course

Two flavours of random variables

Support is finite set

Discrete random variables

X is the number of unfilled seats

$\text{supp}(X) = \{0,1,2,\ldots 200\}$

Y is whether the back row is filled

$\text{supp}(Y) = \{0,1\}$

Support is infinite set

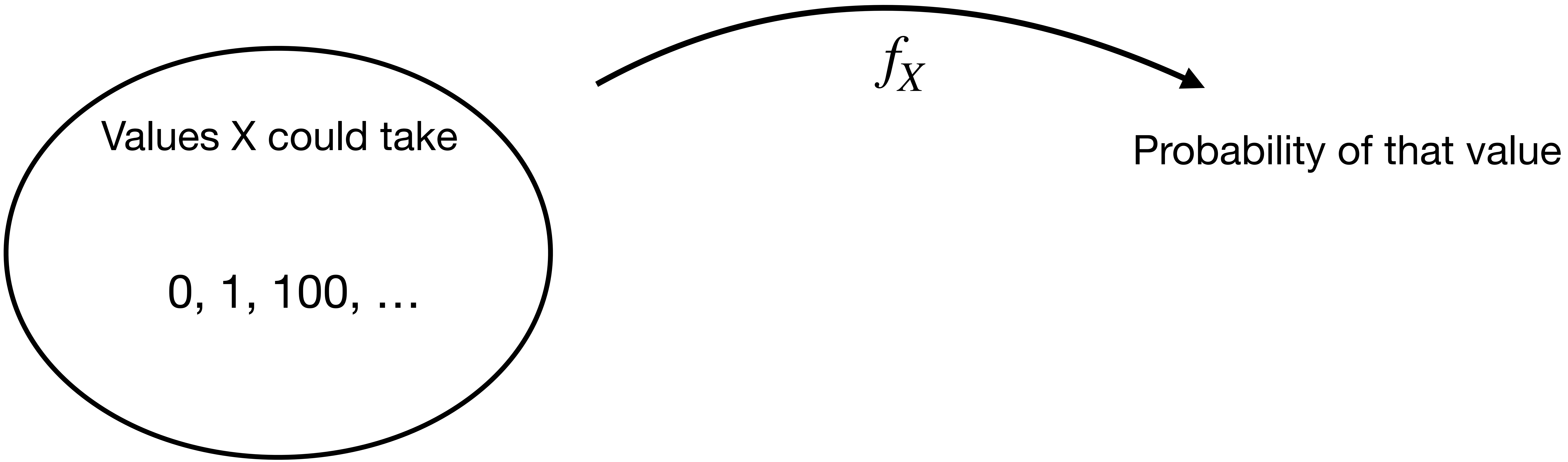
Continuous random variables

Z is height of 2nd person in 3rd row

$\text{supp}(Z) = (l, u)$

Probability mass function of a discrete random variable

$$f_X : \mathbb{R} \to [0,1]$$



Probability mass function of a discrete random variable

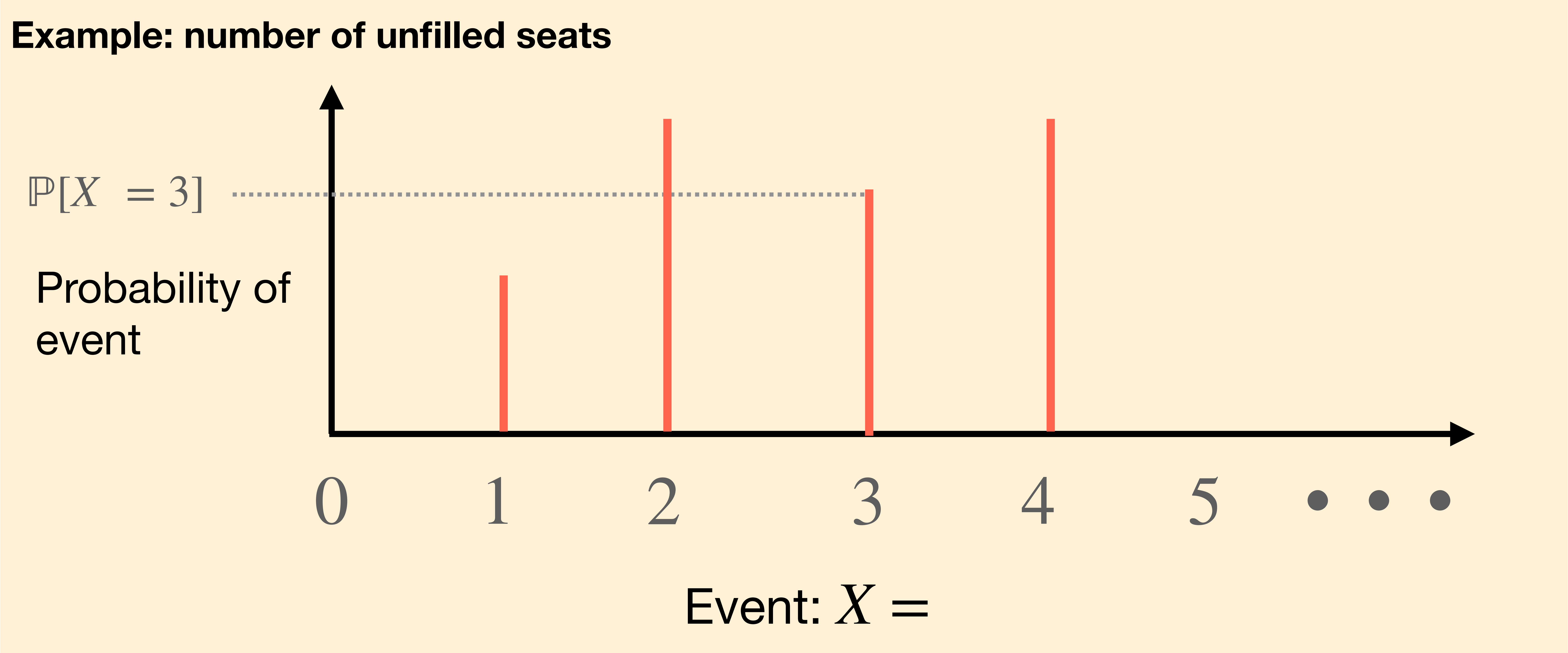
X is the number of unfilled seats

$$f_X(4) = 0.1$$

$\updownarrow$

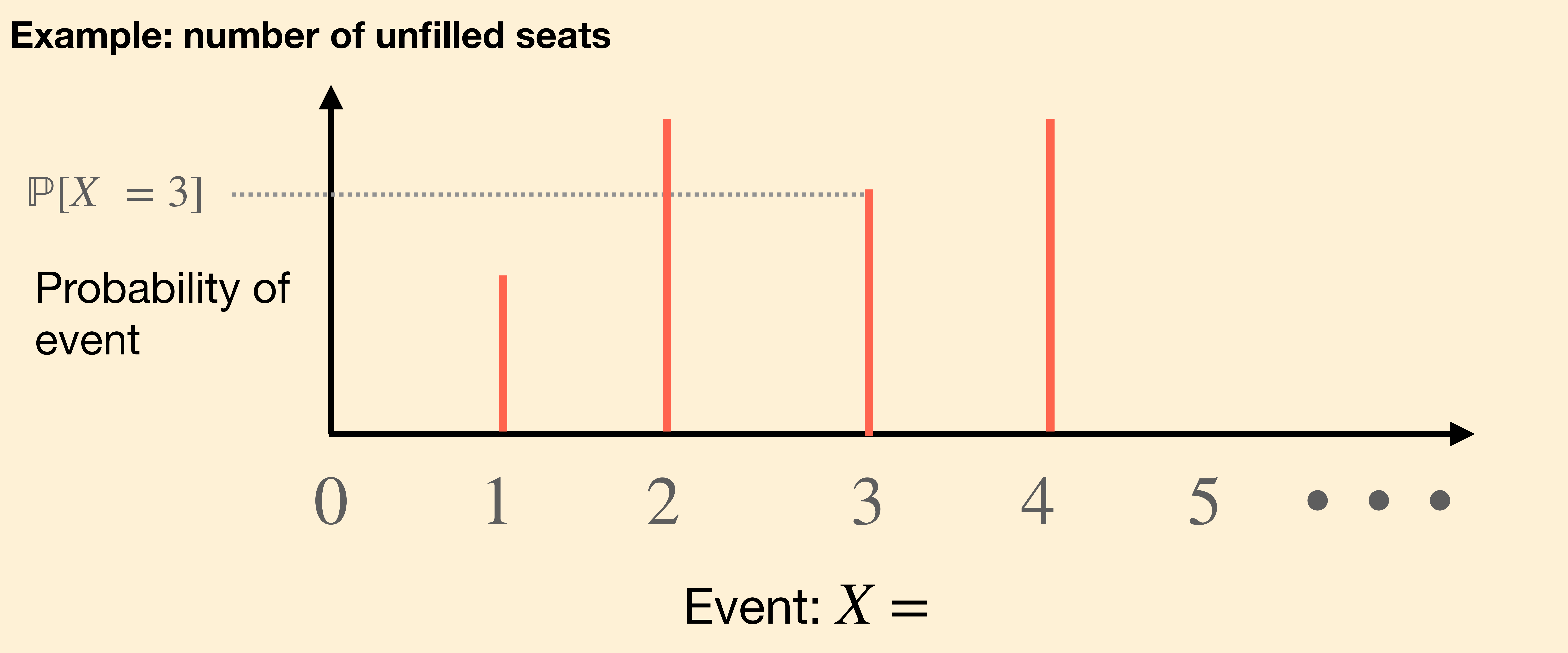
Probability of 4
unfilled seats is 0.1

Graph of probability mass function f_X



Total length of red lines?

Properties of Probability mass functions



Sum of red line lengths adds to one $\Leftrightarrow$ Probability of X taking some value = 1

Random variables with particular PMFs pop up quite often, across experiments

We give then names

Bernoulli random variable

Random variables are quantitative questions about the experiment

Bernoulli random variables are binary questions (yes/no)

Y *is whether the back row is filled*

f_Y?

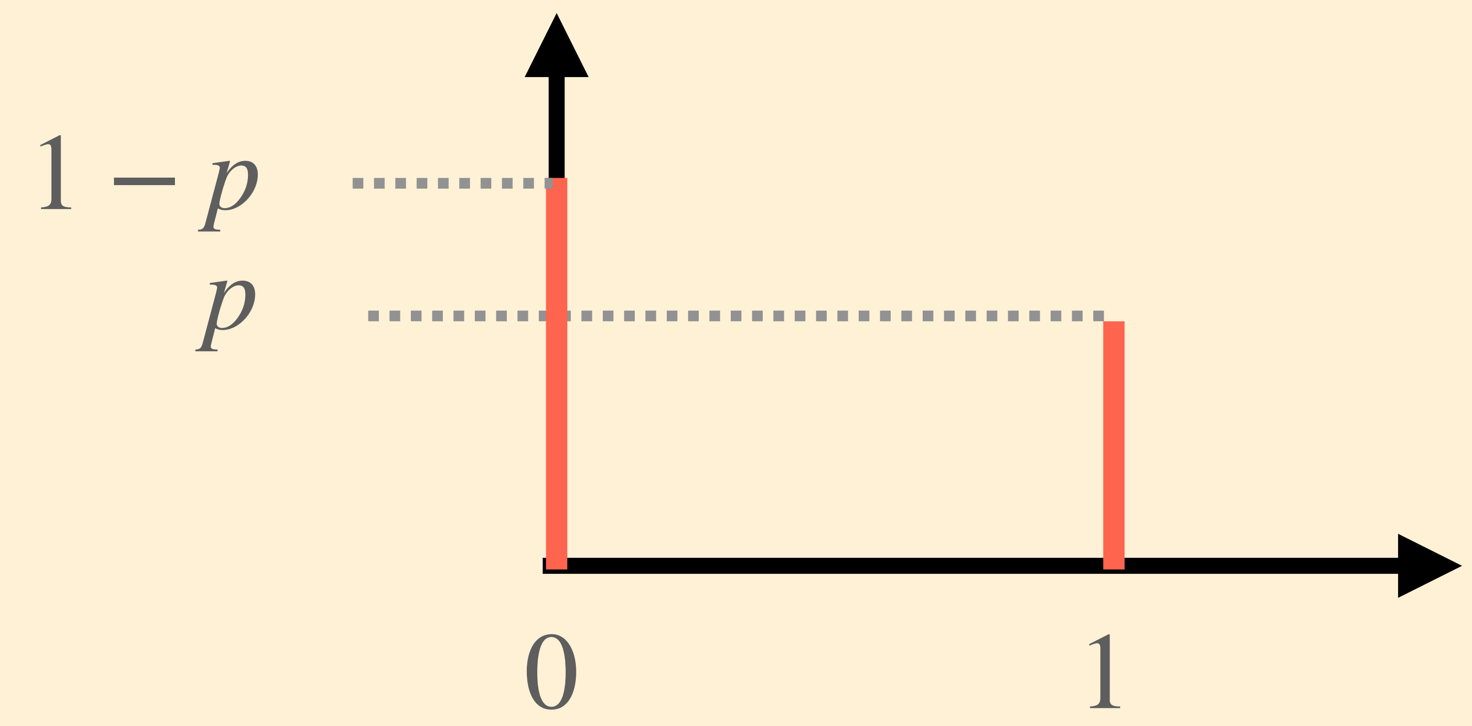
Bernoulli random variable

```
numpy.random.binomial(1,p)
```

$$Y \sim \text{Bern}(p)$$

Y **is distributed as a Bernoulli random variable, with probability** p (of yes)

$$f_Y = \begin{cases} p & \text{if } x = 1 \\ 1 - p & \text{if } x = 0 \end{cases}$$



Uniform random variables

`numpy.random.rand`

Probability of every outcome in support is equal

(Probably means we know very little about experiment)

NB: S is support

X is uniformly distributed on the set S $\Longleftrightarrow$ $X \sim U(S)$

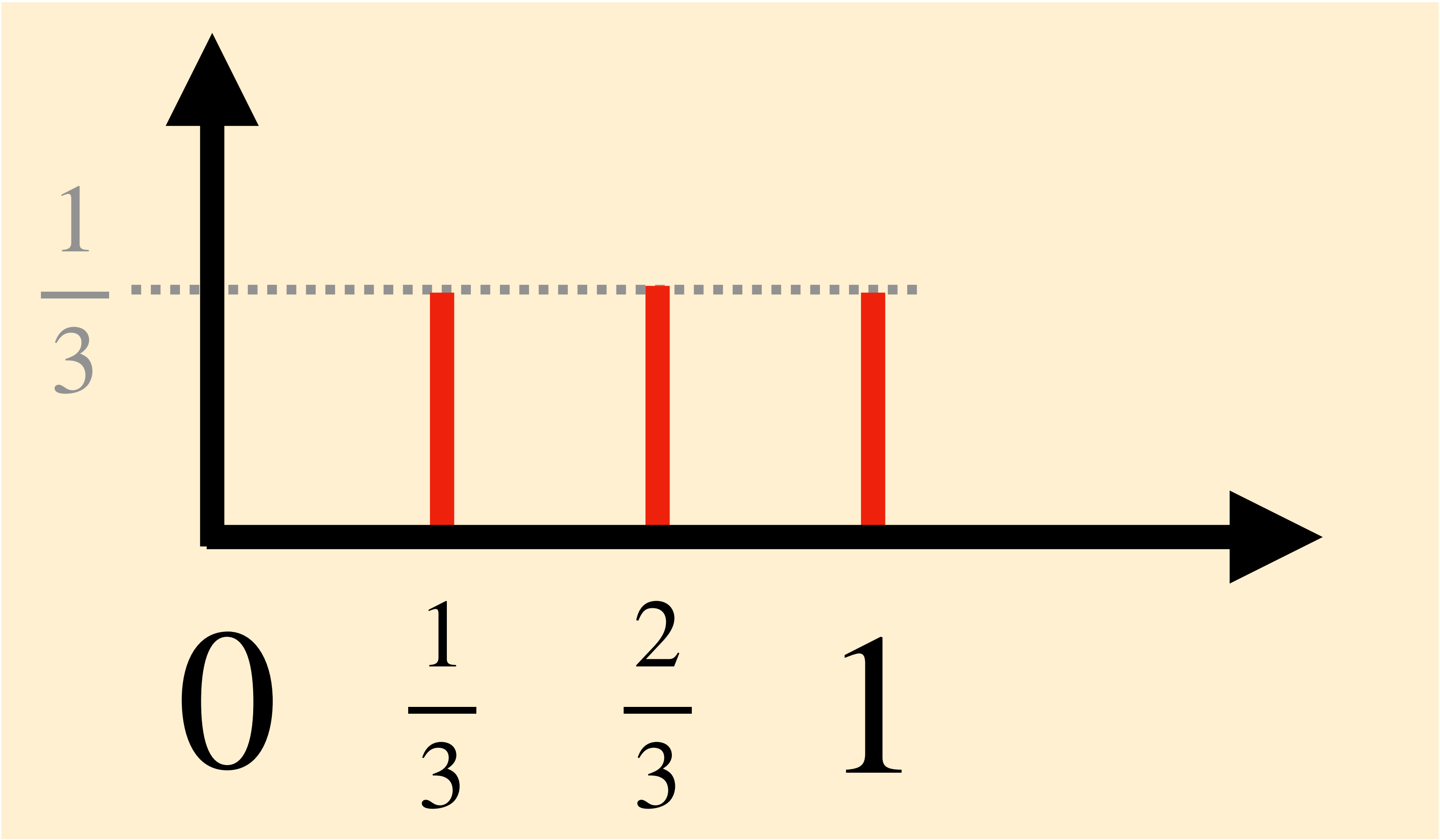
X is a discrete random variable $\Longleftrightarrow$ $|S| < \infty$ (Finite number of elements in set)

Example of a Uniform random variable

$$X \sim U\left(\left\{\frac{1}{3}, \frac{2}{3}, 1\right\}\right)$$

NB: curly brackets
{ • } means set

Probability
mass function



Random variable with continuous support

What’s the probability a human is 180cm tall?

Random variable with continuous support

What’s the probability a human is 180cm tall?

What’s the probability a human is 180.00000000… cm tall?

Random variable with continuous support

Every possible (single) outcome is *impossible*

Only reasonable quantity: ranges of outcomes

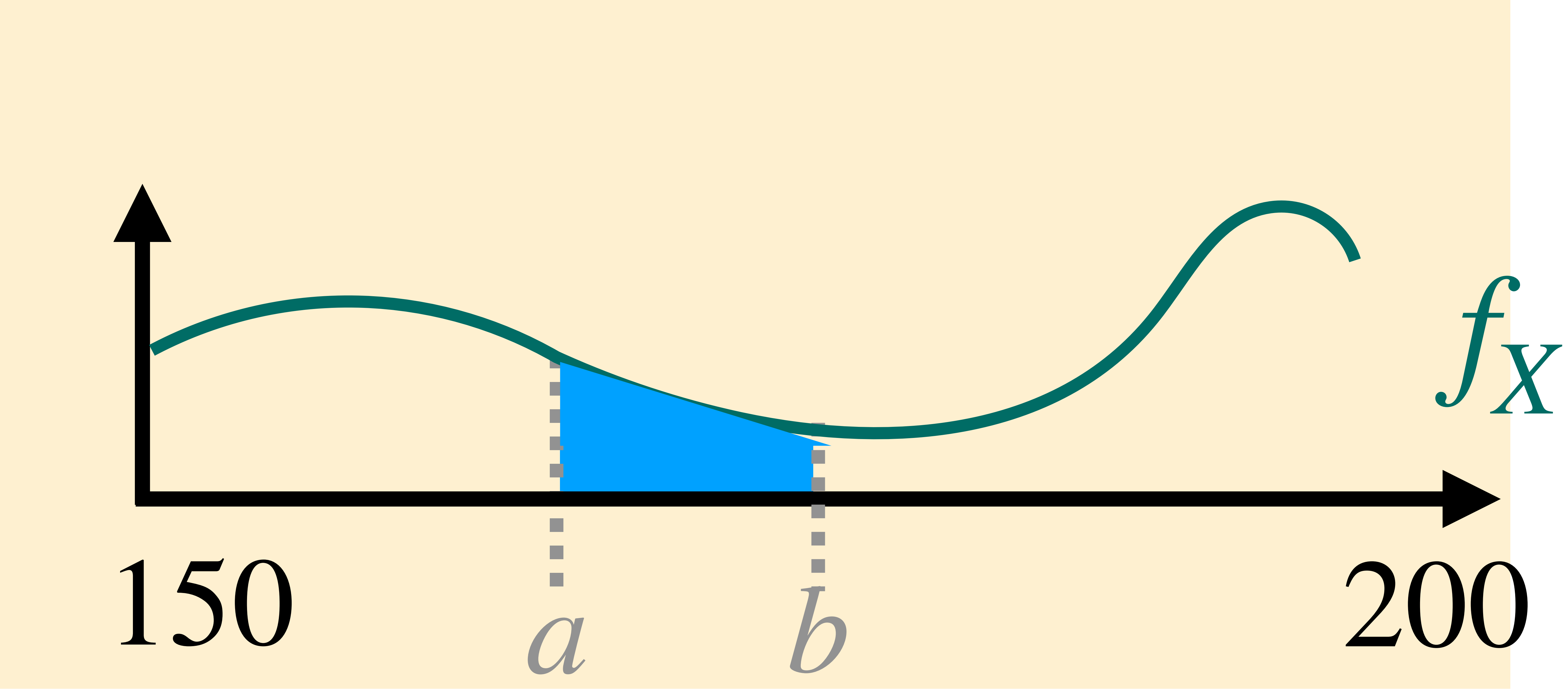
Probability density functions

Continuous random variables only

$\mathbb{P}[X = \text{anything}] = 0$

Instead, look for probability X is between values:

$\mathbb{P}[a \leq X \leq b] = ?$



$$\boxed{\mathbb{P}[x \in (a, b)] = \int_a^b f_X(x)\ dx}$$

Catching breath

https://www.youtube.com/watch?v=8idr1WZ1A7Q

Experiments have outcomes with probabilities

Sets of outcomes are events

Random variables map outcomes to values

Random variable taking particular values is an event

PMFs/PDFS give probabilities of such events

Next up

Statistics of random variables

Expected value

Variance

Functions of RVs

Central limit theorem

Conditional probability

Independent events

Bayes' law

Recap: probability space

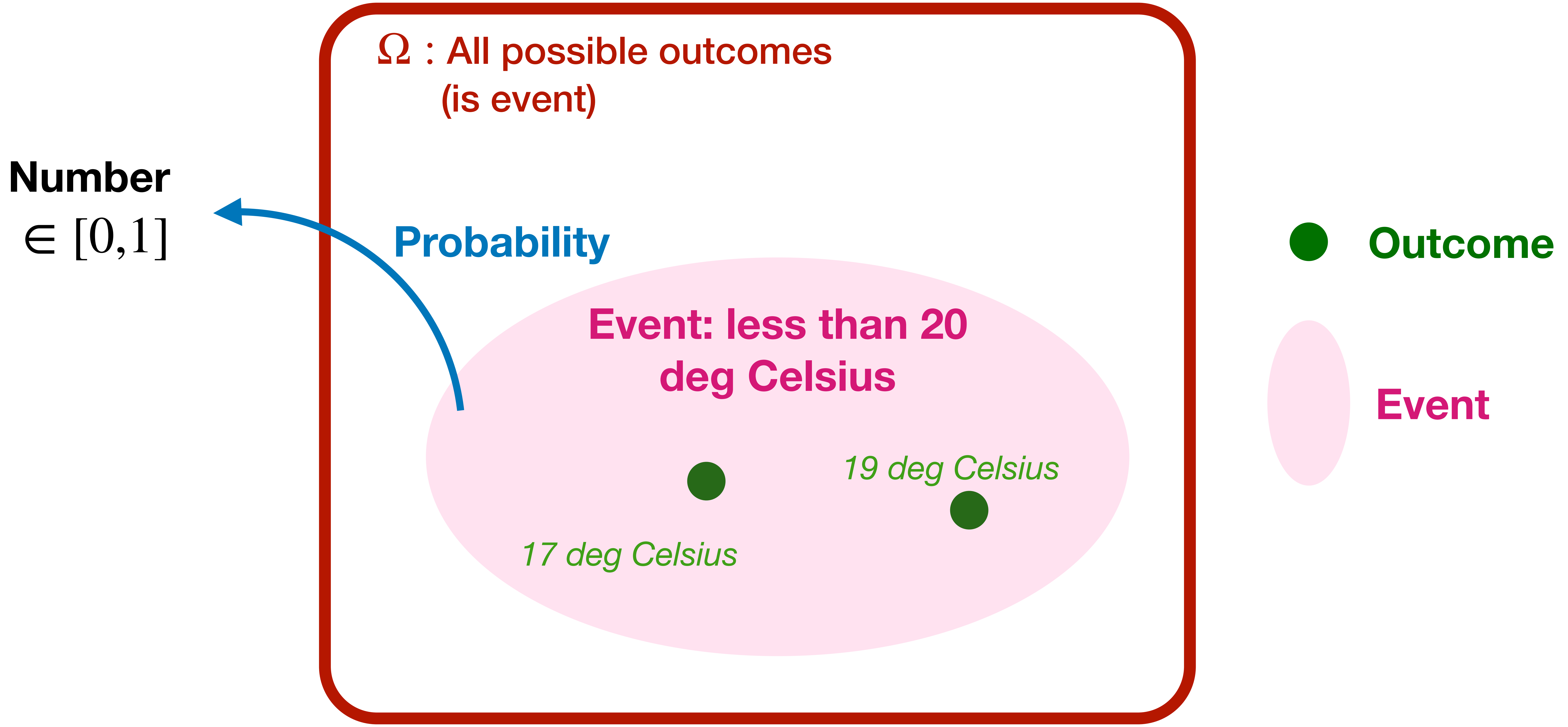
Can only talk about probabilities when you have an experiment

Experiments have outcomes

Sets of outcomes are events

All events have probabilities

Experiment: tomorrow's midday temperature



Random variables

Recap

are quantitative questions
about the experiment

are functions that map from
outcomes to numbers

(or to any "measurable space")

Purpose

Say something quantitative about
a situation we can't model fully

(e.g. lecture seating next week)

Practice

$$\begin{bmatrix} \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet \end{bmatrix}$$

$$\bullet \sim U[-1,1]$$

is distributed as

Uniform RV

On set of numbers between
-1 and 1

Outcome space?

What type of RV is:
"Is the matrix invertible?"

Probability of event?
"Matrix is invertible"

Hint: invertible $\leftrightarrow$ nonzero eigenvalues

Terminology

Spot the difference

$X(\omega)$: RV is a function

X_i: is a value for the function on trial i

Experimental trial

Run the experiment once: $(\Omega, \mathscr{F}, \mathbb{P})$

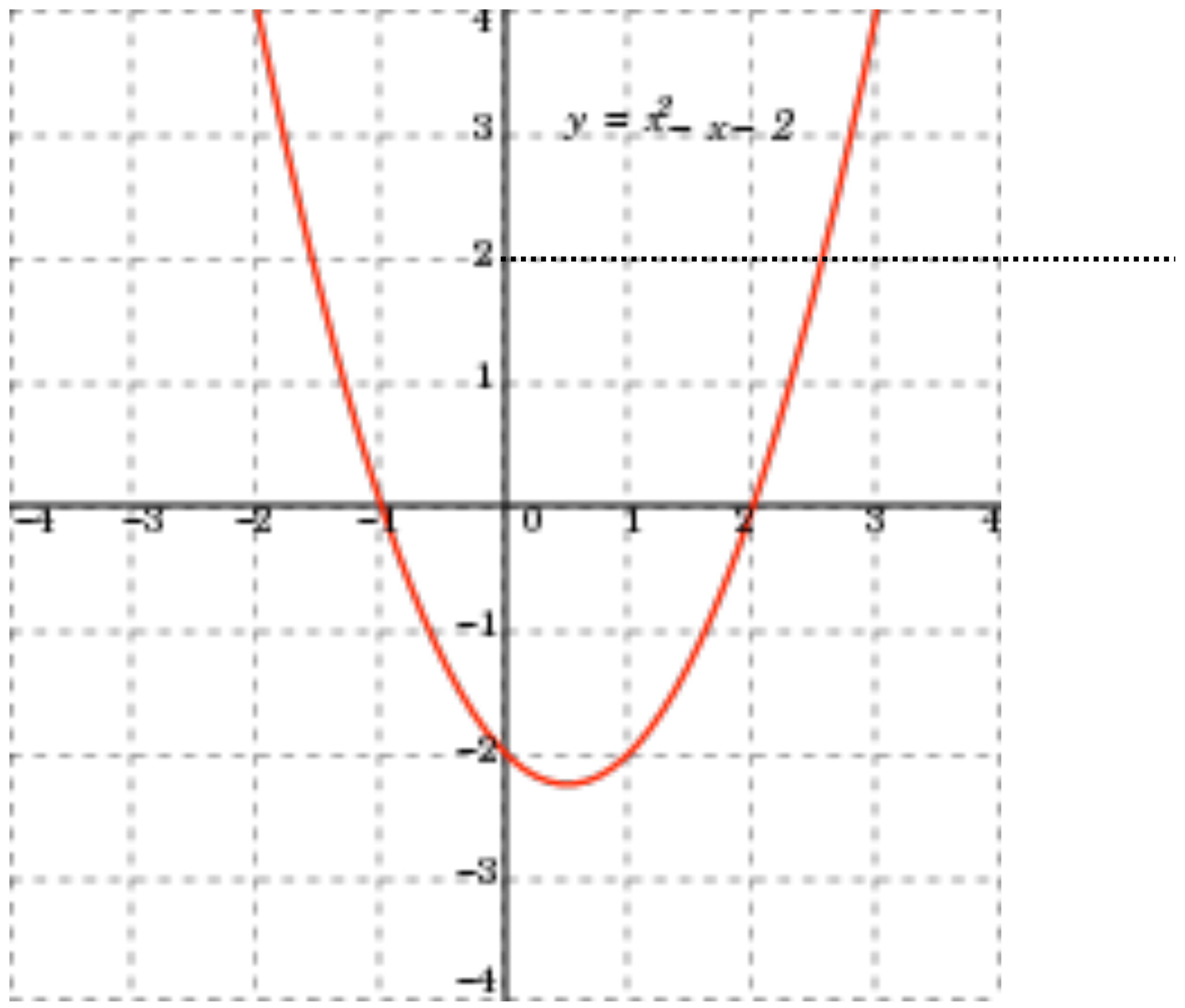
ω_i = *outcome on trial* i

$X_i =$ *Value of RV* X *on trial* i

EG $X_i =$ *Filled seats on week* i

Just like….

Function



Value at ….

Statistics

Summaries of a random variable

Statistical function:

Random variable

$\downarrow$

Number

Expectation

"On average, there are 23 unfilled seats"

Variance

"Number of empty seats typically varies this much across samples"

$\vdots$

Sacrifice information for interpretability

Population vs sample statistics

Experiment

Pick somebody at
Sussex uni

Statistic

Average
height?

Population statistic

Statistic over all
outcomes (people)

Outcome
$\omega \in \Omega$
Person

RV $X(\omega)$

Height

Sample statistic

Statistic over limited
outcomes (some people)

Event $f \in \mathscr{F}$
Sets of people

Population vs sample statistics

Population statistic

Statistic over all outcomes (people)

Sample statistic

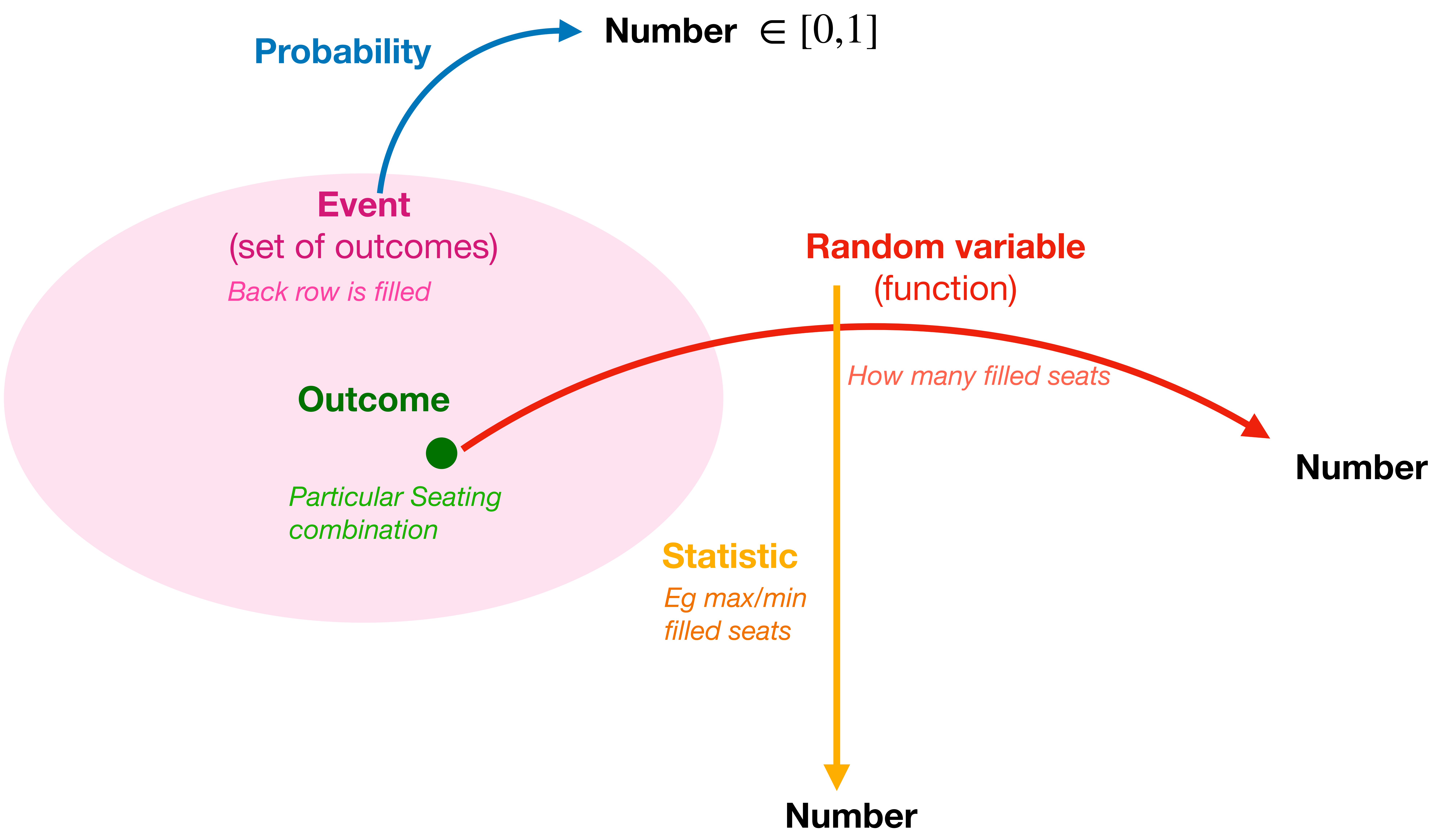
Statistic over limited outcomes (some people)

Hard to measure

Easier to measure

Use sample statistic to estimate?

Sample statistic called estimator


Probability
Number $\in [0,1]$
Event
(set of outcomes)
Back row is filled
Random variable
(function)
How many filled seats
Outcome
Number
Particular Seating
combination
Statistic
Eg max/min
filled seats
Number

Statistic 1: expectation by example

(Mean/average)

Ω : all outcomes

All possible seating combinations

$S \subset \Omega$: subset of outcomes

Seating combinations sampled during course

$X : \Omega \to \mathbb{Z}$

RV from outcomes to number of filled seats

Samples:

$$S = \{\omega_1, \omega_2, \ldots \omega_5\}$$
$$S = \{\omega_i\}_{i=1}^5$$

Samples of RV:

$$\{X(\omega_1, X(\omega_2), \ldots\} =$$
$$\{90,\ 90,\ 110,\ 80,\ 130\}$$

Sample expectation

$$\bar{\mu}(X) = \sum_{\omega \in S} X(\omega)\mathbb{P}(\omega)$$

$$= 90 \times \frac{1}{5} + 90 \times \frac{1}{5} + 80 \times \frac{1}{5} + \ldots$$
$$= 100$$

Statistic 1: expectation

(Mean/average)

Ω : all outcomes

All possible seating combinations

$S \subset \Omega$: subset of outcomes

Seating combinations sampled during course

$X : \Omega \to \mathbb{Z}$

RV from outcomes to number of filled seats

Sample expectation

$$\bar{\mu}(X) = \sum_{\omega \in S} X(\omega)\mathbb{P}(\omega)$$

Population expectation

$$\mu(X) = \sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\omega)$$

Average over *all* possible outcomes

Law of large numbers

"As you take more samples, the sample expectation converges to the population expectation"

Statistic 1: expectation

Alternative (better) formula

Sample expectation

$$\bar{\mu}(X) = \sum_{\omega \in S} X(\omega)\mathbb{P}(\omega)$$

Samples of RV:

$$\{X(\omega_1, X(\omega_2), \dots\} =$$
$$\{90,\ 90,\ 110,\ 80,\ 130\}$$

5 outcomes

4 values

Population expectation

$$\mu(X) = \sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\omega)$$

Average over *all* possible outcomes

Sum over probability of values of RV. Instead of outcomes

$$\mu(X) = \sum_{x \in \text{supp}(X)} x \times \mathbb{P}[X(\omega) = x]$$

$$= 90 \times \frac{2}{5} + 80 \times \frac{1}{5} + \dots$$

Statistic 1: expectation

Alternative (better) formula

Sample expectation

$$\bar{\mu}(X) = \sum_{\omega \in S} X(\omega)\mathbb{P}(\omega)$$

$$\bar{\mu}(X) = \sum_{x \in \text{supp}(X)} x \times \mathbb{P}_{\omega \in S}[X(\omega) = x]$$

Population expectation

$$\mu(X) = \sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\omega)$$

$$\mu(X) = \sum_{x \in \text{supp}(X)} x \times \mathbb{P}[X(\omega) = x]$$

Average over *all* possible outcomes

This is probability mass function!

Statistic 1: expectation

Alternative (better) formula

https://www.youtube.com/watch?v=OvTEhNL96v0

Sample expectation

$$\bar{\mu}(X) = \sum_{\omega \in S} X(\omega)\mathbb{P}(\omega)$$

$$\bar{\mu}(X) = \sum_{x \in \mathsf{supp}(X)} x \times \mathbb{P}_{\omega \in S}[X(\omega) = x]$$

Population expectation

$$\mu(X) = \sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\omega)$$

$$\mu(X) = \sum_{x \in \mathsf{supp}(X)} x \times f_X(x)$$

Average over *all* possible outcomes

This is probability mass function!

Lévy distribution

How big will it be?

$$\mathbb{E}[X] = \infty$$

Google if interested

Usual terminology

Functions: Inputs → Outputs

Functionals: Functions → Outputs

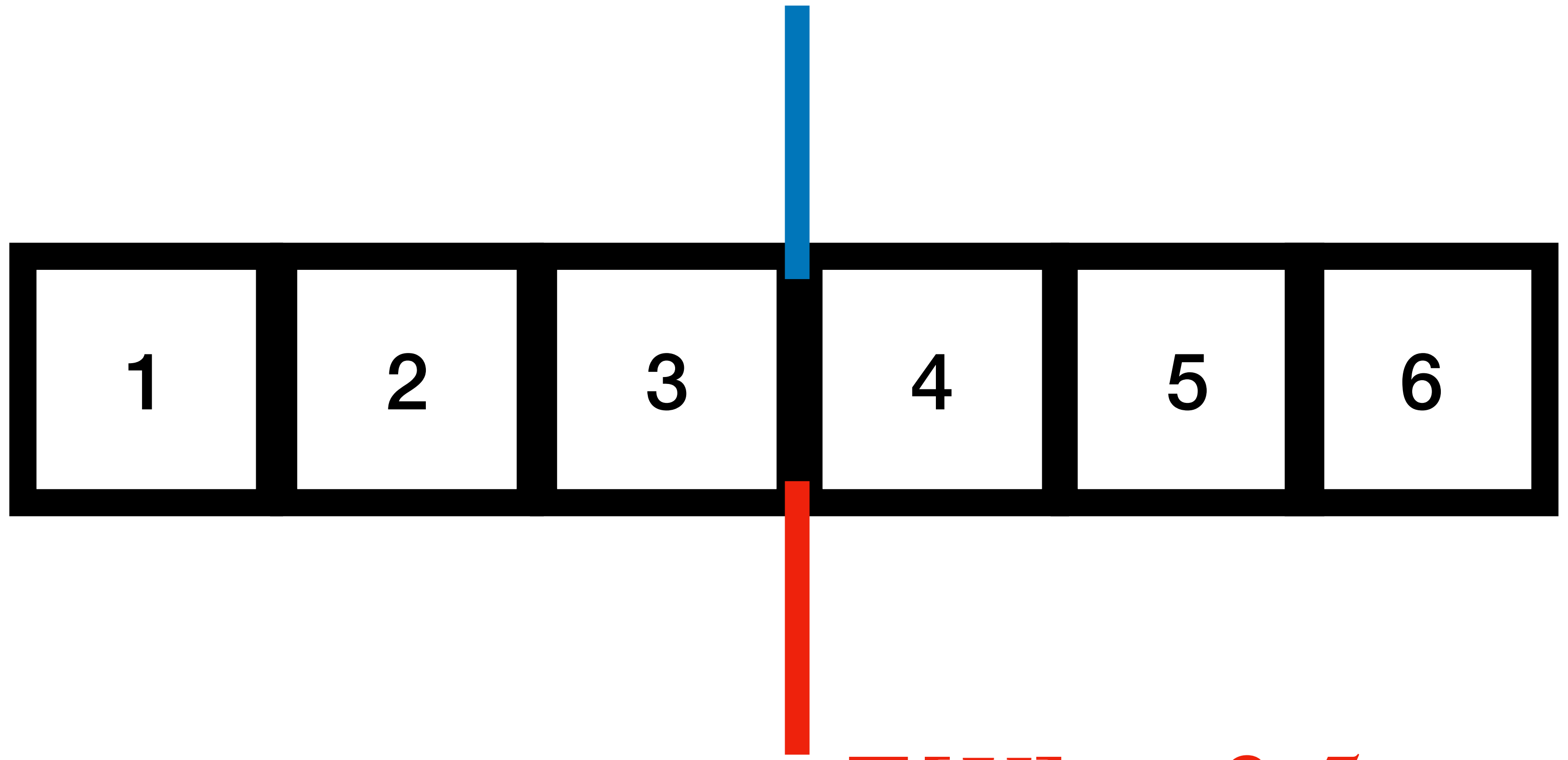
Higher-order function in programming

Think of a statistic as a functional

https://en.wikipedia.org/wiki/Higher-order_function

Expected value is sensitive to outliers

Median = 3.5 (or e.g. 3.8)



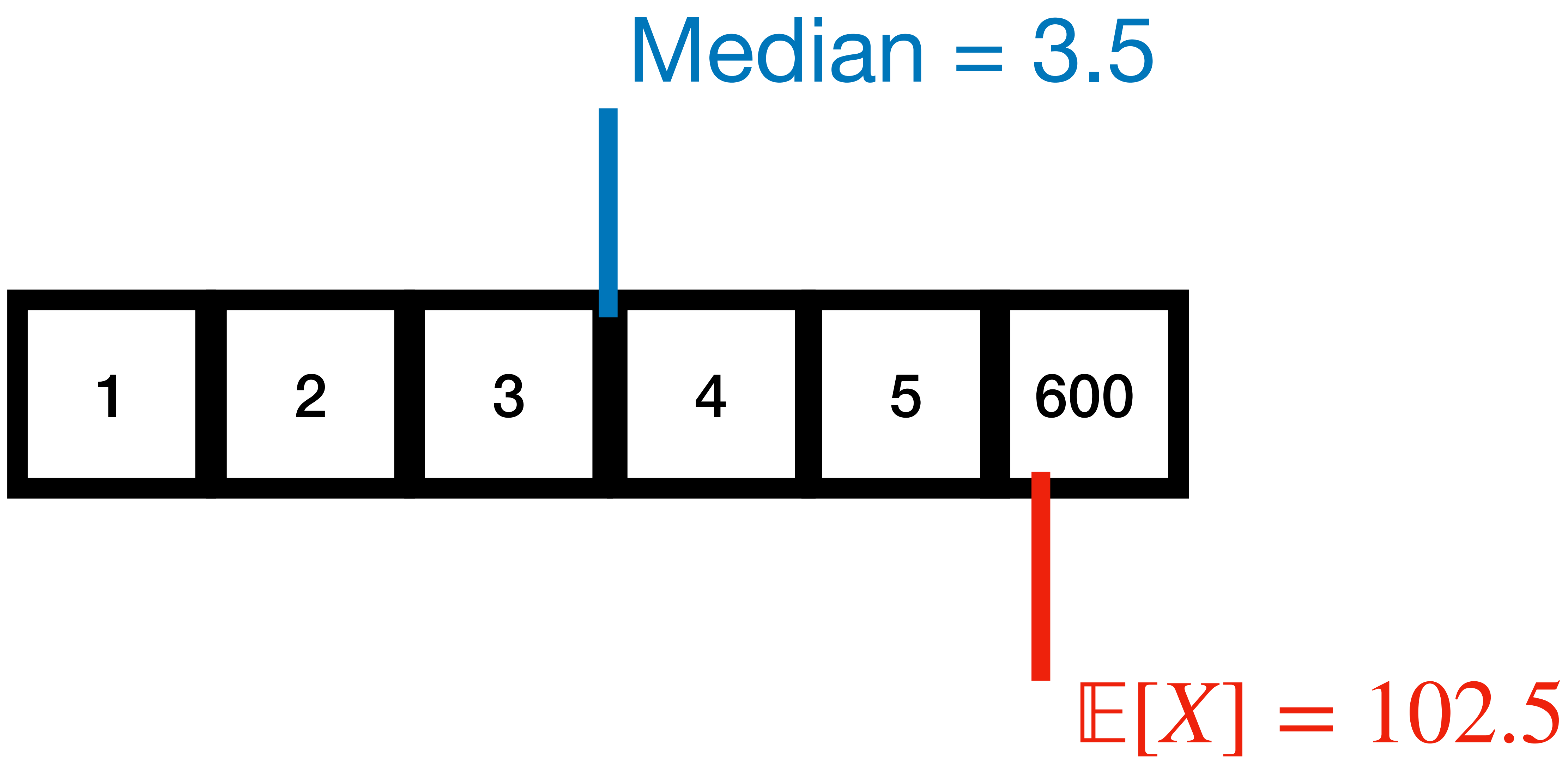
$\mathbb{E}[X] = 3.5$

$$\frac{1}{6} \times 1 + \frac{1}{6} \times 2 + \ldots \frac{1}{6} \times 6 = 3.5$$

Experiment: roll dice

Random variable:
outcome -> number on die

Expected value is sensitive to outliers

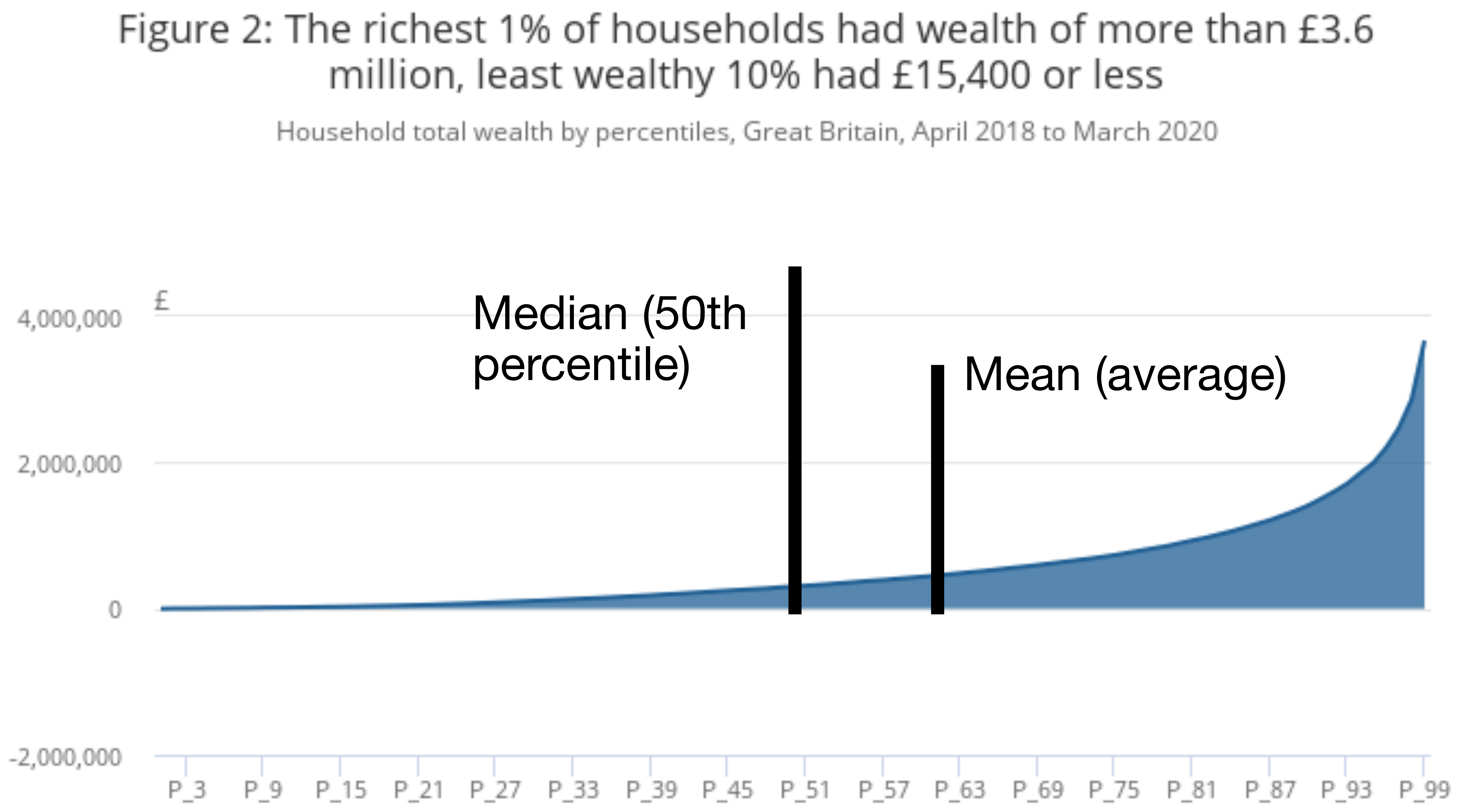


$$\frac{1}{6} \times 1 + \frac{1}{6} \times 2 + \ldots \frac{1}{6} \times 600 = 102.5$$

Experiment: roll dice

Random variable:
outcome -> number on die

Expectation of a random variable



Experiment: pick random UK household

Random variable: household -> net wealth

Source: Office for National Statistics

Expectation of function of random variable

What's the mean of $h(X)$ over many trials?

Example

Mean of $h(X) = X^2$ where X is outcome of dice roll

$$\frac{1}{6} \times 1^2 + \frac{1}{6} \times 2^2 + \ldots \frac{1}{6} \times 6^2 = 15\frac{1}{6}$$

$h(1)$ $h(2)$ $\ldots$

Experiment: roll dice

Random variable:
outcome -> number on die

Expectation of function of random variable

What's the mean of $h(X)$ over many trials?

$$\boxed{\mathbb{E}[h(X)] = \sum_{x \in \mathrm{supp}(X)} h(x) f_X(x)}$$

Value of $h(x)$

Probability of $X = x$

Expectation of function of random variable

Do it yourself! (At home)

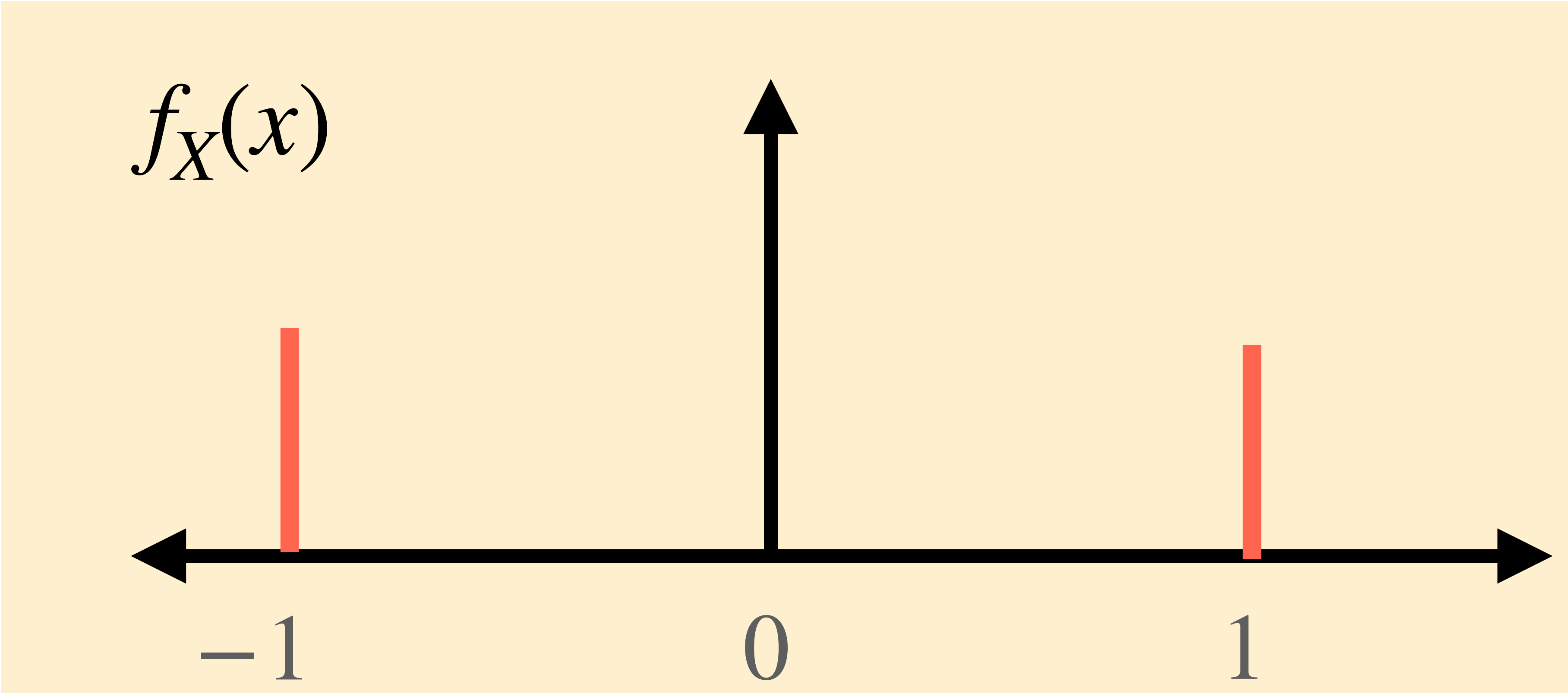
1. New random variable $Y = h(X)$

2. $\mathbb{E}[Y] = \sum_{\text{supp}(Y)} y\mathbb{P}[Y = y]$

3. Rewrite RHS in terms of X (i.e. get rid of Y)

Getting intuition on expectations

Let's consider $X \sim U(\{-1, 1\})$



$\mathbb{E}[X]$?

$\mathbb{E}[X^2]$?

Getting intuition on expectations

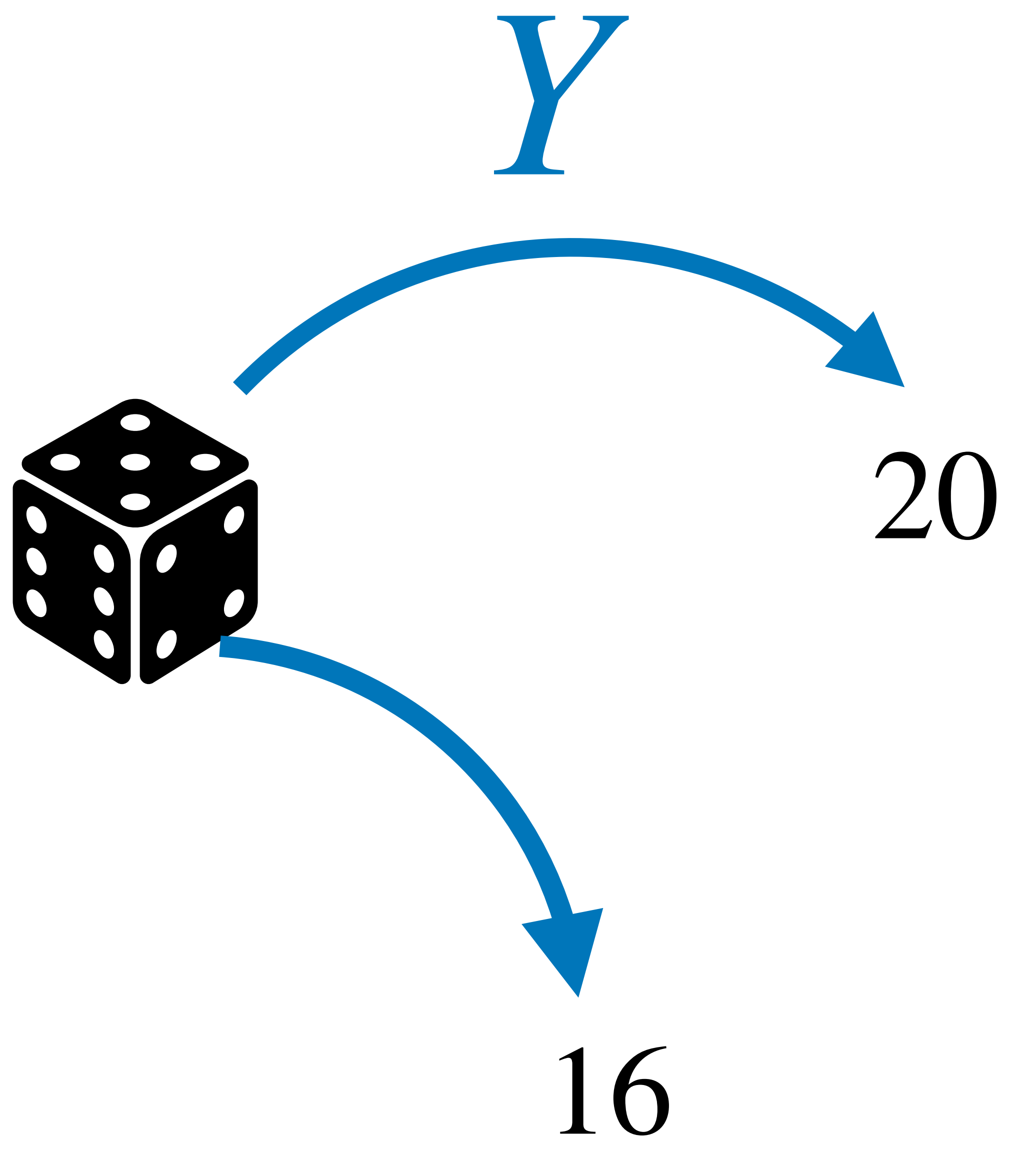
Expectations don’t play nicely with nonlinearities!

$$\mathbb{E}[X] = 0$$
$$\mathbb{E}[X^2] = 1$$

Expectation satisfies linearity

$X =$ Dice number

$Y = 4X$



$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$$

Linearity means I can separate out
the expectations under addition!

$$\mathbb{E}[Y] = \mathbb{E}[4X] = 4\mathbb{E}[X]$$

Linearity means I can take the
scalar out of the expectation!

Expectation satisfies linearity

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y] \qquad \forall a, b \in \mathbb{R}$$

(For any real numbers a and b)

…for any random variables X, Y

Examples

$$\mathbb{E}[\sum_i a_i X_i] = \sum_i a_i \mathbb{E}[X_i] \qquad \forall a_i \in \mathbb{R}$$

Expectation satisfies linearity

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y] \qquad \forall a, b \in \mathbb{R}$$

(For any real numbers a and b)

…for any random variables X, Y

"If you're adding random variables or multiplying them by a constant number, you can do the same to their expectations"

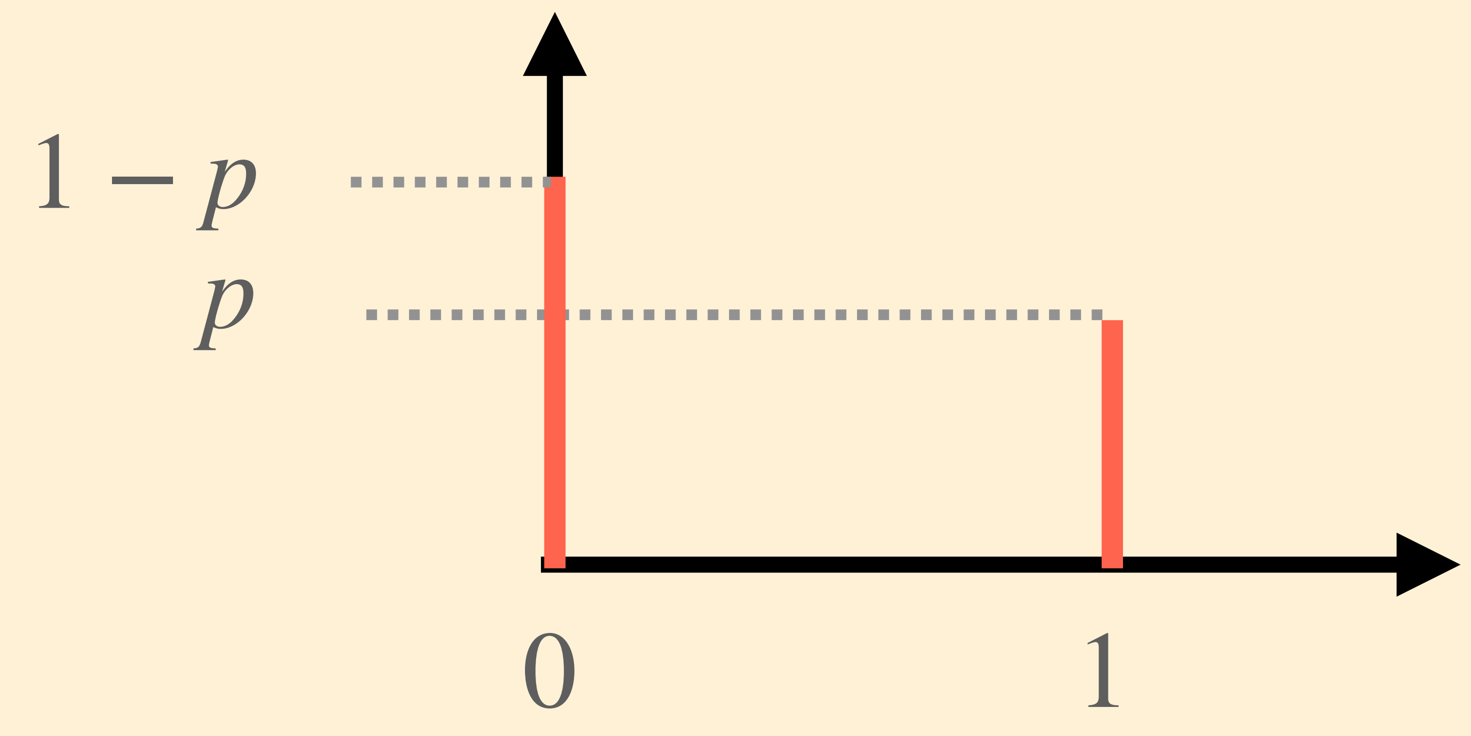
Expectation preserved under multiplication?

$$\underset{p}{\mathbb{E}}[XY] = \underset{p}{\mathbb{E}}[X]\underset{p}{\mathbb{E}}[Y]?$$ ✗

EG X is Bernoulli:

$$f_X(x) = \begin{cases} p & \text{if } x = 1 \\ 1 - p & \text{if } x = 0 \end{cases}$$

EG $Y = X$, so $XY = X^2$



Summarising random variables

How big is a random variable (on average)

Expected value: $\mathbb{E}[X]$ or $\mu(X)$

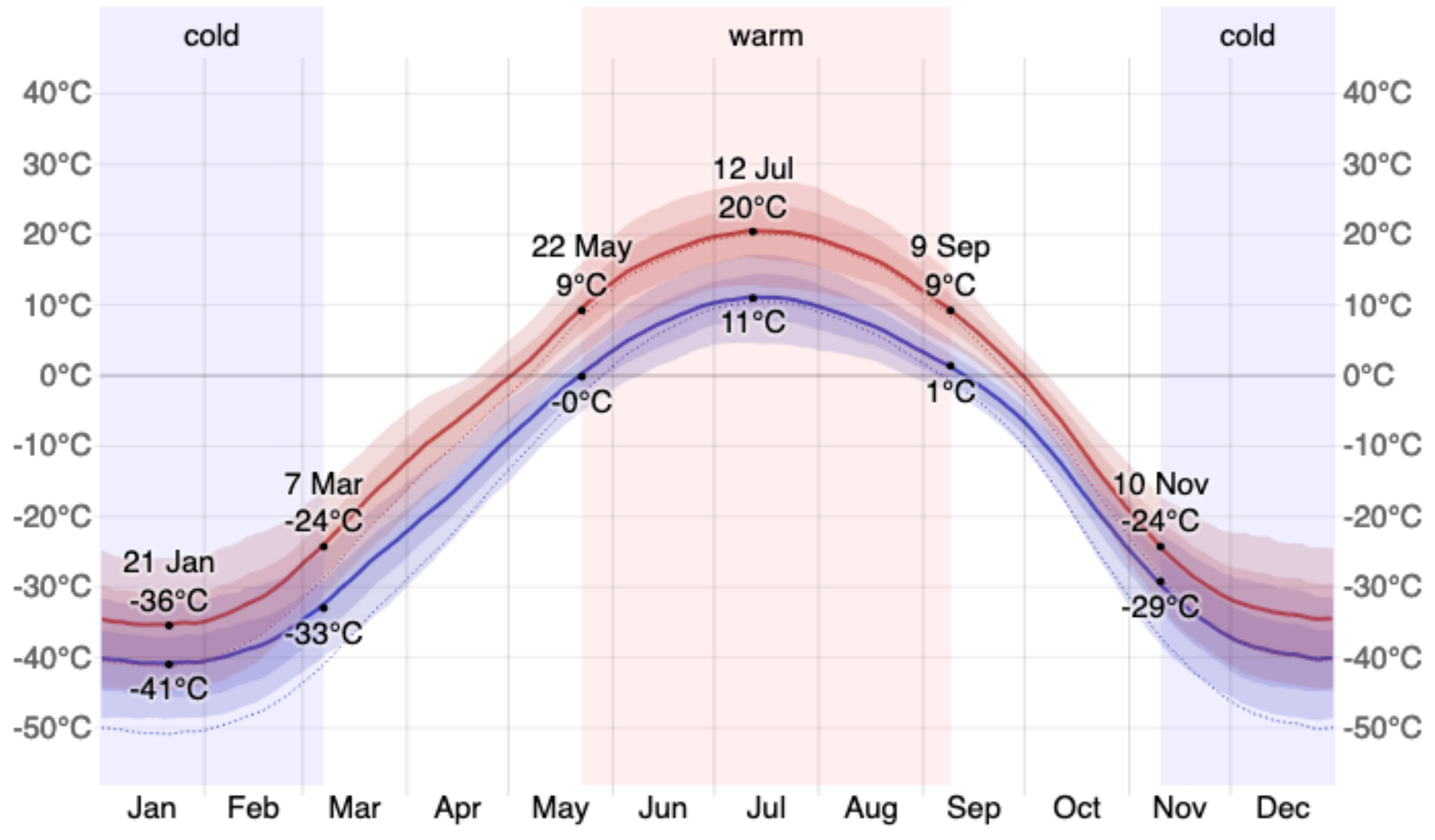
How variable is a random variable (on average)

Variance: $\mathrm{Var}[X]$ or $\sigma^2(X)$

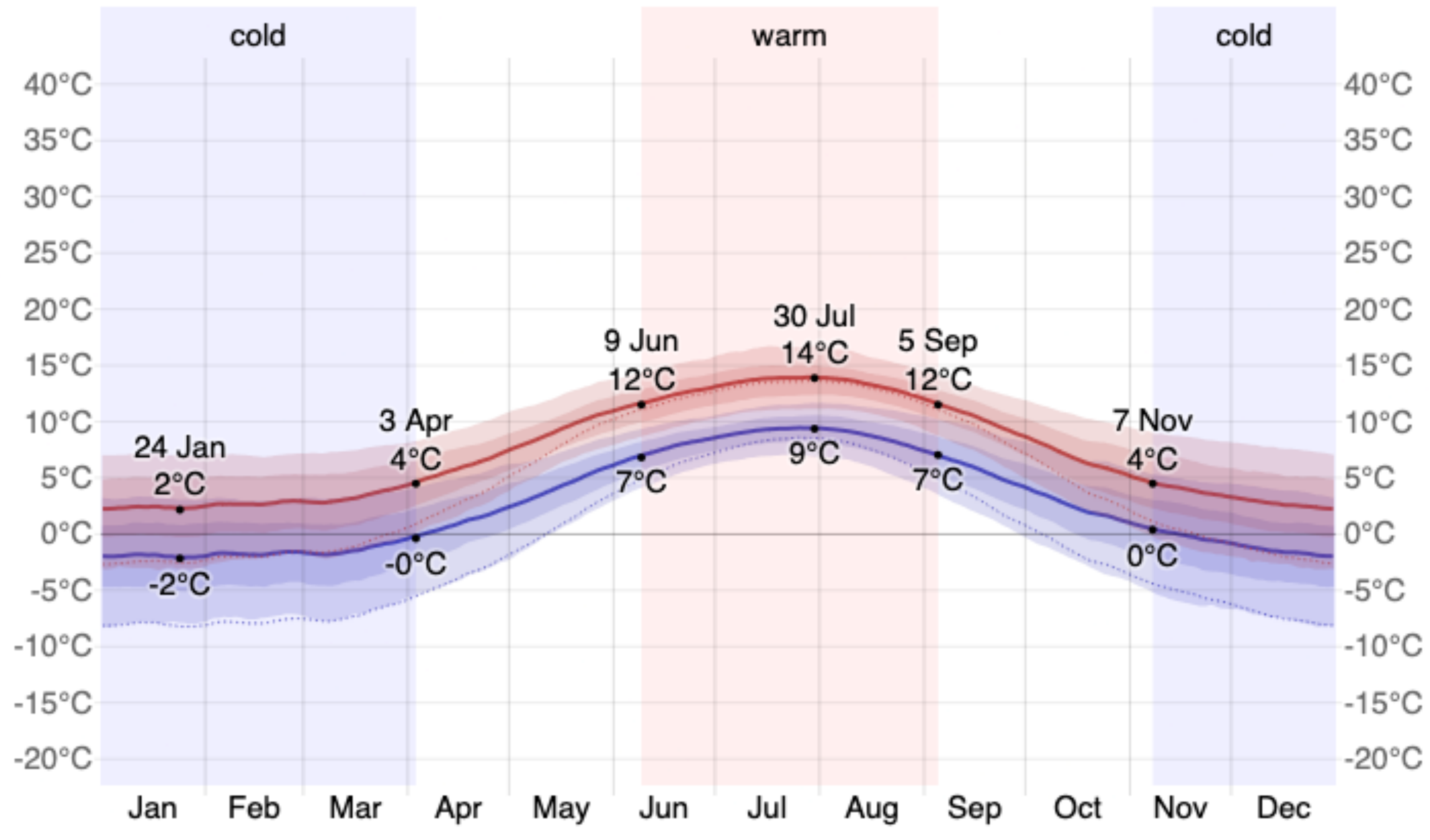
Temperature

Experiment: take temperature on a random day in…

Verkhoyansk



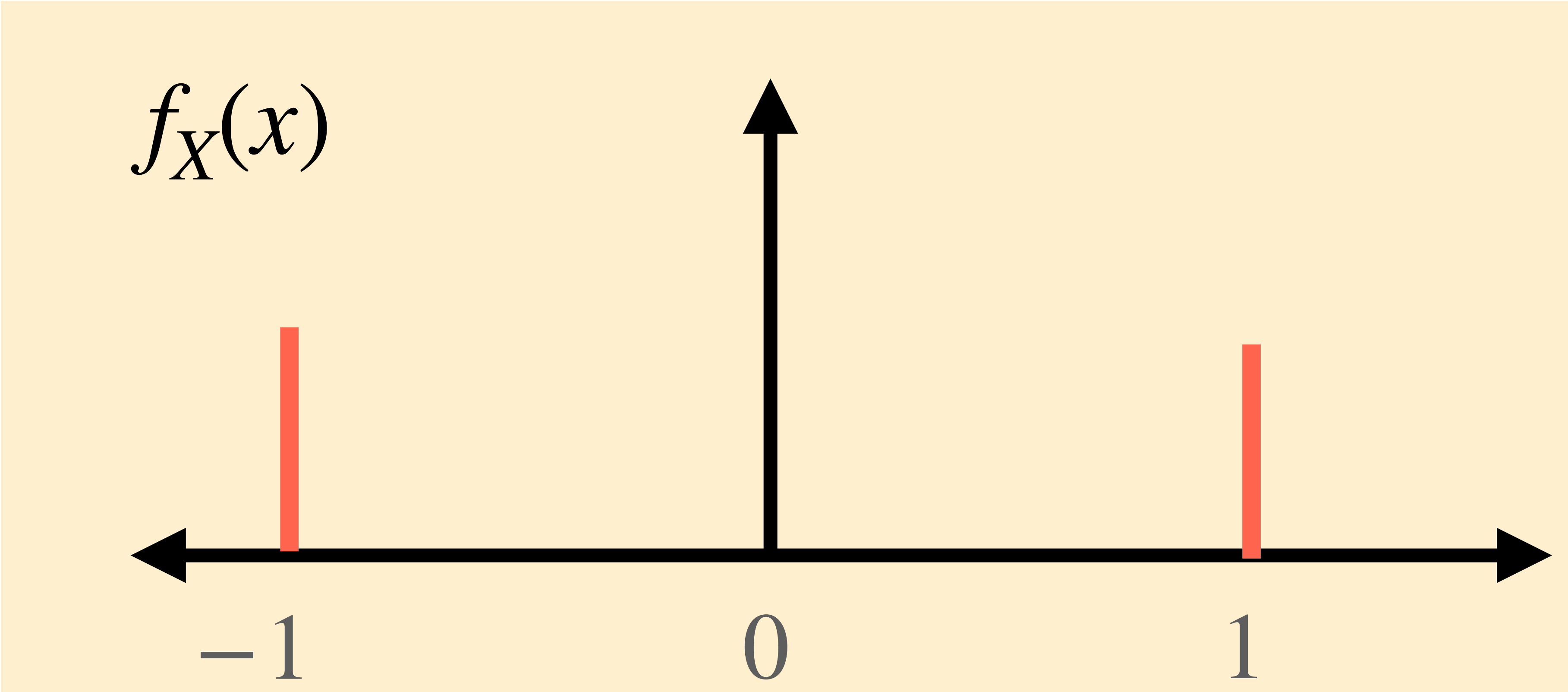
Reykjavik



Variance of an example random variable

Let's consider $X \sim U(\{-1,1\})$ again

How far from mean do we expect?



Expectation of:

$(X - \mathbb{E}[X])^2$? 1

Why not $X - \mathbb{E}[X]$?

Conceptualising variance

How far does a random variable usually fall from its expected value?

$\Updownarrow$

How much uncertainty in the random variable?

$\Updownarrow$

$$\boxed{\mathbb{E}[(X - \mathbb{E}[X])^2]}$$

Conceptualising variance

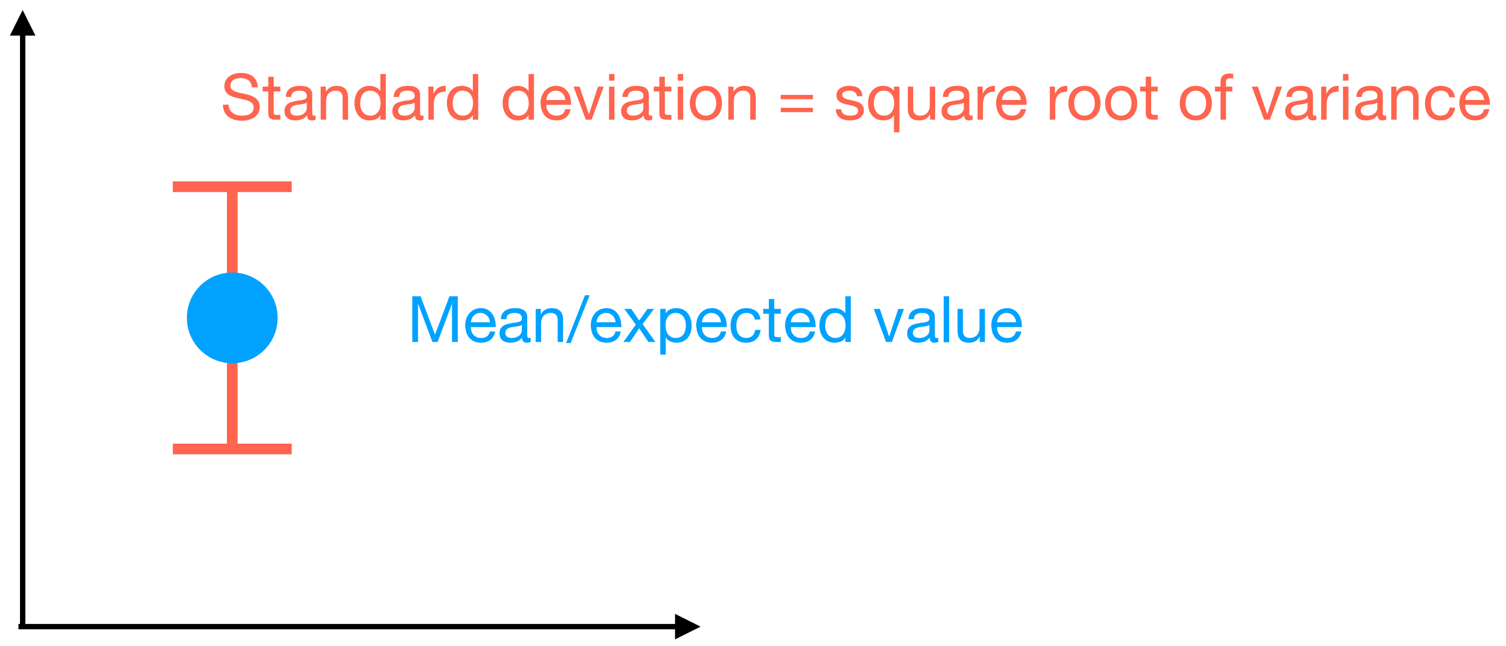
How far does a random variable usually fall from its expected value?

↕

How much uncertainty in the random variable?

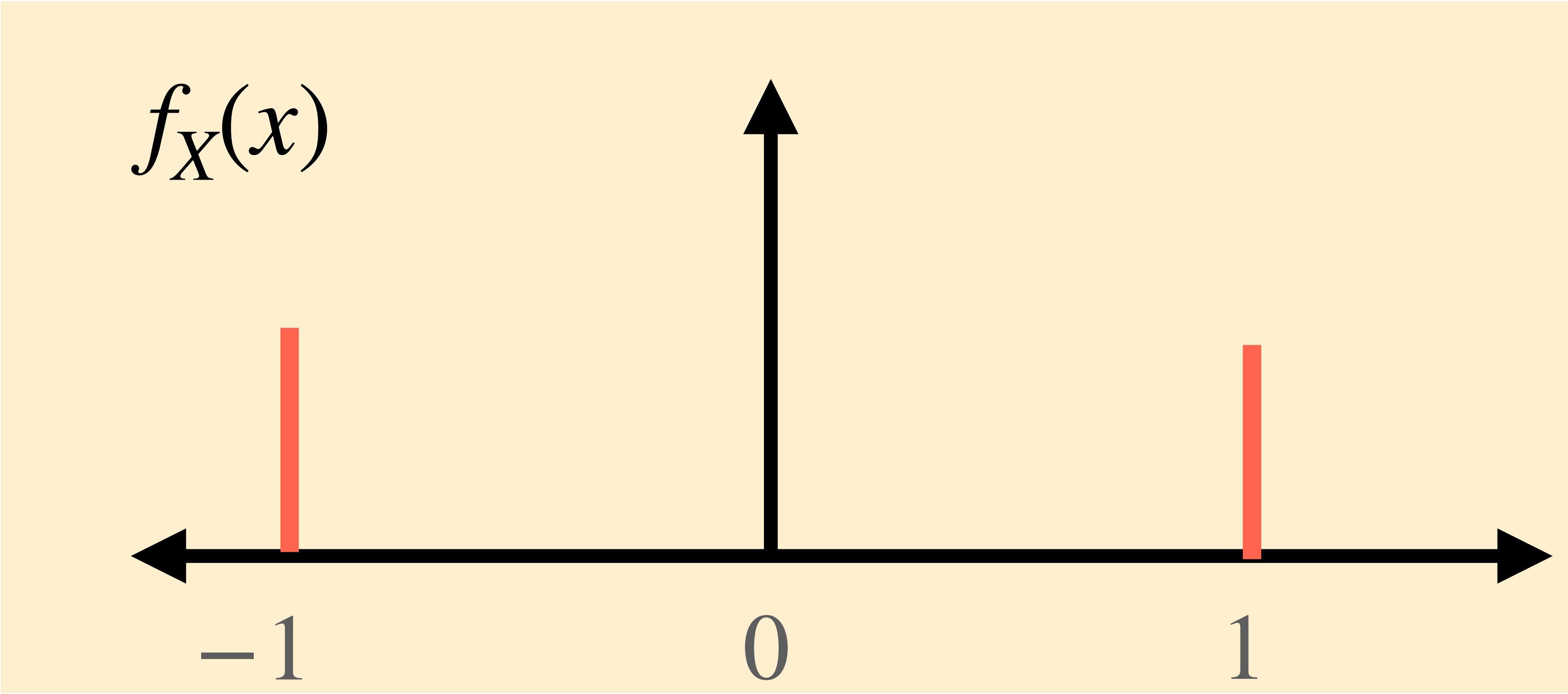
↕

$$\boxed{\mathbb{E}[(X - \mathbb{E}[X])^2]}$$



Variance of an example random variable

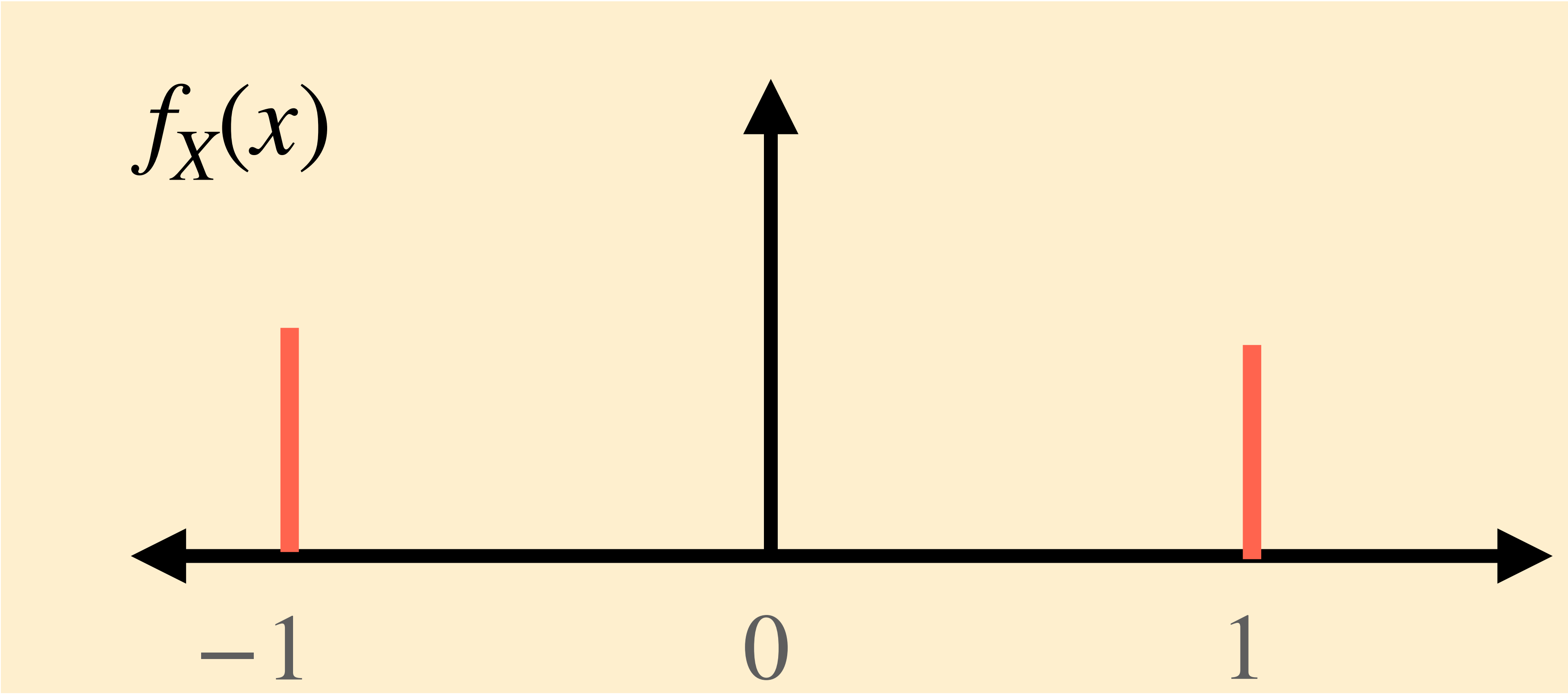
Let's consider $X \sim U(\{-1,1\})$ again



$$\mathbb{E}[(X - \mathbb{E}[X])^2]?$$

Variance of an example random variable

Let's consider $X \sim U(\{-1, 1\})$ again



$$\mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] = 1$$

Algebra on the expected value

$$\mathbb{E}[(X - \mathbb{E}[X])^2]$$

$$= \mathbb{E}\left[X^2 - 2X\mathbb{E}[X] + \mathbb{E}[X]^2\right]$$

$$= \mathbb{E}[X^2] - 2\mathbb{E}\left[X\mathbb{E}[X]\right] + \mathbb{E}\left[\mathbb{E}[X]^2\right]$$

(Linearity of expectation)

Algebra on the expected value

How to simplify $\mathbb{E}\left[\mathbb{E}[X]\right]$ and similar?

Well $\mathbb{E}[X]$ is a number, not a random variable! Apply linearity

$$= \mathbb{E}[X^2] - 2\mathbb{E}\left[X\mathbb{E}[X]\right] + \mathbb{E}\left[\mathbb{E}[X]^2\right]$$

$$= \mathbb{E}[X^2] - 2\mathbb{E}[X]^2 + \mathbb{E}[X]^2$$

Two equivalent expressions for the variance

$$\boxed{\mathbb{E}[(X - \mathbb{E}[X])^2 = \mathbb{E}[X^2] - \mathbb{E}[X]^2}$$

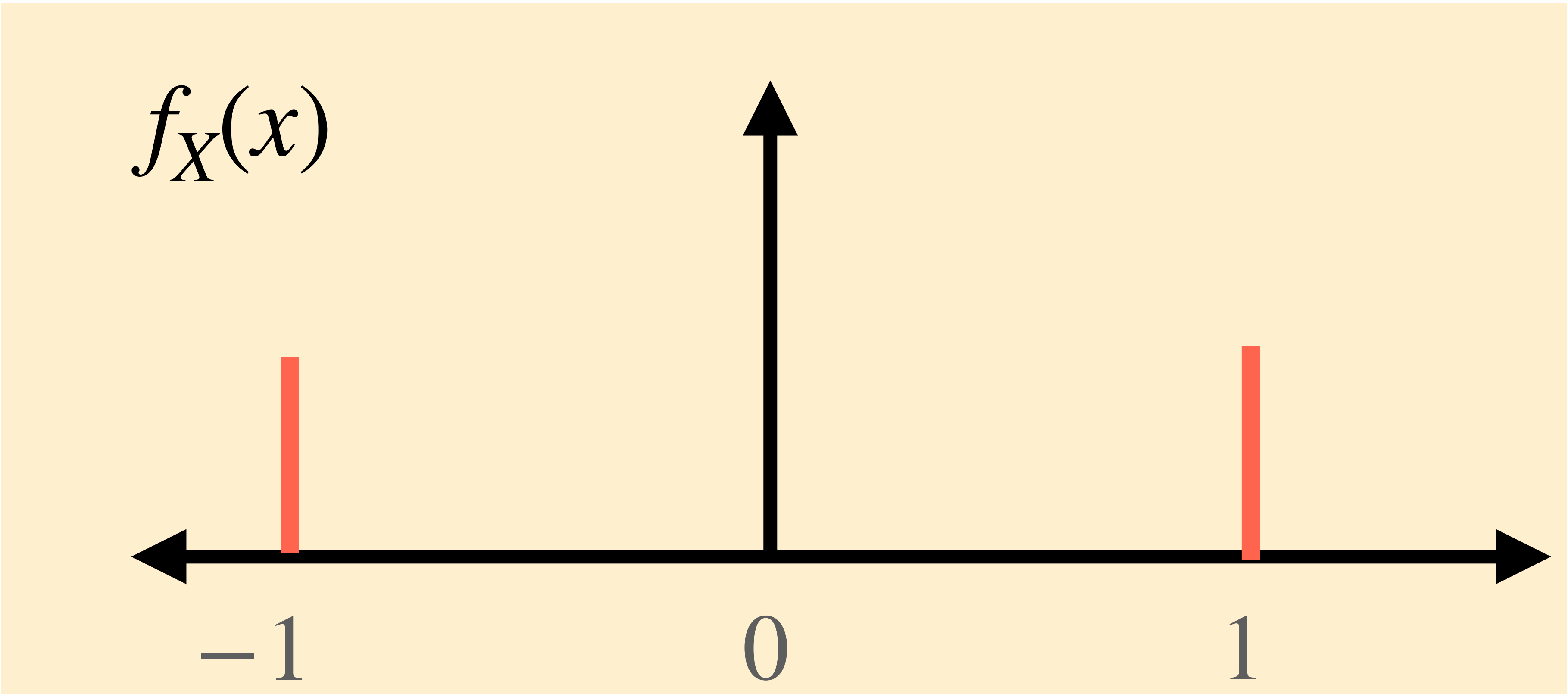
$$= \mathbb{E}[X^2] - 2\mathbb{E}\left[X\mathbb{E}[X]\right] + \mathbb{E}\left[\mathbb{E}[X]^2\right]$$

$$= \mathbb{E}[X^2] - 2\mathbb{E}[X]^2 + \mathbb{E}[X]^2$$

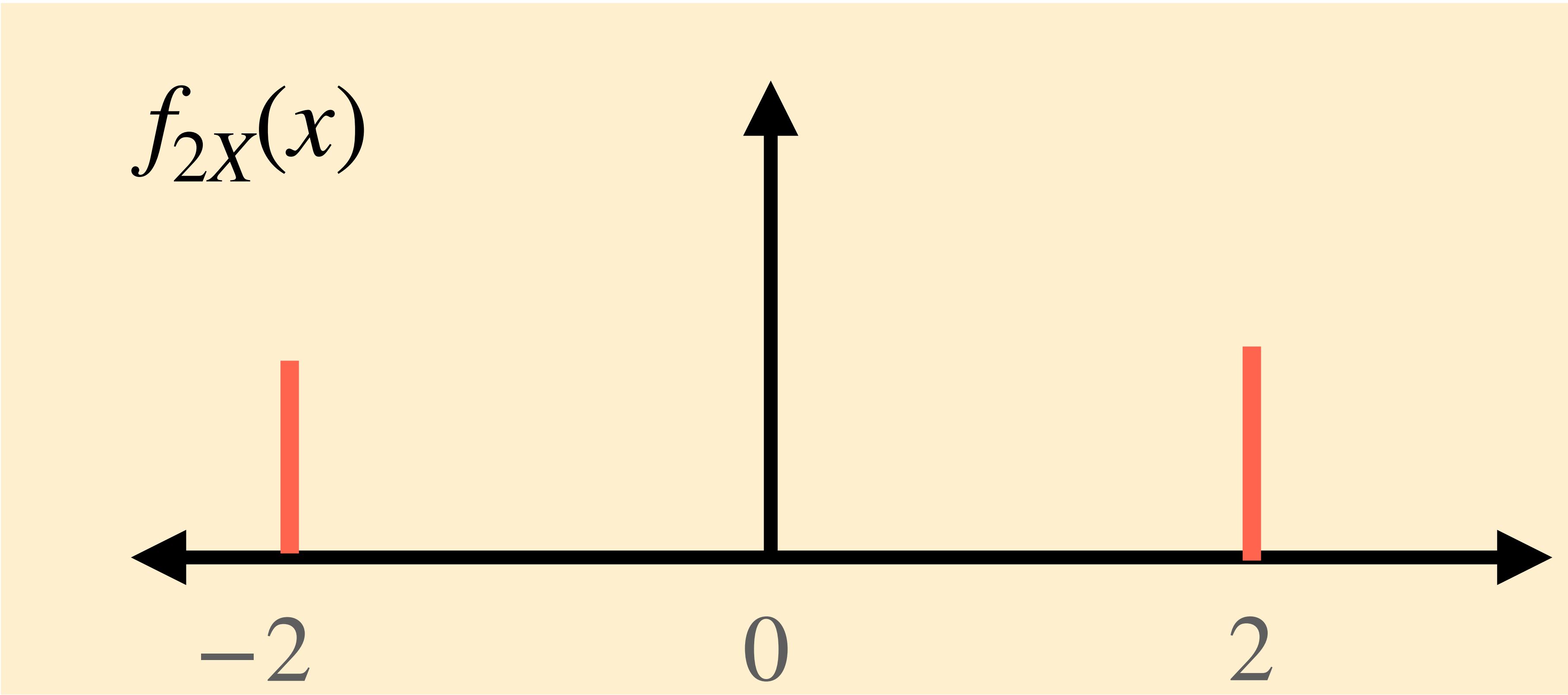
How does the variance scale?

Let's consider $X \sim U(\{-1,1\})$ again

X



$2X$



$\mathrm{Var}[X] =$
$\mathbb{E}[(X - \mathbb{E}[X])^2] = 1$

$\mathrm{Var}[2X] = ?$

4!

Algebra on the variance

Homework

Use linearity of expectation to convince yourself that:

$$\text{Var}(cX) = c^2\text{Var}(X)$$

$$c \in \mathbb{R}$$

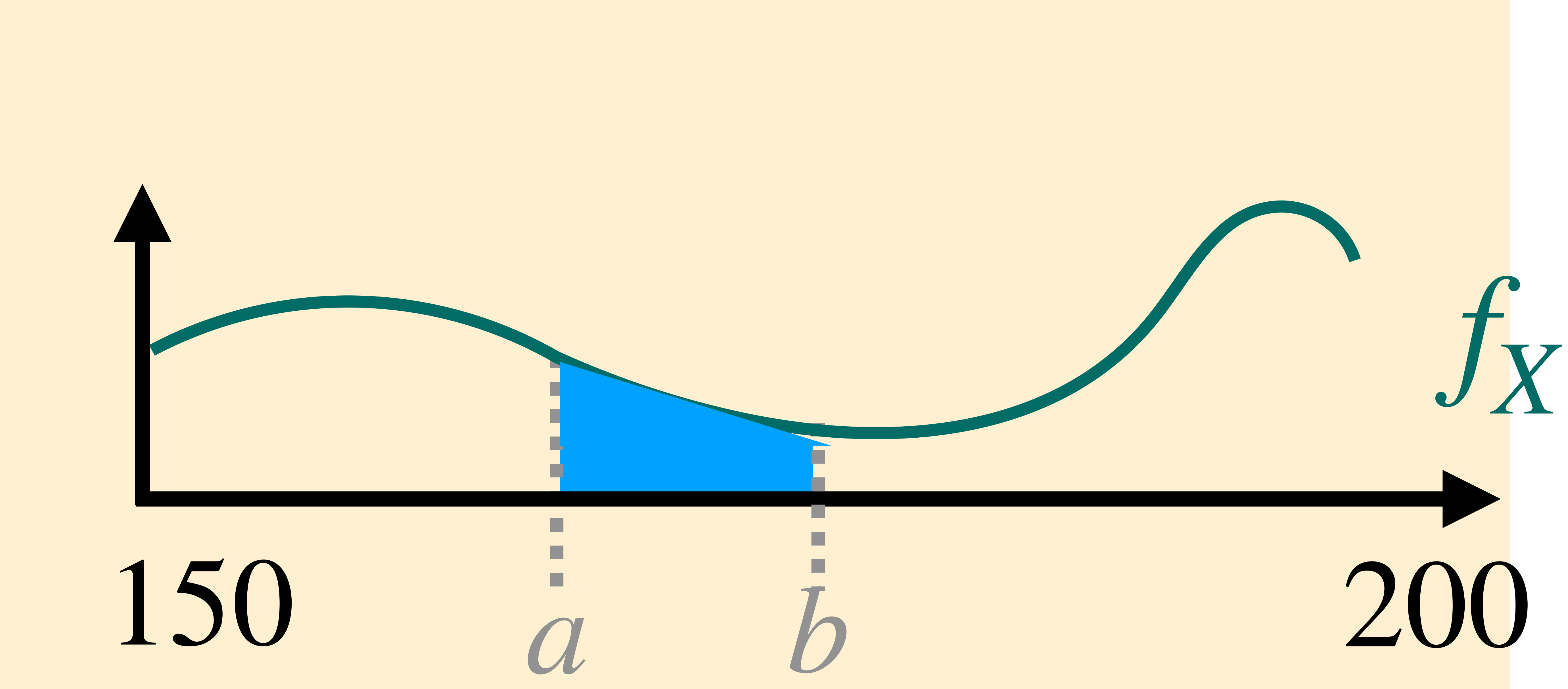
Recap: Probability density functions

Continuous random variables only

$\mathbb{P}[X = \text{anything}] = 0$

Instead, look for probability X is between values:

$\mathbb{P}[a \leq X \leq b] = ?$



$$\boxed{\mathbb{P}[x \in (a, b)] = \int_a^b f_X(x)\ dx}$$

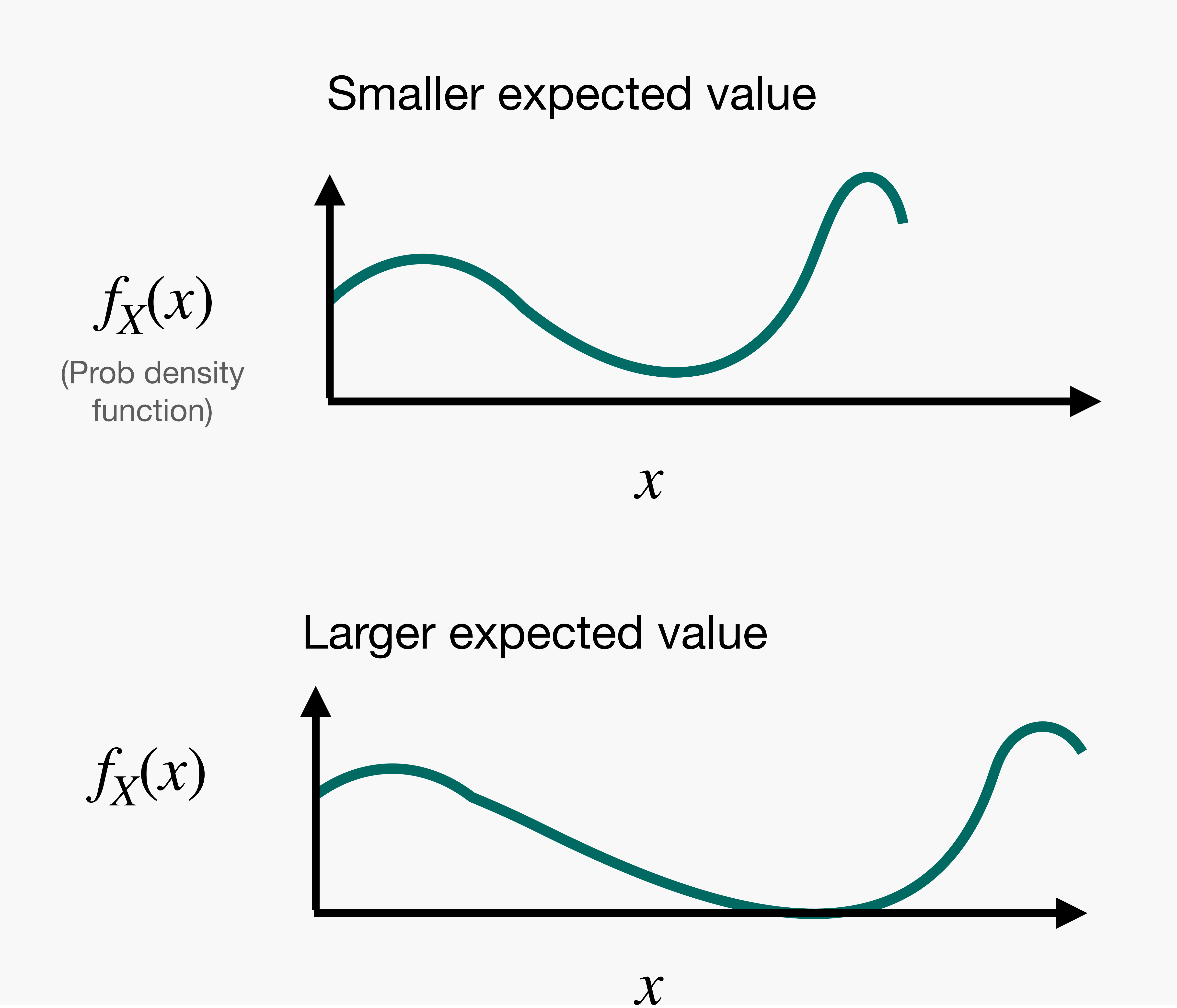
Expectation of a continuous random variable

Discrete:

$$\mathbb{E}[X] = \sum_{x \in \text{supp}(X)} x \times \mathbb{P}[X = x]$$

Continuous:

$$\mathbb{E}[X] = \int_{x \in \text{supp}(X)} x \times f_X(x)\ dx$$

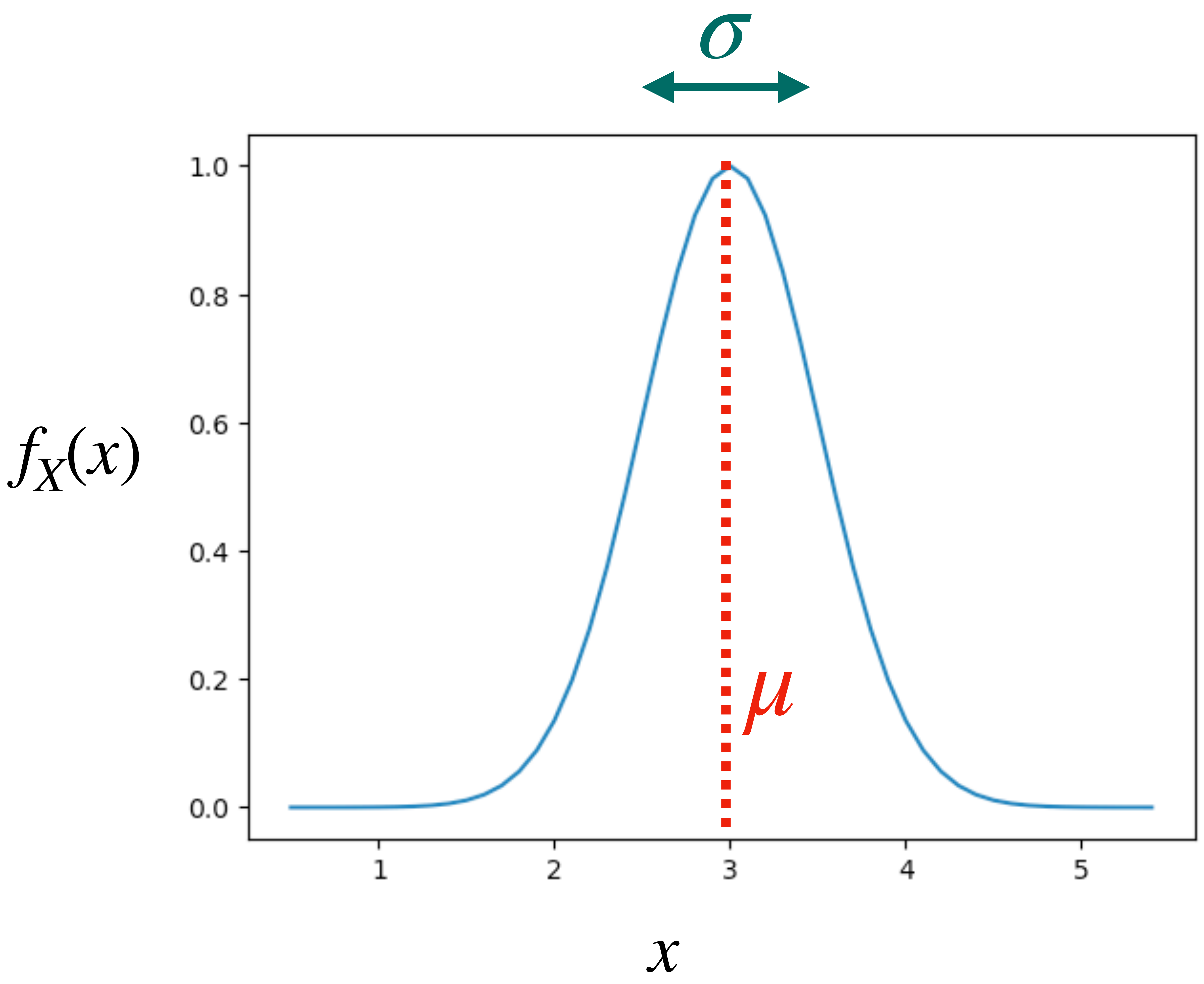


Gaussian random variable are special

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

$\Updownarrow$

"X is a Gaussian with mean μ and variance σ^2"



Also known as: bell curve, Normal distribution

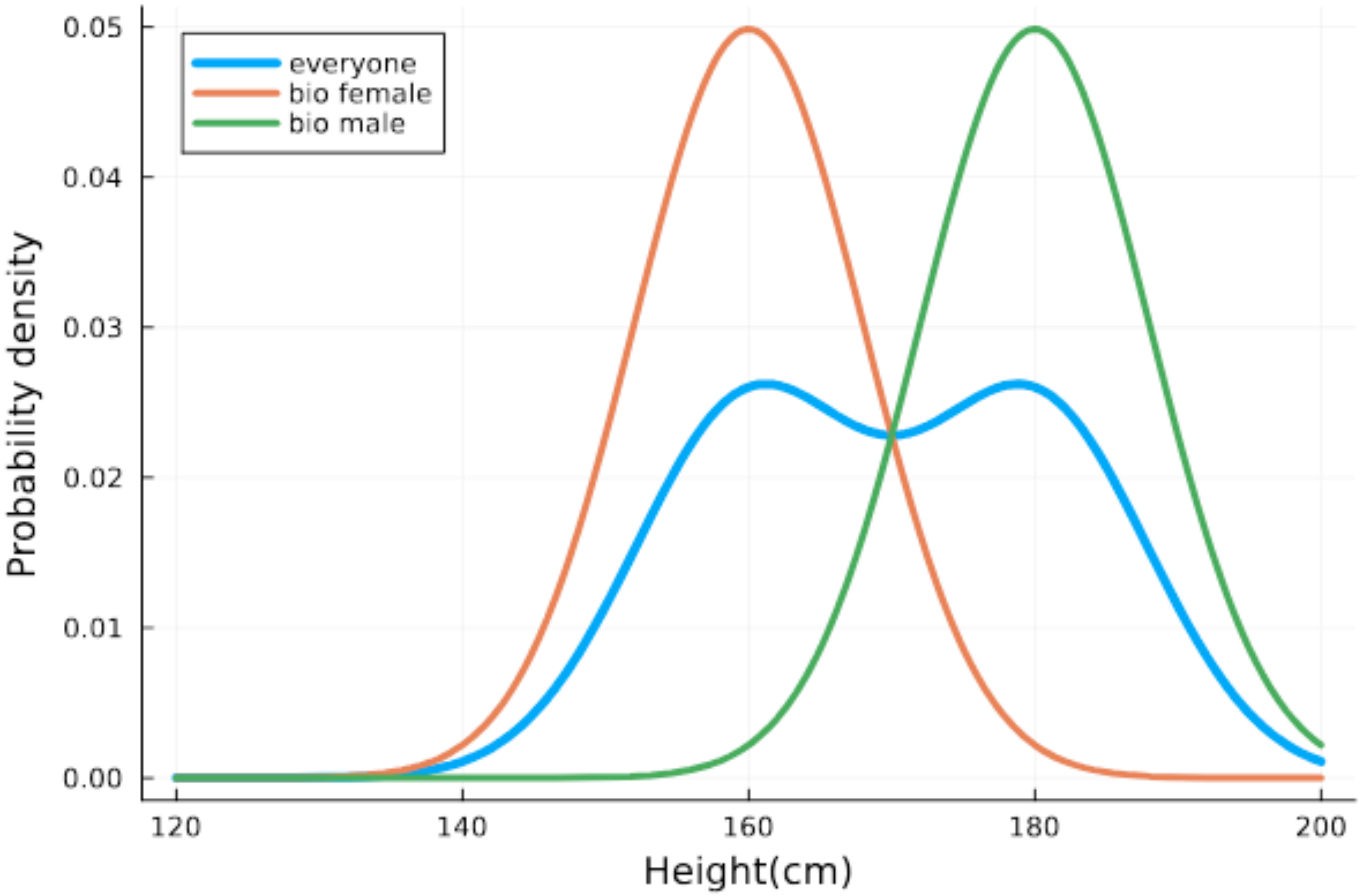
Gaussian random variable are special: Central Limit Theorem

Is height approximately normally (gaussian) distributed?

Is the average height of groups of 100 people normally distributed?

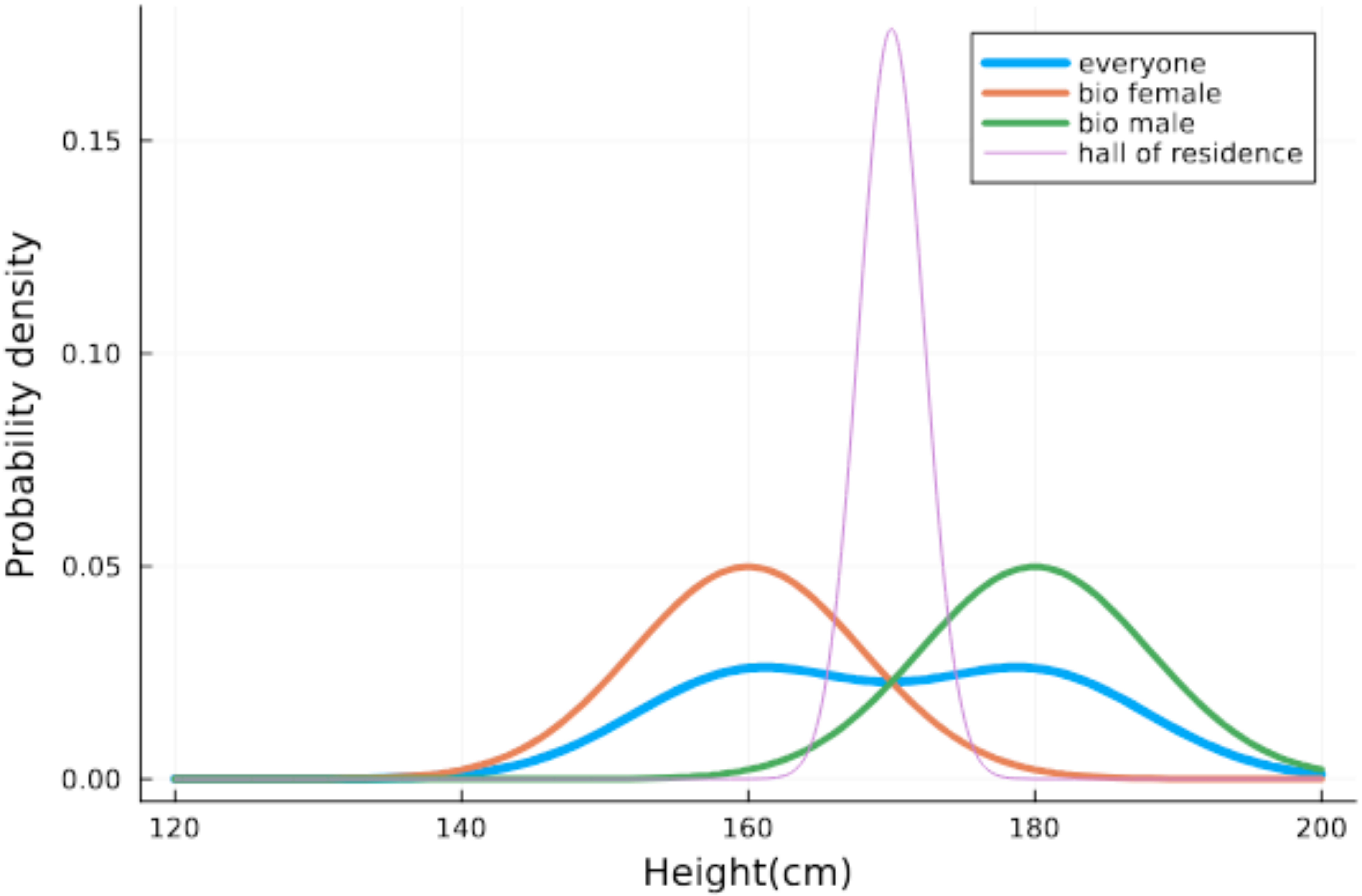
(e.g. hall of residence)

Height probably isn't Gaussian



Halls of residences might be

Why lower variance?



Central limit theorem

Random variables depend
on single experiments

$$X_i = i^{th} \text{ height}$$

Experiment E1: pick random person

Outcome: a person

Random var: person -> height

CLT is about groups of *independent, identically distributed (i.i.d)* experiments

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i = \text{mean height}$$

$n =$ group size

Experiment: run E1 100 times

Outcome: 100 people

Random var: people -> mean height

Central limit theorem

Sample mean on (large enough) groups of i.i.d experiments has a Gaussian distribution

CLT is about groups of *independent, identically distributed (i.i.d)* experiments

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i = \text{mean height}$$

n = group size

Experiment: run E1 100 times

Outcome: 100 people

Random var: people -> mean height

Central limit theorem

"Independent, identically distributed"
(remember abbreviation)

$\{X_1 \ldots X_n\}$ are i.i.d.

$\mathbb{E}[X_i] = \mu$

$\text{Var}[X_i] = \sigma^2$

Converges in distribution

$\bar{X}_n \to^d \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$

regardless of X_i distribution

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Average outcome over groups of n experiments

Central limit theorem

Caveat

How large should n be for the Central Limit Approximation to be reasonable?

Pragmatic: n=30
Theoretical: arbitrarily large! (depends on how 'non-Gaussian' X is)

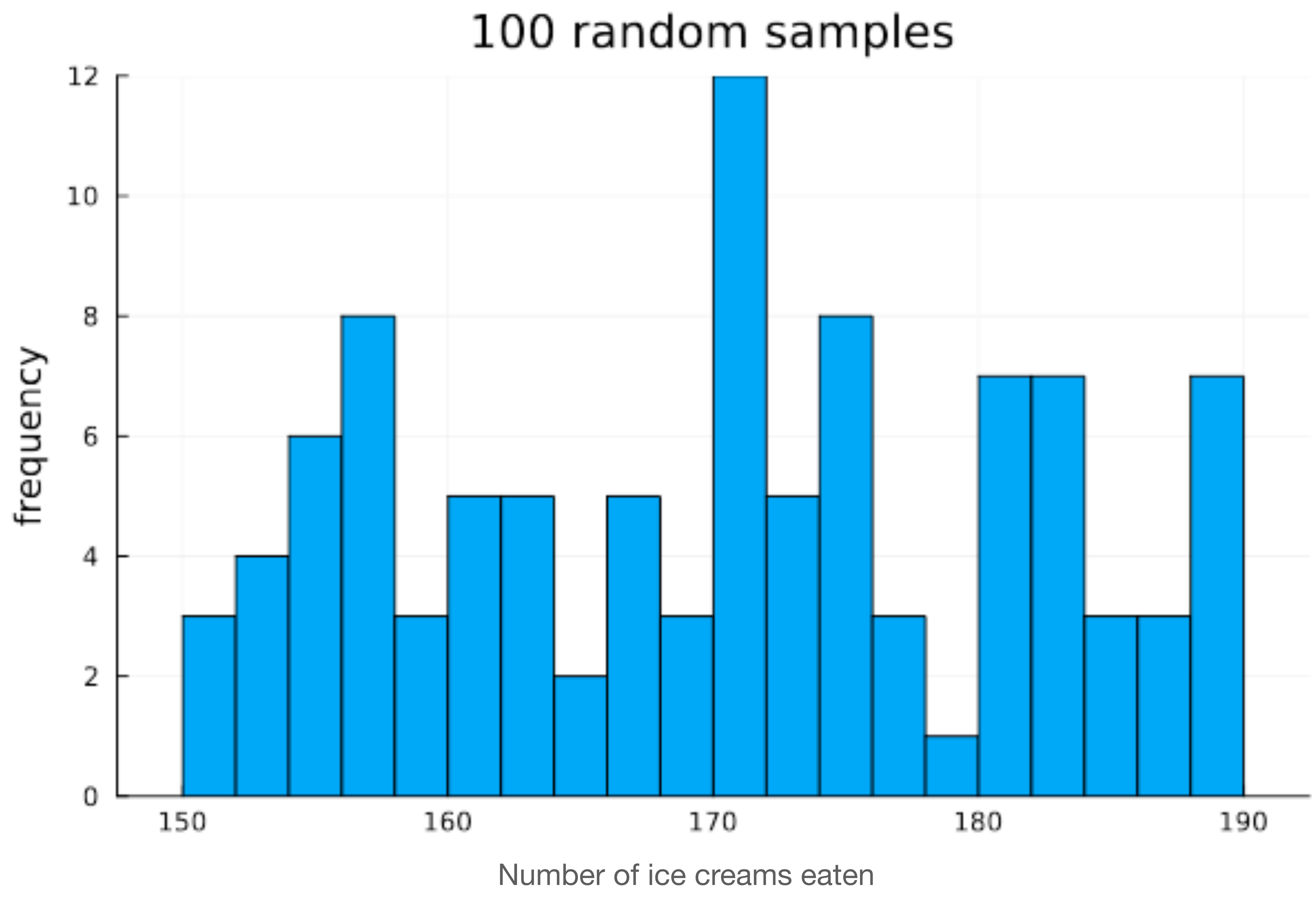
Central Limit Theorem

$\bar{X}_n \rightarrow^d \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$ regardless of X_i distribution

Converges in distribution

Central limit theorem

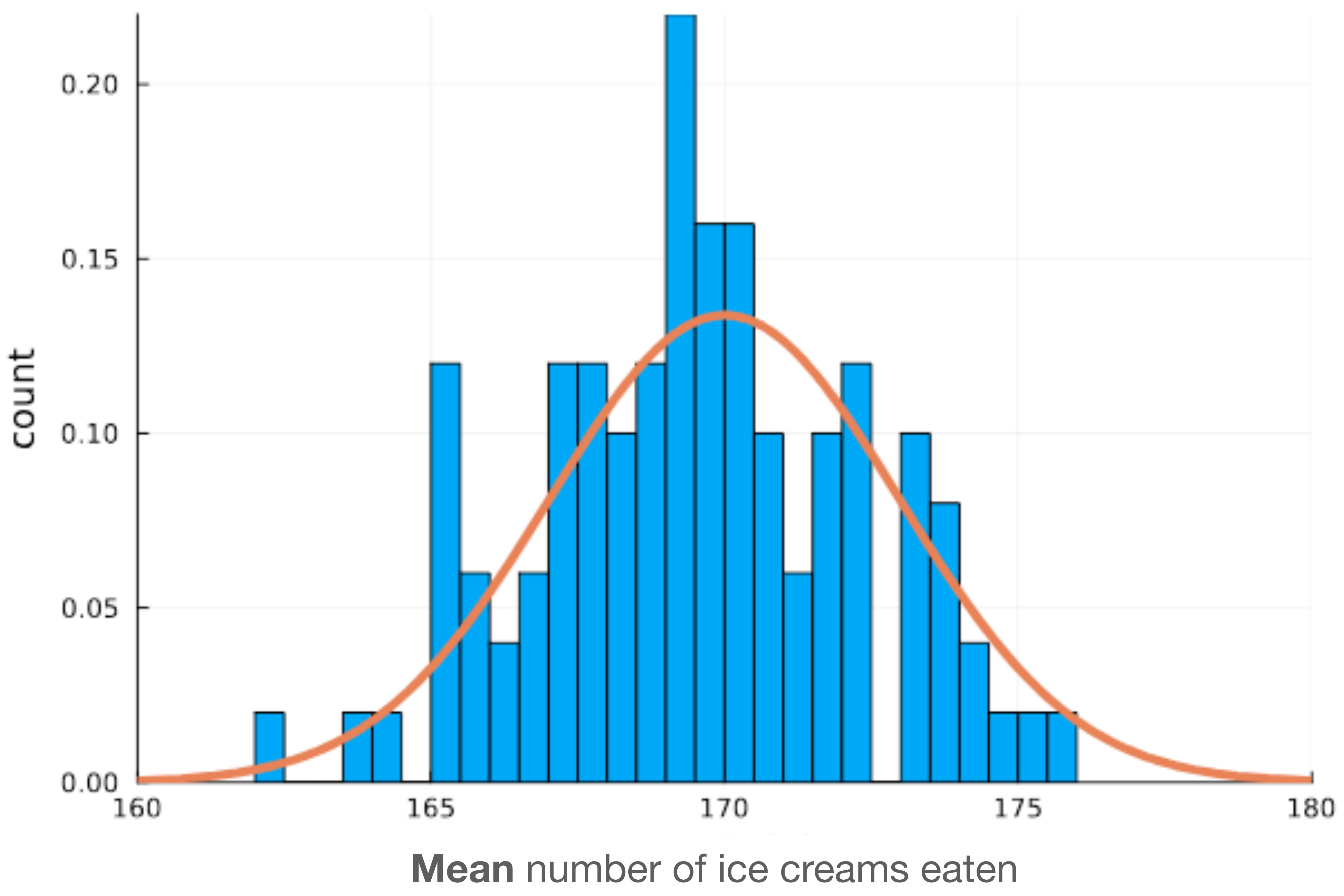
Toy assumption: Ice creams per year $\sim U(150, 190)$



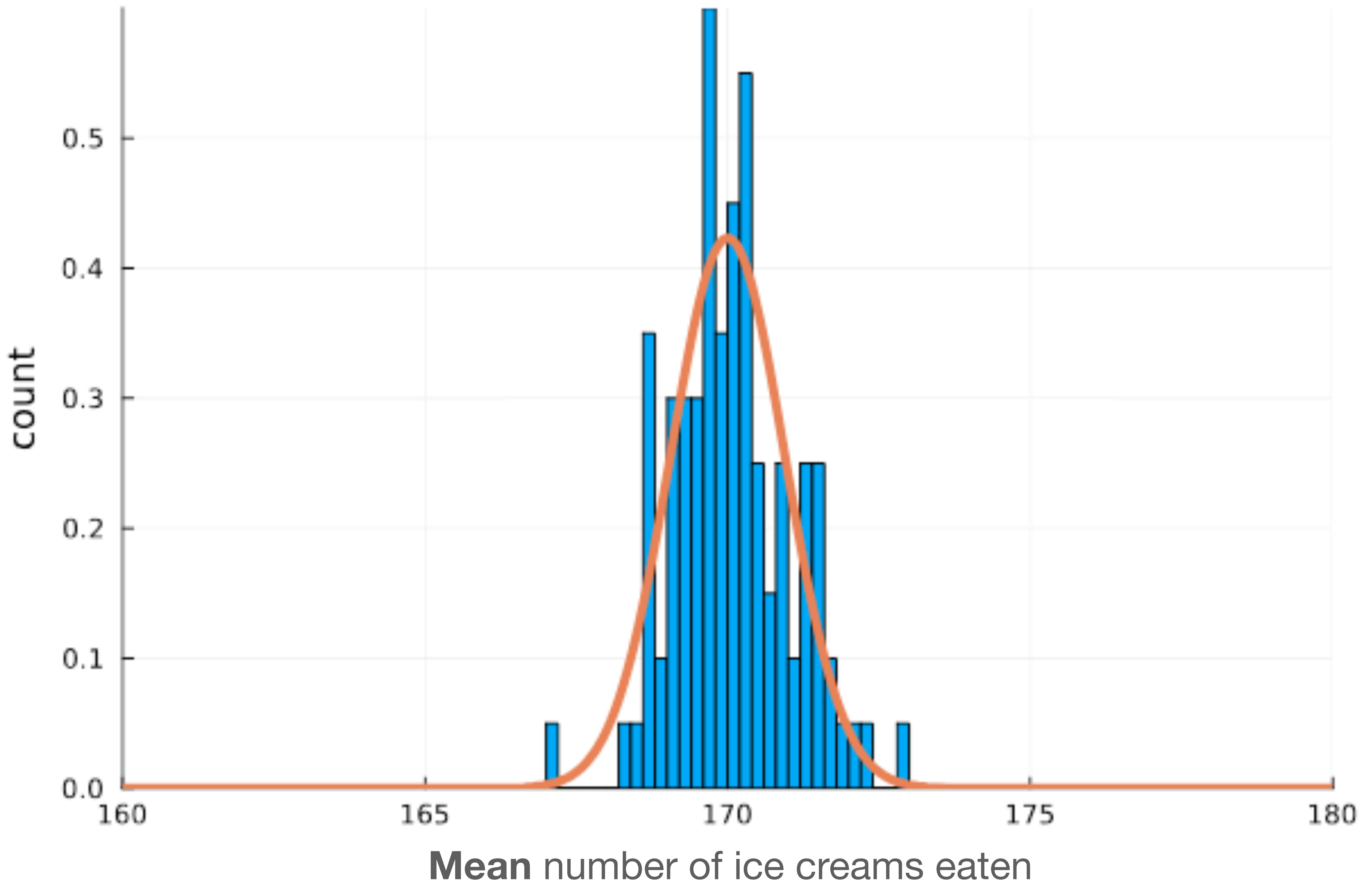
Central limit theorem

Toy assumption: Ice creams per year $\sim U(160{,}180)$

Groups of 15 people



Groups of 150 people



Why do we care about the probability density of experimental outcomes?

(Optional interlude)

Why do we care about the probability density of experimental outcomes?

Quantify and weight confidence in summaries/predictions/decisions

Can I really cluster this population into three groups?

What's their probability of paying back the loan?

Low risk/reward or high risk/reward action?

Problem: don’t know distribution of real random variables

What’s the actual probability distribution of height?

Solution:
learn from data!

Option 1 for learning distributions

Parametric estimation

1. Assume data comes from distribution (e.g. Gaussian)

2. Estimate necessary statistics of distribution (e.g. mean, variance)

Example:

Data $\{X_i\}_{i=1}^{N}$ on N people's heights

```
from scipy.optimize import curve_fit
curve_fit(gaussian, ...)
```

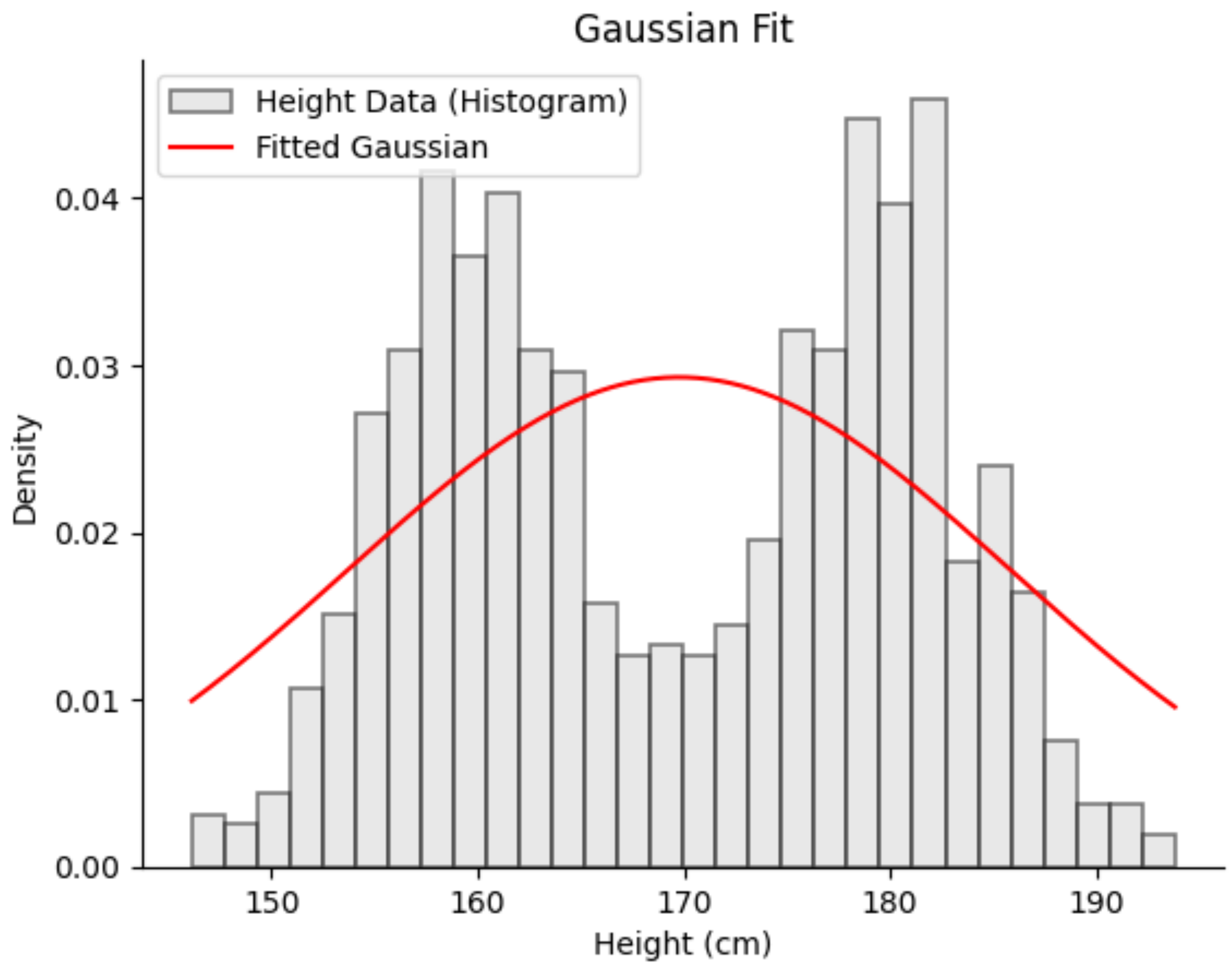
=> Answer questions like
"what's the probability of >190cm"

Option 1 for learning distributions

Parametric estimation

1. Assume data comes from distribution (e.g. Gaussian)

2. Estimate necessary statistics of distribution (e.g. mean, variance)



What's the problem?

Learning distributions

Parametric estimation

1. Assume data comes from distribution (e.g. Gaussian)
2. Estimate necessary statistics of distribution (e.g. mean, variance)

Vs

Nonparametric estimation

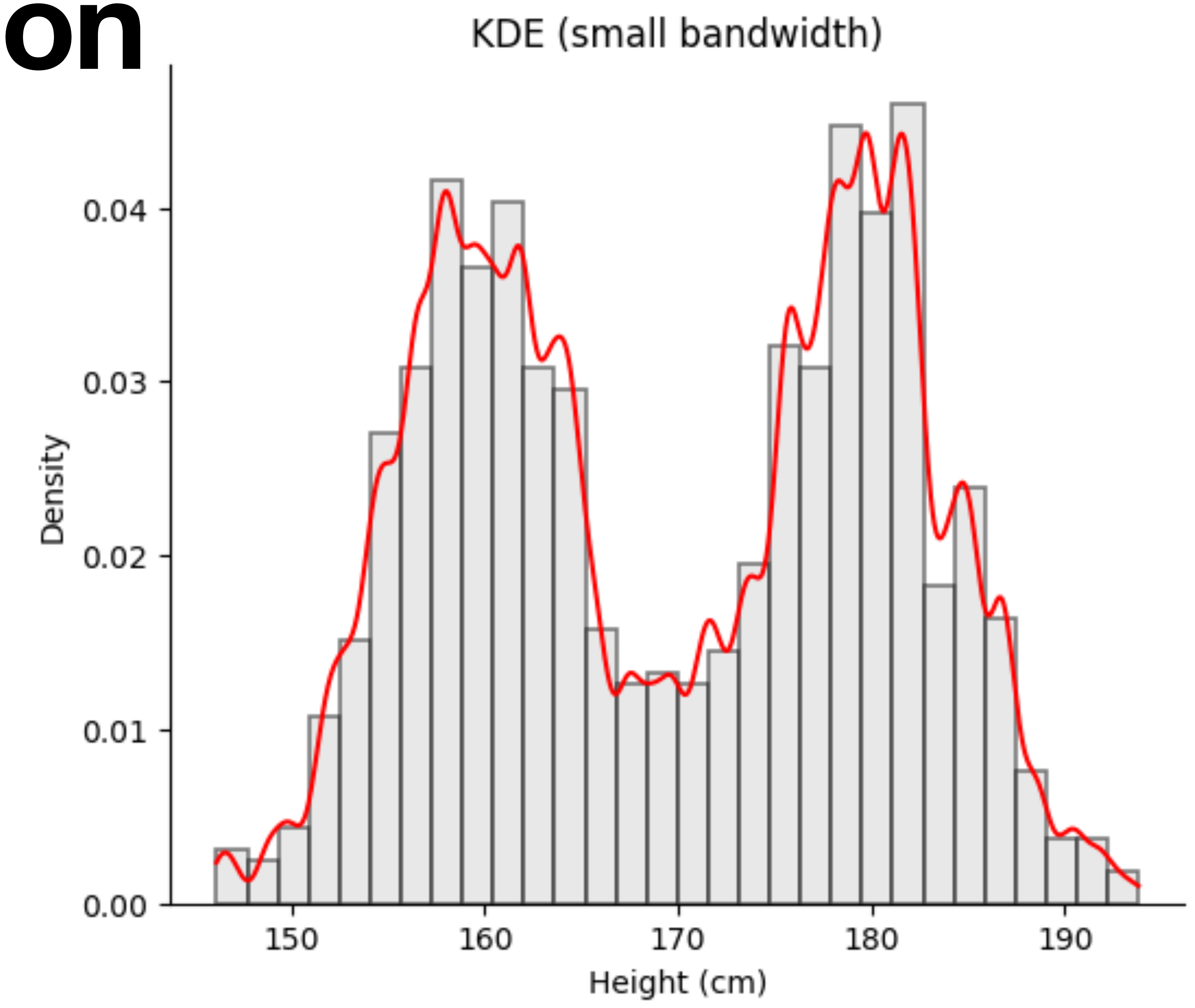
1. Don't assume predefined distribution
2. Jointly estimate statistics and distribution

Nonparametric estimation

One solution

Kernel density estimation (KDE)

"Draw a line over the histogram"



What's bad about this?

Models noisy fluctuations

```

from scipy.stats import gaussian_kde
gaussian_kde(heights, bw_method=0.05)

```

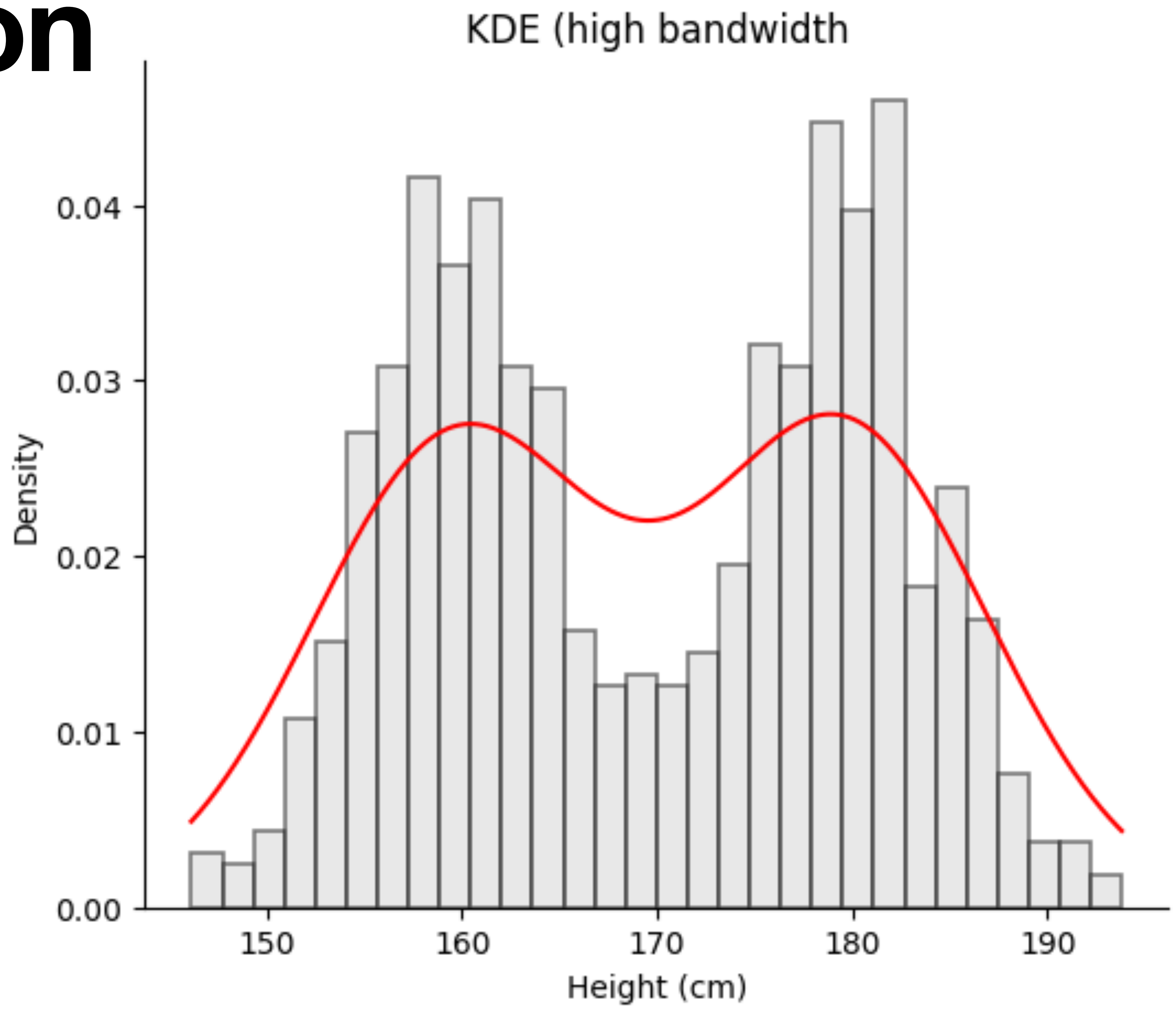
Nonparametric estimation

Better?

Not if they are meaningful fluctuations!

How can we decide?

More data!




```
from scipy.stats import gaussian_kde
gaussian_kde(heights, bw_method=0.5)
```

Fundamental tradeoff in all of applied maths

Techniques with fewer assumptions need more data

Learning distributions

Parametric estimation

1. Assume data comes from distribution (e.g. Gaussian)
2. Estimate necessary statistics of distribution (e.g. mean, variance)

Vs

Nonparametric estimation

1. Don't assume predefined distribution
2. Jointly estimate statistics and distribution

More assumptions

More data

Which to use? Subjective!

Multivariate distributions

Random variable

Quantitative question about outcome of experiment

Experiment

Choose random person

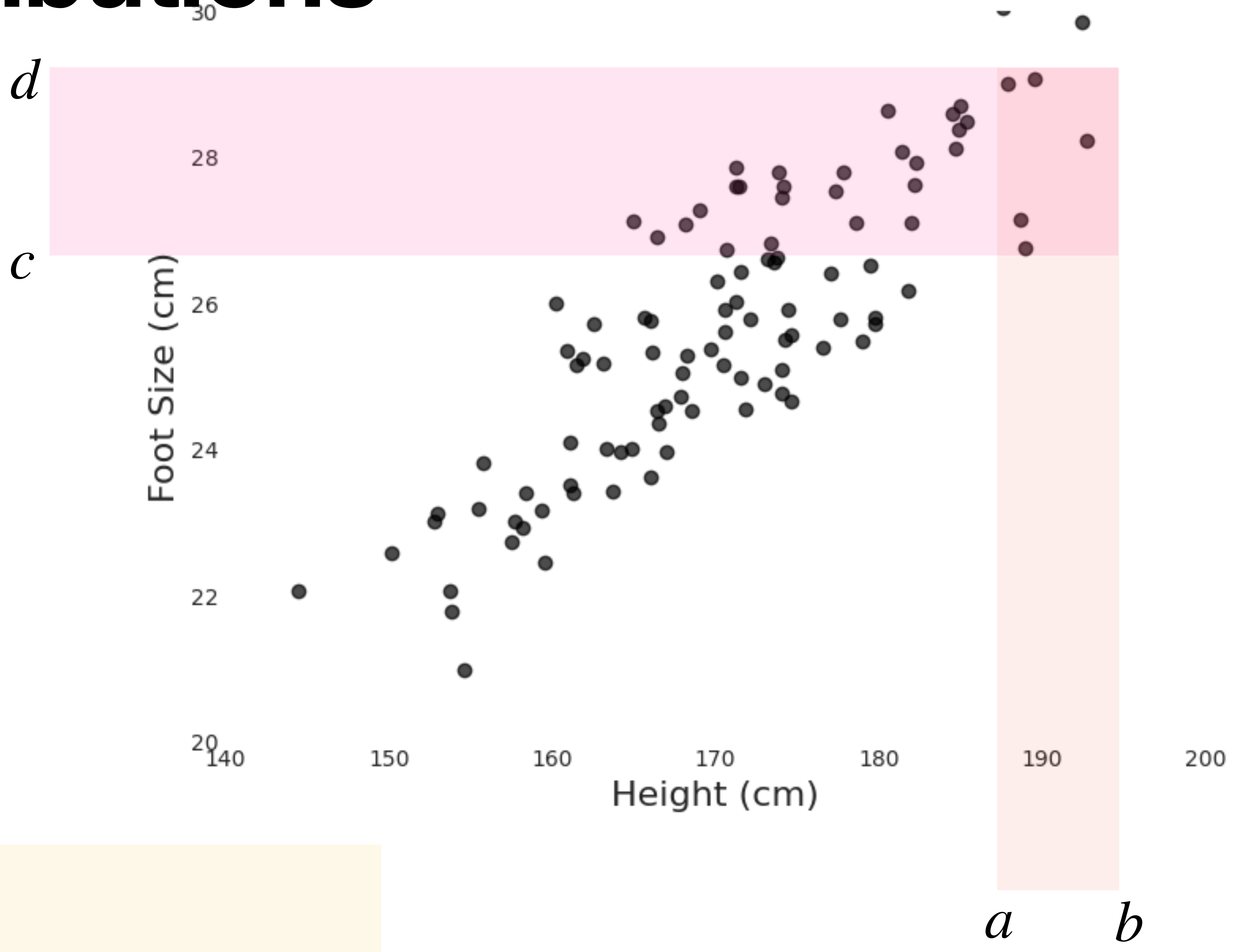
Random variables

Height, foot size

What if we have two related questions?

Multivariate distributions

Probability of one is influenced by the other



Joint PMF (discrete)

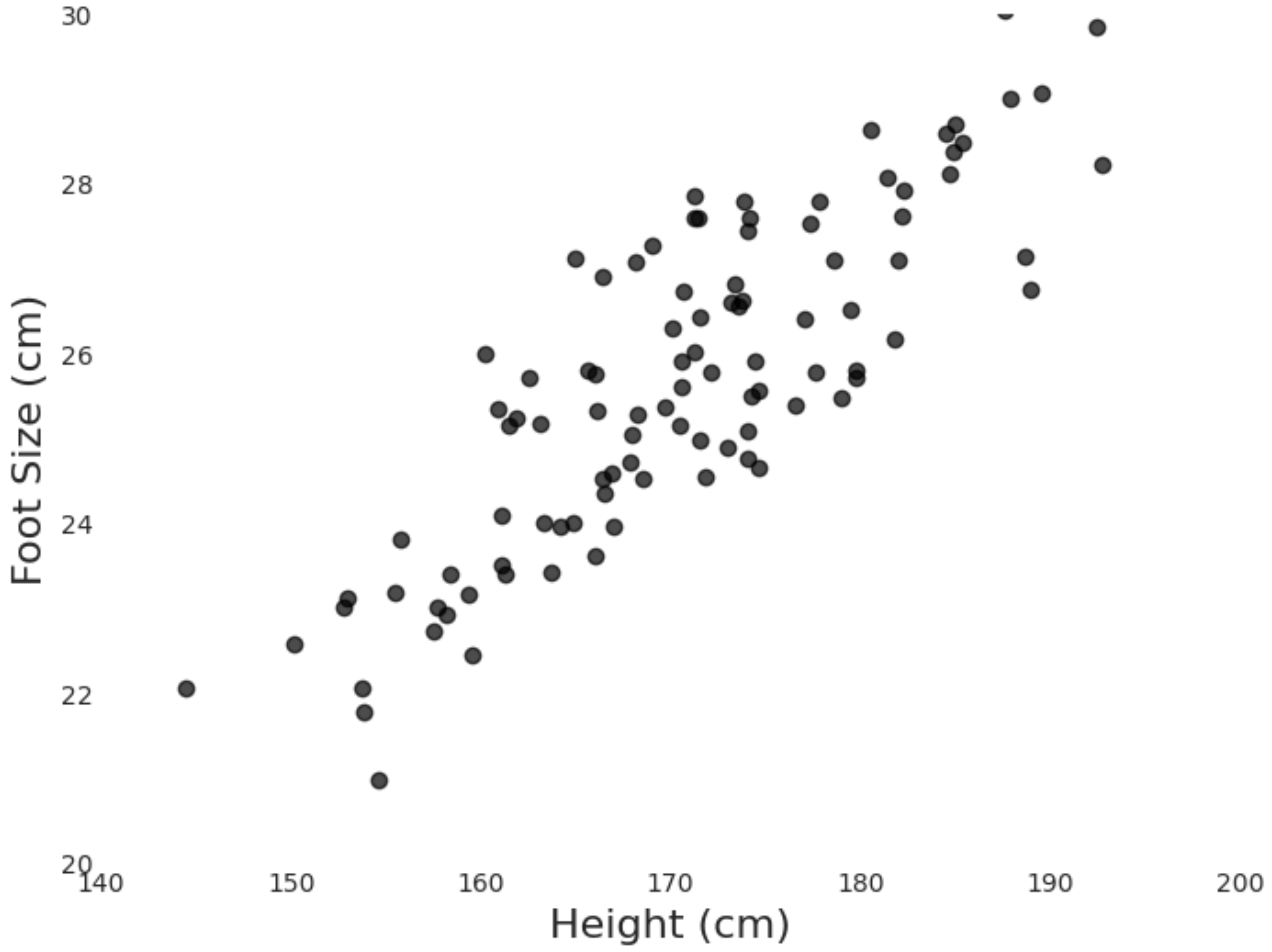
$$\mathbb{P}[X = x, Y = y] = f_{XY}(x, y)$$

Joint PDF (continuous)

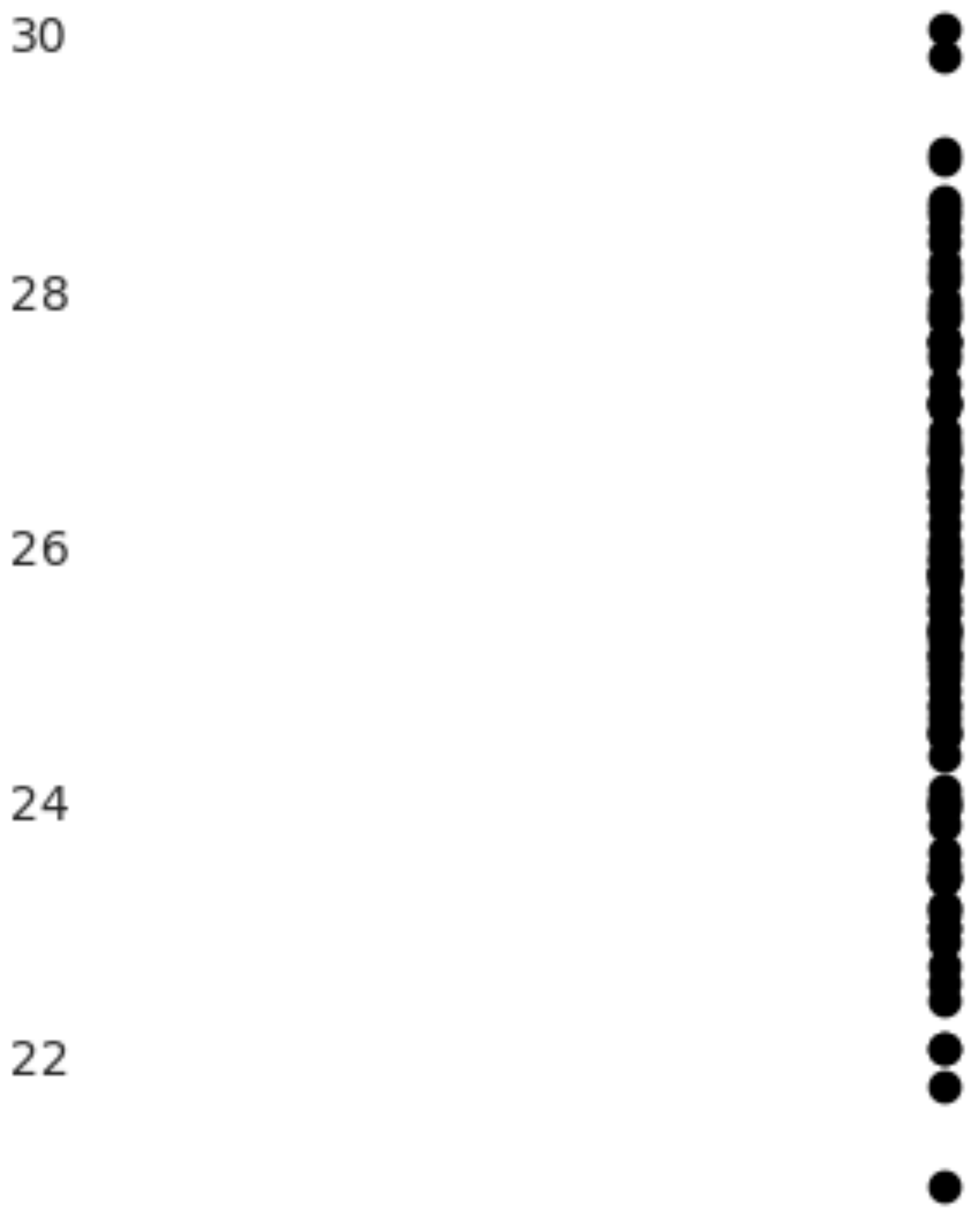
$$\mathbb{P}\Big[X \in [a, b], Y \in [c, d]\Big] = \int_a^b \int_c^d f_{XY}(x, y)\ dx\ dy$$

Marginalising

Scatter of X and Y



Squeeze X axis to scatter of Y



Joint pdf: $f_{XY}(x, y)$

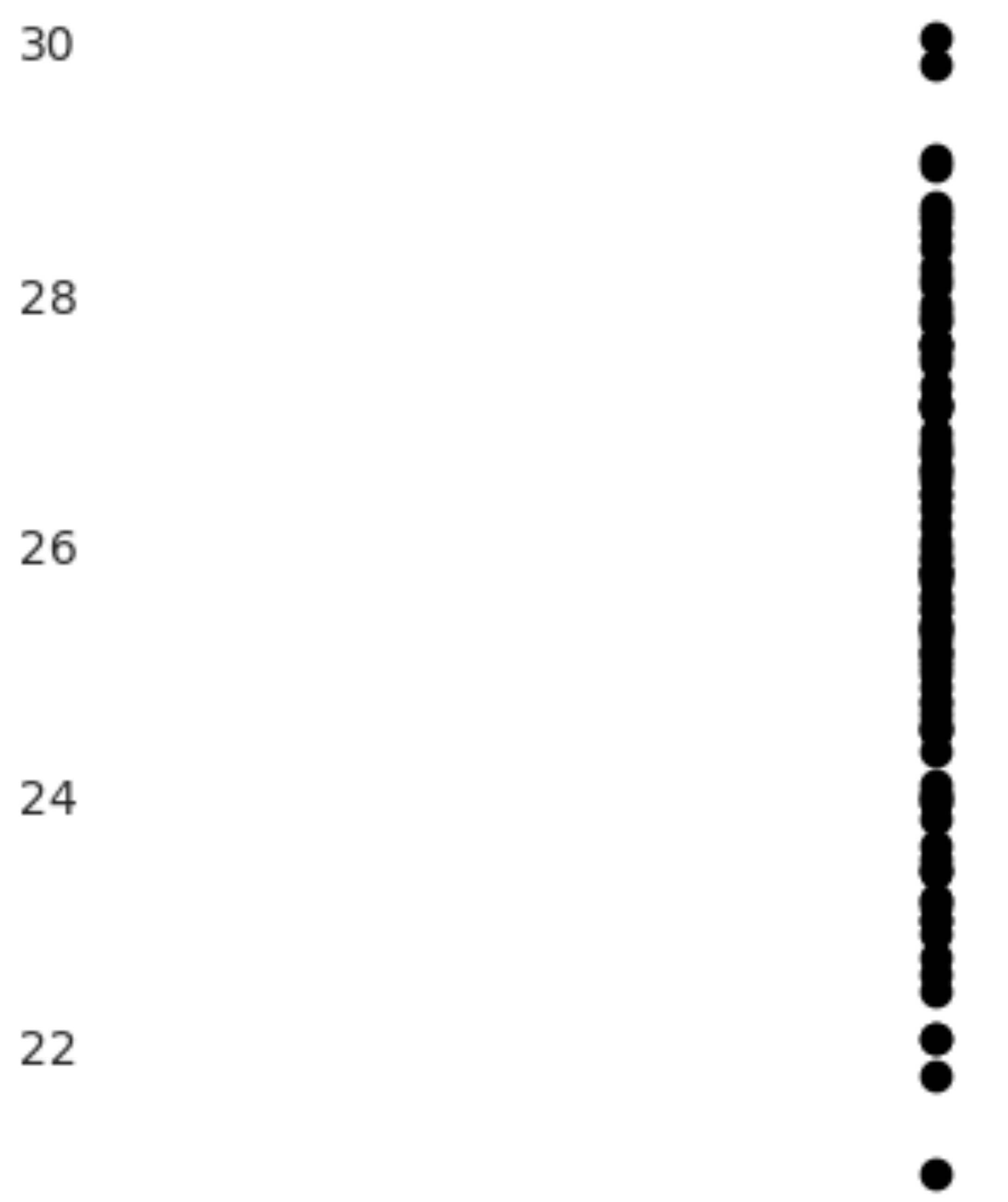
Marginalised pdf: $f_Y(y)$

Marginalising

PDF if we didn't have any information on X (height)

(*Squeeze Y axis to marginalise the other way*)

Squeeze X axis to scatter of Y



Marginalised pdf: $f_Y(y)$

Covariance of variables X and Y

Meaning:

Quantitative measure of how much they influence each other

$$\operatorname{cov}(X, Y)$$

Positive covariance

"Knowing that X is big increases how big we expect Y to be"

Negative covariance

"Knowing that X is big decreases how big we expect Y to be"

Zero covariance

"Knowing that X is big doesn't affect how big we expect Y to be"

Covariance of variables X and Y

Meaning:

Quantitative measure of how much they influence each other

Formula:

$$\boxed{\mathbb{E}\left[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\right]}$$

Or equivalently:

$$\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

…why?

Understanding covariance formula

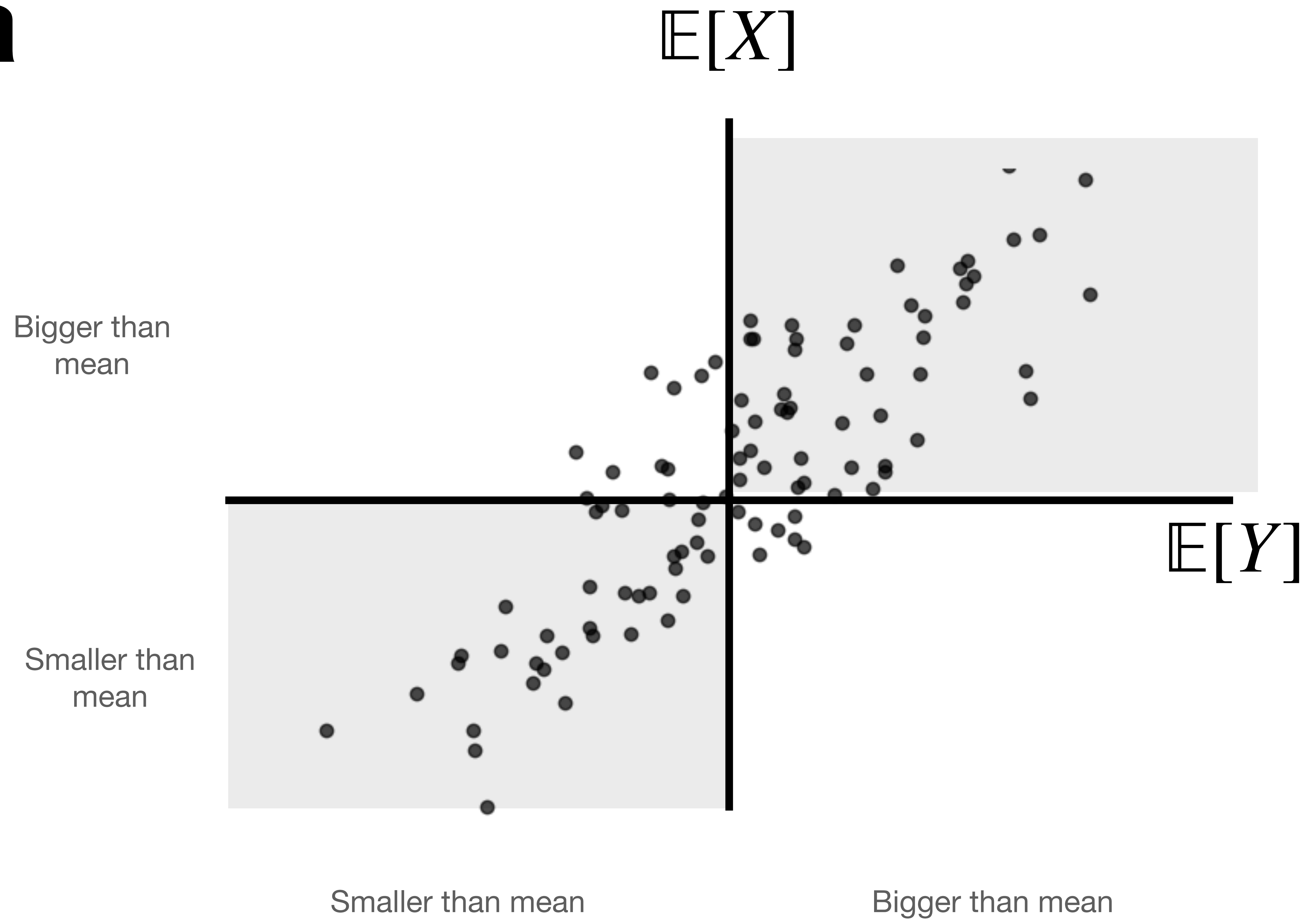
$$\mathbb{E}\left[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\right]$$

$(X - \mathbb{E}[X])(Y - \mathbb{E}[Y]) > 0$

when…

X and Y are bigger than their means:

X and Y are smaller than their means:



Understanding covariance formula

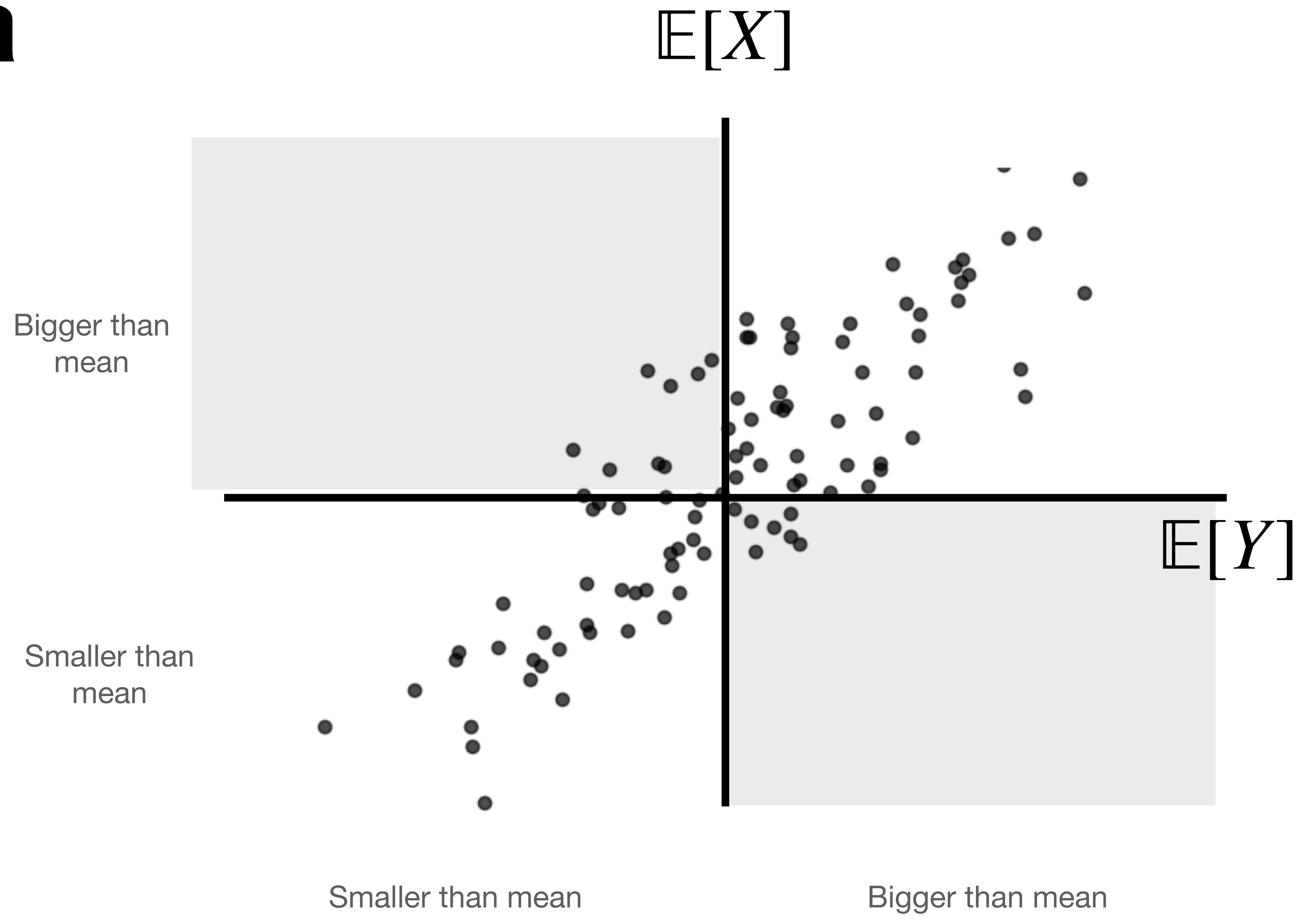
$$\mathbb{E}\left[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\right]$$

$$(X - \mathbb{E}[X])(Y - \mathbb{E}[Y]) < 0$$

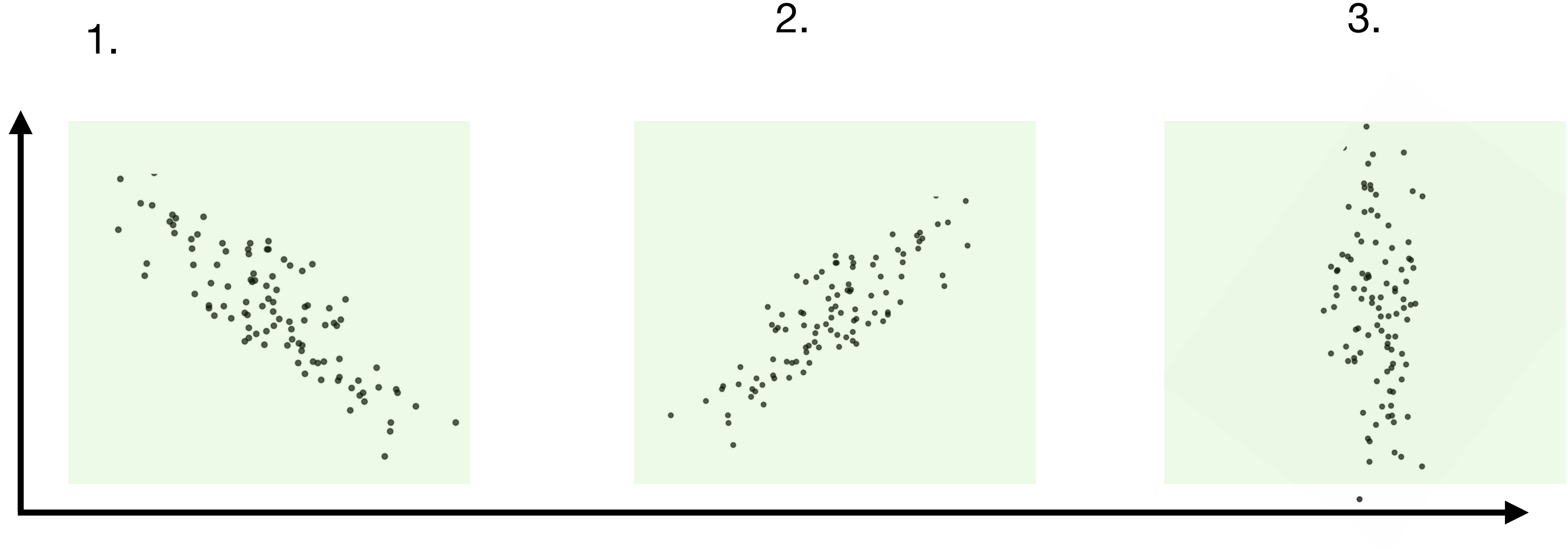
when…

One is bigger than mean

One is smaller than mean



What's the sign of the covariance?



Conditional probability

Partial knowledge of an outcome

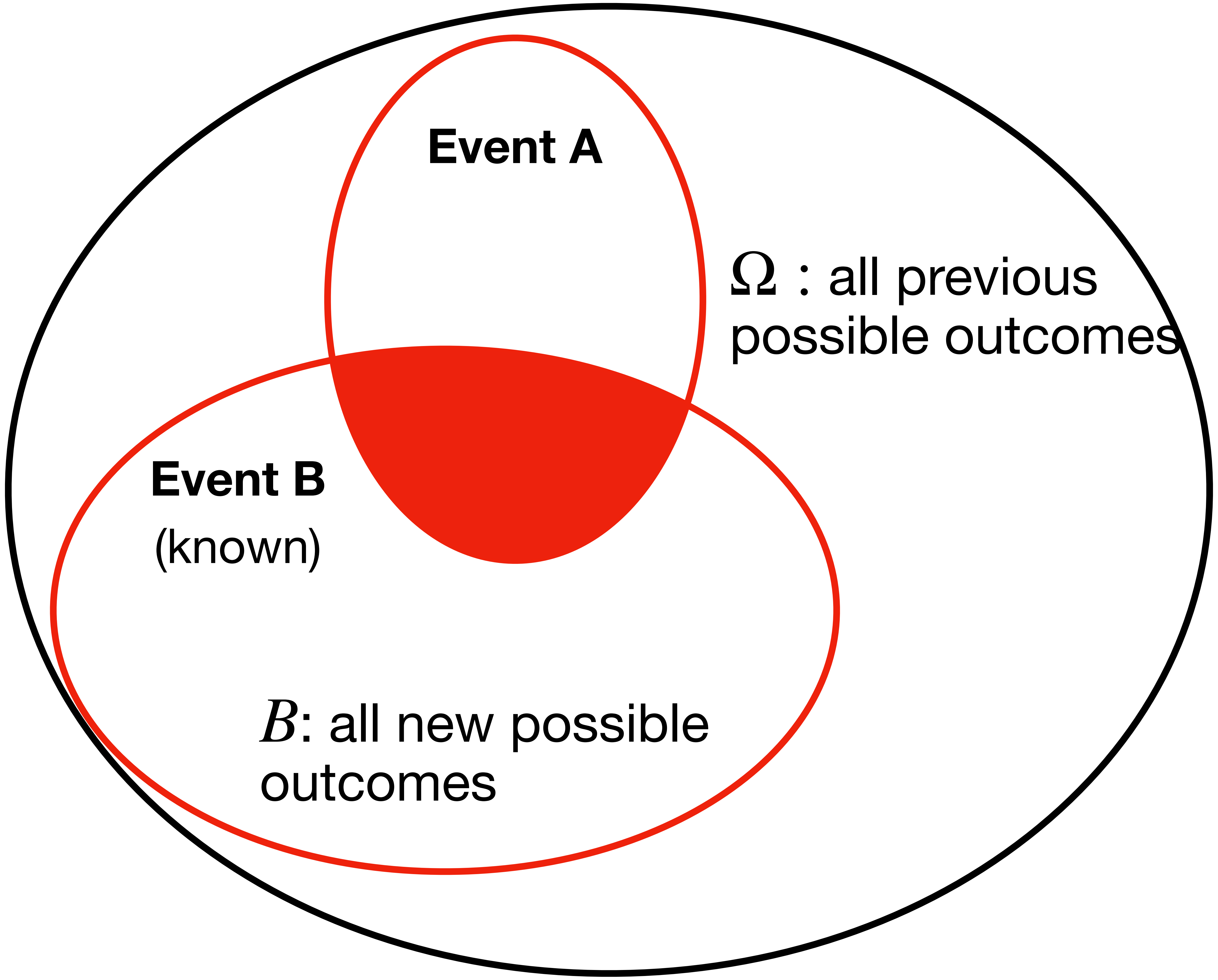
Other events are still uncertain, but their probability has changed

Y : first seat in back row filled?

Z : all seats unfilled?

(Draw on Venn diagram)

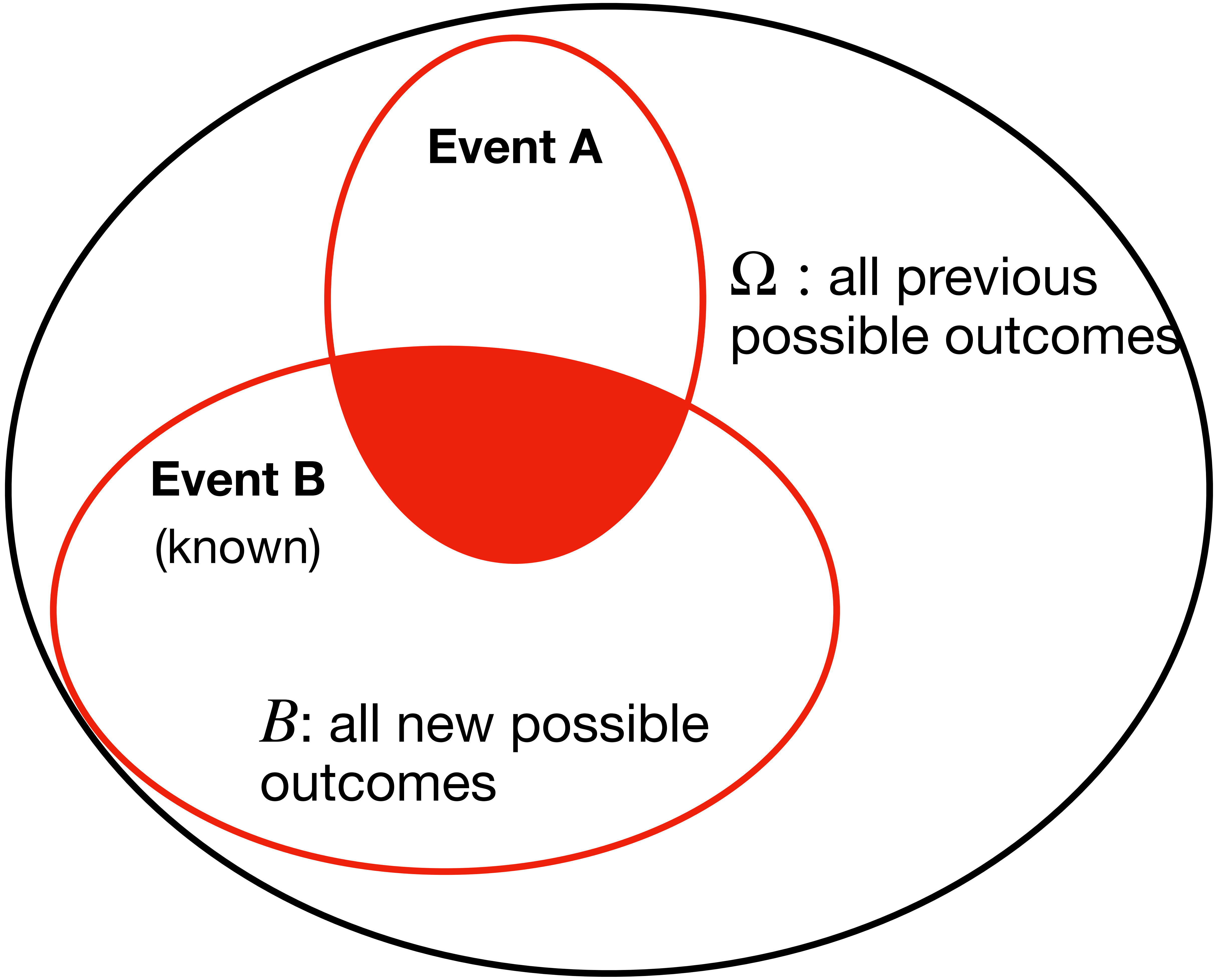
New Probability Space when Event B occurred



New set of outcomes in which event A happens: $A \cap B$

New set of all possible outcomes: B

New Probability Space when Event B occurred

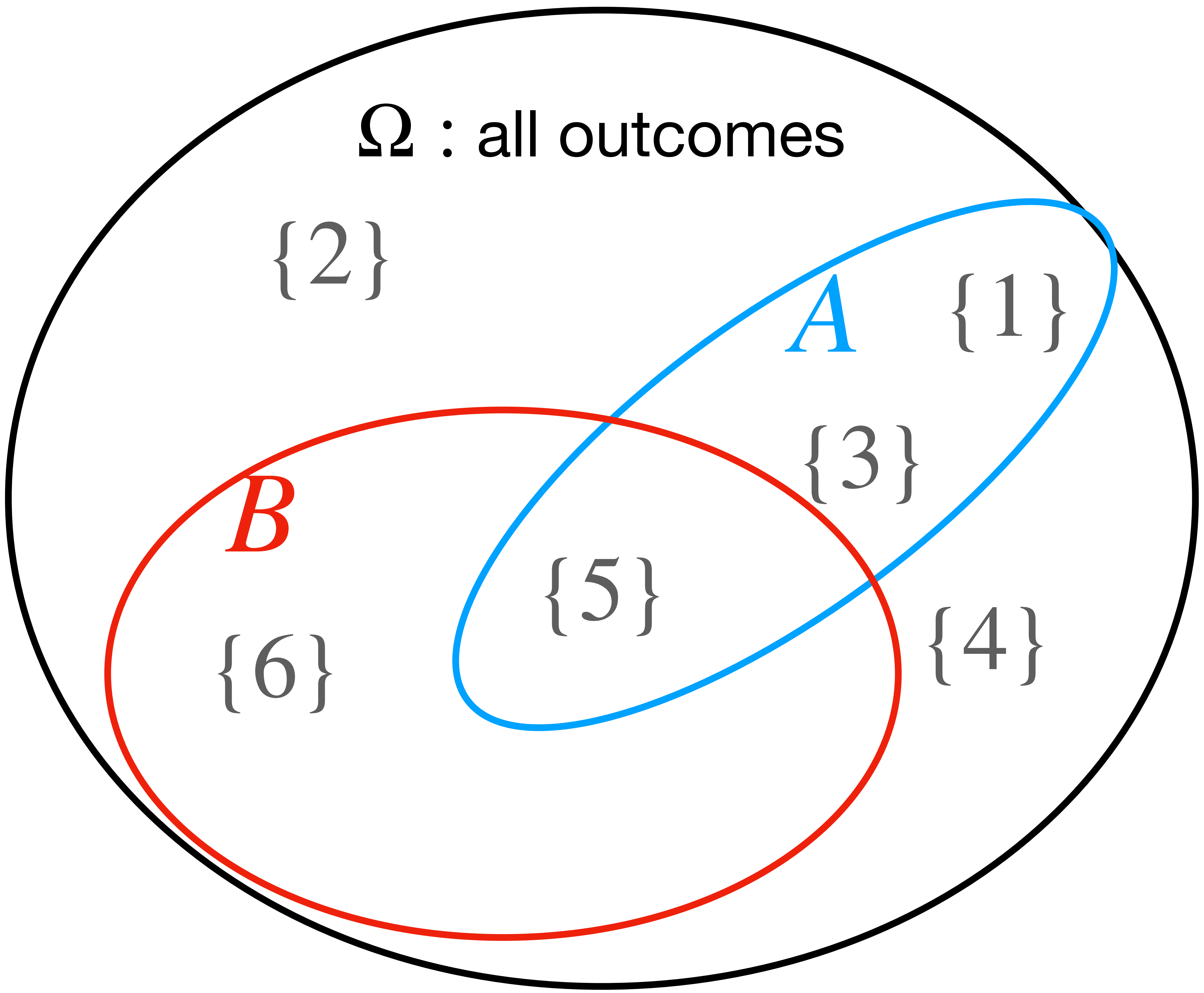


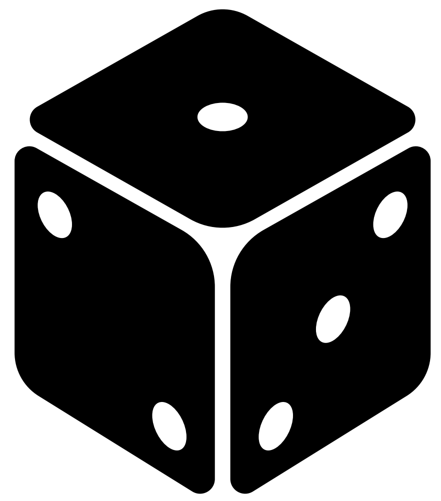
New set of outcomes in which event A happens: $A \cap B$

New set of all possible outcomes: B

$$\boxed{\mathbb{P}[A \mid B] = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}}$$

Partial knowledge of an outcome





$$\boxed{\mathbb{P}[A \mid B] = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}}$$

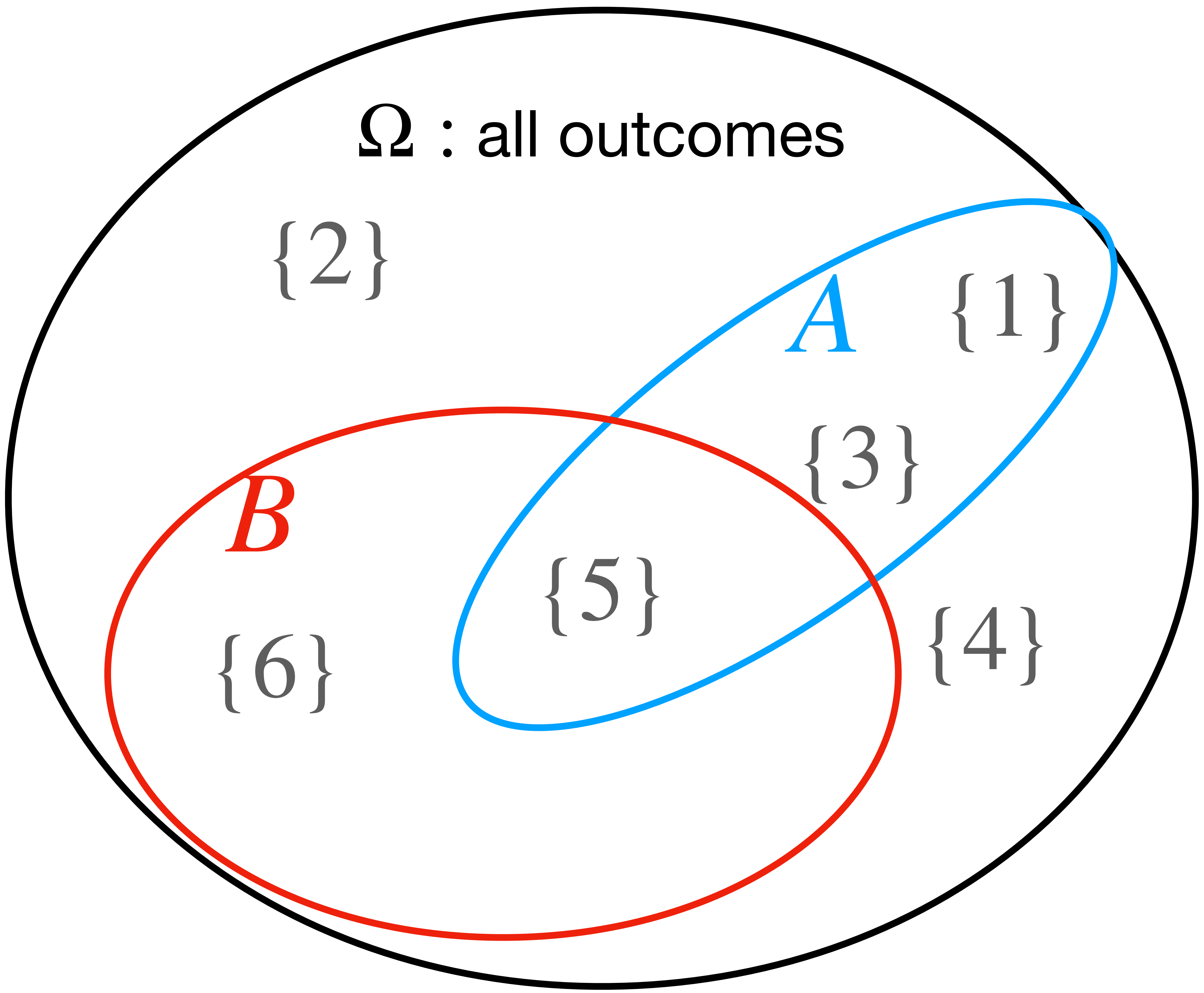
Law of conditional probability

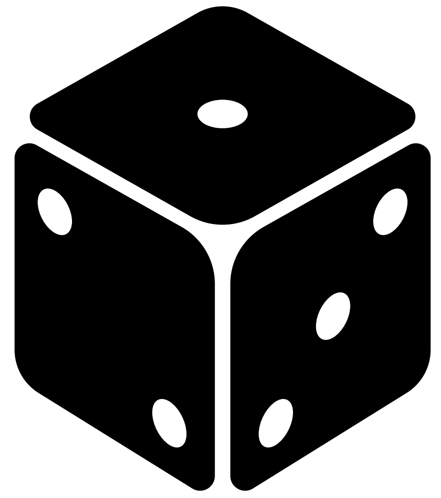
$$\boxed{\mathbb{P}[A \mid B] = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}}$$

Independent events

$$\boxed{\begin{aligned} &\mathbb{P}[A \mid B] = \mathbb{P}[A] \\ \Rightarrow\ &\mathbb{P}[A \cap B] = \mathbb{P}[A]\mathbb{P}[B] \end{aligned}}$$

Independent events?





$$\frac{1}{6} = \frac{1}{3} \times \frac{1}{2} \quad ✔$$

$$\boxed{\begin{aligned} &\mathbb{P}[A \mid B] = \mathbb{P}[A] \\ \Rightarrow\ &\mathbb{P}[A \cap B] = \mathbb{P}[A]\mathbb{P}[B] \end{aligned}}$$

Independent events?

Can exclusive events ever be independent?

Independent

$$\mathbb{P}[A \mid B] = \mathbb{P}[A]$$

$$\mathbb{P}[B \mid A] = \mathbb{P}[B]$$

Exclusive

$$\mathbb{P}[A \mid B] = 0$$

$$\mathbb{P}[B \mid A] = 0$$

Exclusive events:

$$\boxed{\mathbb{P}[A \mid B] = \frac{\overbrace{\mathbb{P}(A \cap B)}^{0}}{\mathbb{P}(B)}}$$

…only if one of their initial probabilities are zero

Bayes' Theorem

$$\boxed{\mathbb{P}[A \mid B] = \frac{\mathbb{P}(B \mid A)\mathbb{P}[A]}{\mathbb{P}(B)}}$$

Aim is to understand and apply

Bayes' Theorem is useful

$$\boxed{\mathbb{P}[A \mid B] = \frac{\mathbb{P}(B \mid A)\mathbb{P}[A]}{\mathbb{P}(B)}}$$

A B

Probability of COVID | sore throat?

$$= \frac{0.3 \times 0.01}{0.05}$$

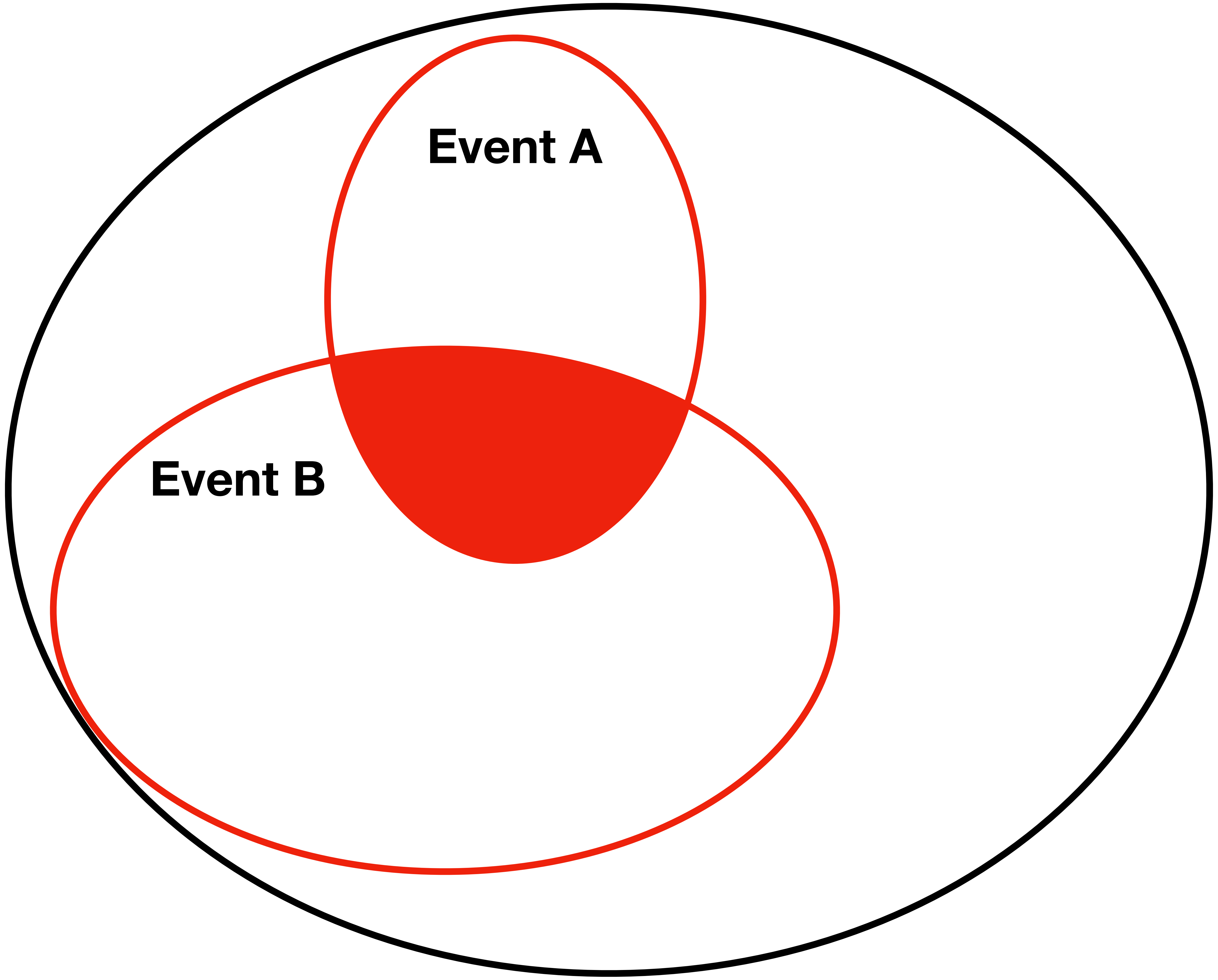
$$= 6\,\%$$

Probability of COVID = 1% $\mathbb{P}[A]$

Probability of sore throat = 5% $\mathbb{P}[B]$

Percentage of Covid patients with sore throat = 30% $\mathbb{P}[B \mid A]$

Understanding Bayes' Theorem

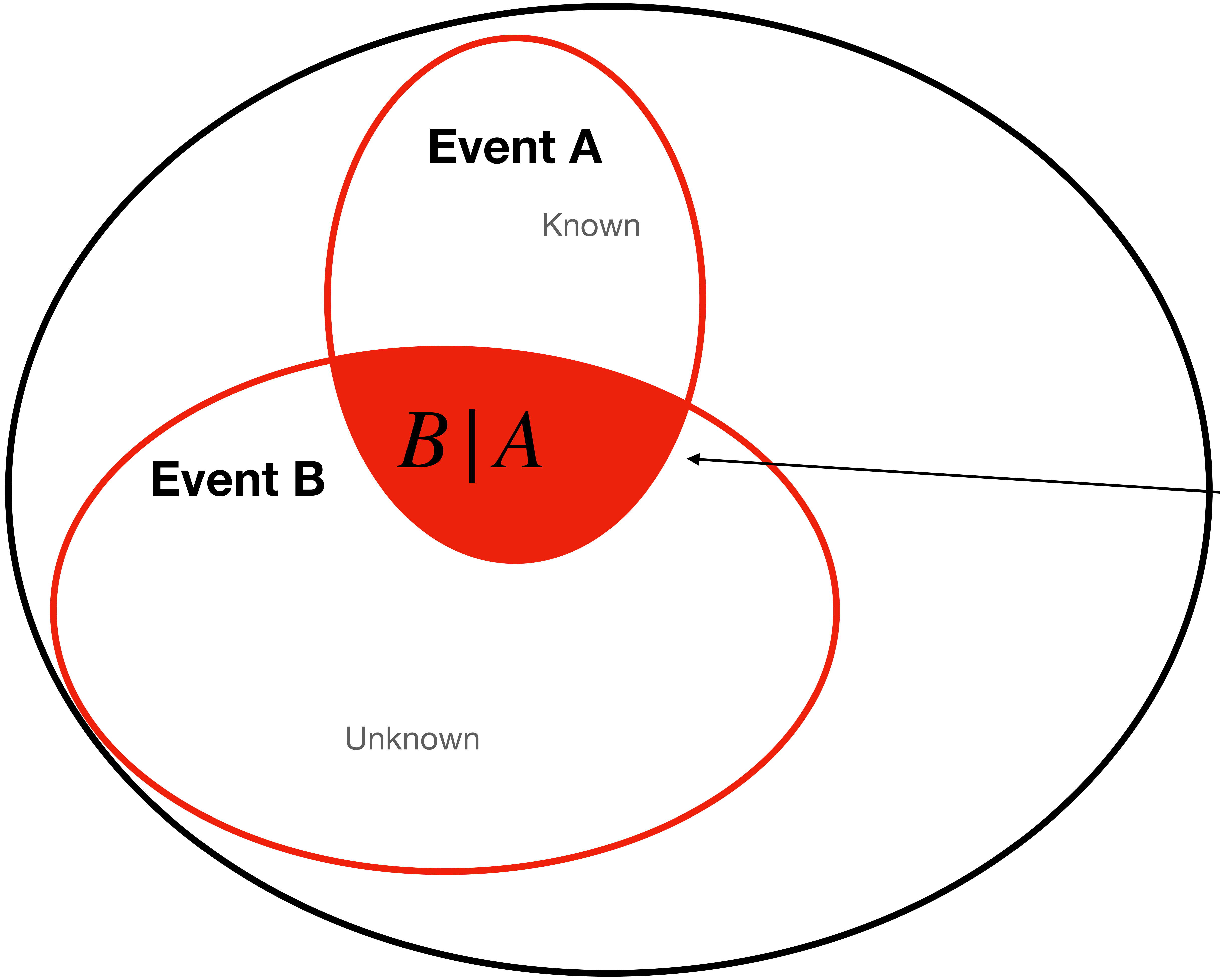


$$\boxed{\mathbb{P}[A \mid B] = \frac{\mathbb{P}(B \mid A)\mathbb{P}[A]}{\mathbb{P}(B)}}$$

Going to think about the intersection in two different ways…

Bayes' Theorem

$B \mid A:$



$$\boxed{\mathbb{P}[A \mid B] = \frac{\mathbb{P}(B \mid A)\mathbb{P}[A]}{\mathbb{P}(B)}}$$

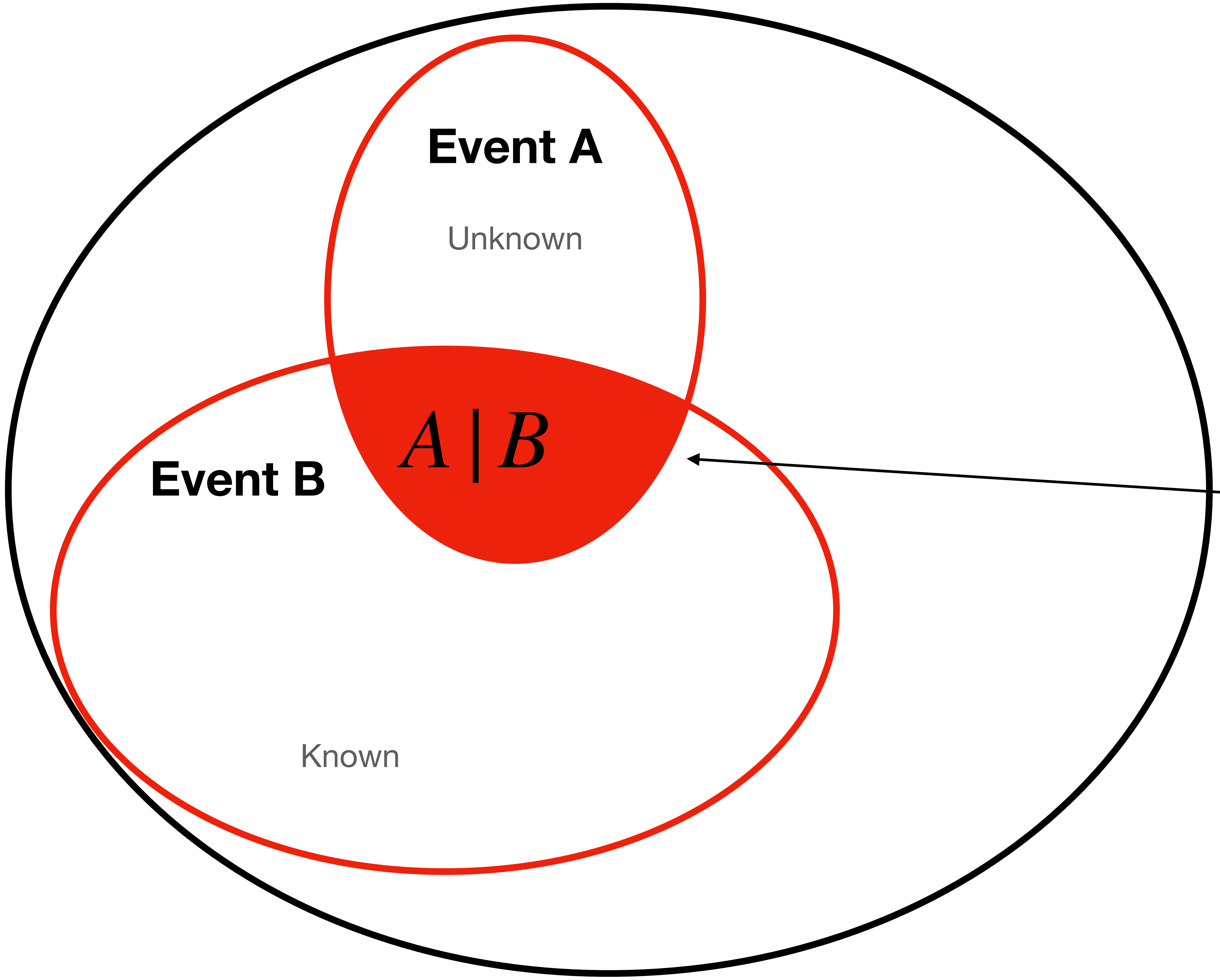
$$\mathbb{P}[B \mid A] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[A]}$$

$$\Downarrow$$

$$\mathbb{P}[A \cap B] = \mathbb{P}[B \mid A]\mathbb{P}[A]$$

Bayes' Theorem

$$A \mid B:$$



$$\boxed{\mathbb{P}[A \mid B] = \frac{\mathbb{P}(B \mid A)\mathbb{P}[A]}{\mathbb{P}(B)}}$$

$$\mathbb{P}[A \mid B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}$$

$$\Downarrow$$

$$\mathbb{P}[A \cap B] = \mathbb{P}[A \mid B]\mathbb{P}[B]$$

Altogether…

$$\boxed{\mathbb{P}[A \mid B] = \frac{\mathbb{P}(B \mid A)\mathbb{P}[A]}{\mathbb{P}(B)}}$$

$A \mid B :$

$$\mathbb{P}[A \cap B] = \mathbb{P}[A \mid B]\mathbb{P}[B]$$

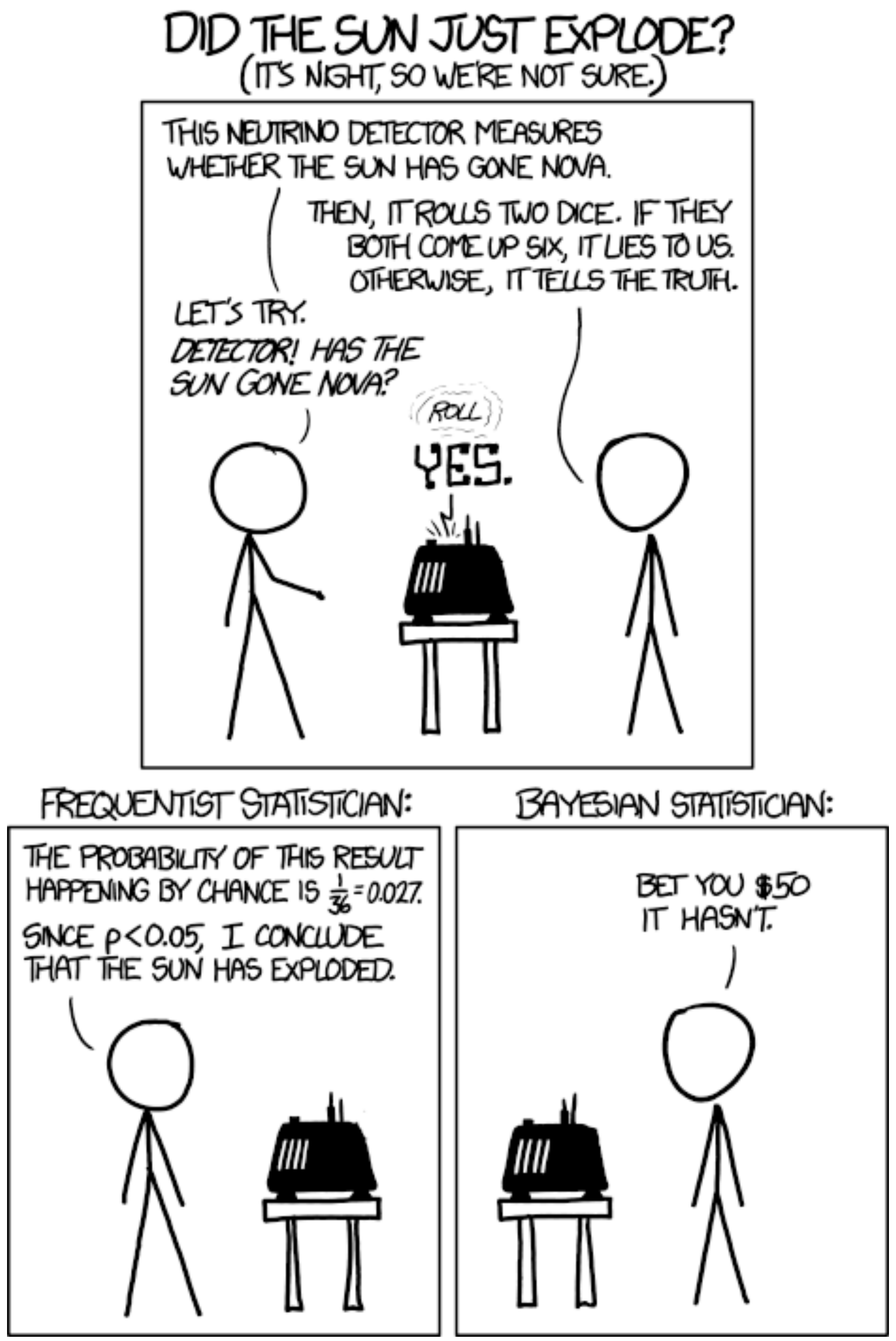
$B \mid A :$

$$\mathbb{P}[A \cap B] = \mathbb{P}[B \mid A]\mathbb{P}[A]$$

$$\Rightarrow$$

$$\mathbb{P}[A \mid B]\mathbb{P}[B] = \mathbb{P}[B \mid A]\mathbb{P}[A]$$

Homework: tell me why this joke is funny



$$\boxed{\mathbb{P}[A \mid B] = \frac{\mathbb{P}(B \mid A)}{\mathbb{P}(B)}}$$