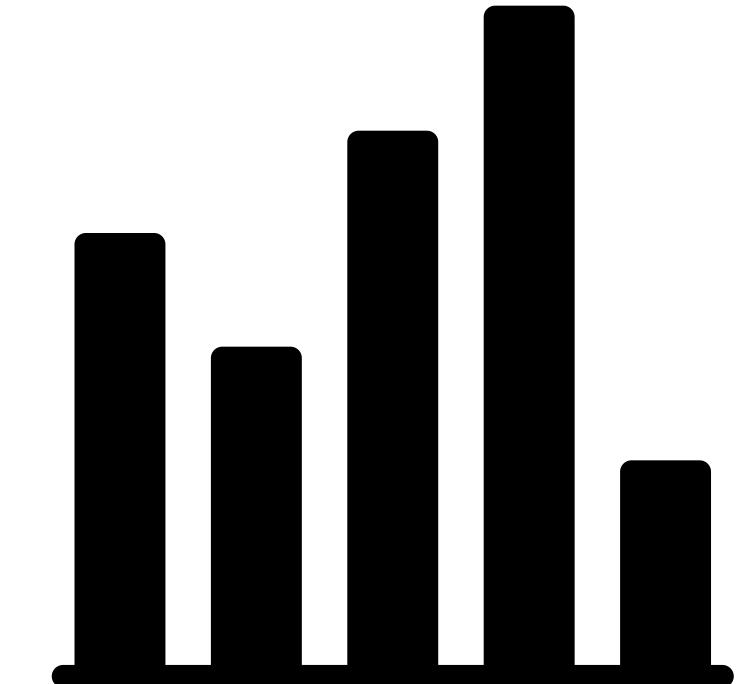
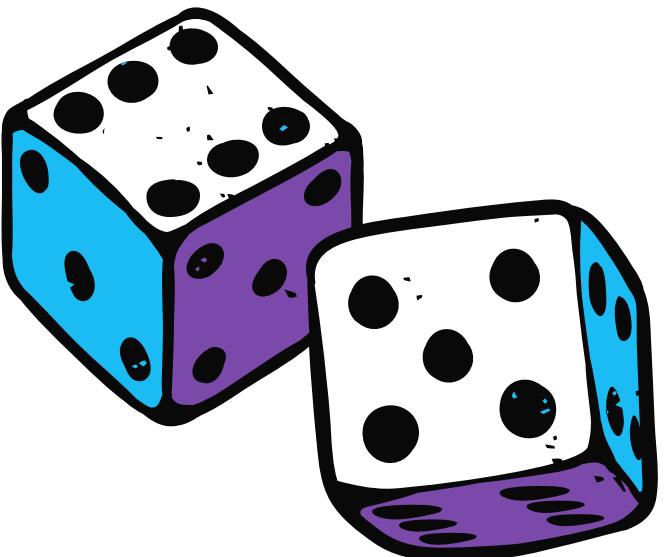


Bayesian probability theory as a logic of science

Gui Araujo



Coin flip: what is probability?



Coin flip: what is the meaning of probability?



- Frequency of outcomes
- Natural propensity of the coin
- A degree of uncertainty

Coin flip: what is the meaning probability?



- A degree of uncertainty
- Randomness in nature

Coin flip: what is the meaning probability?



- A degree of uncertainty
- Randomness in nature
- Initial state

Coin flip: what is the meaning probability?



- A degree of uncertainty
- Randomness in nature
- Initial state
- Model's perspective

Coin flip: what is the meaning probability?



- A degree of uncertainty
- Randomness in nature
- Initial state
- Model's perspective
- Evaluation of statements

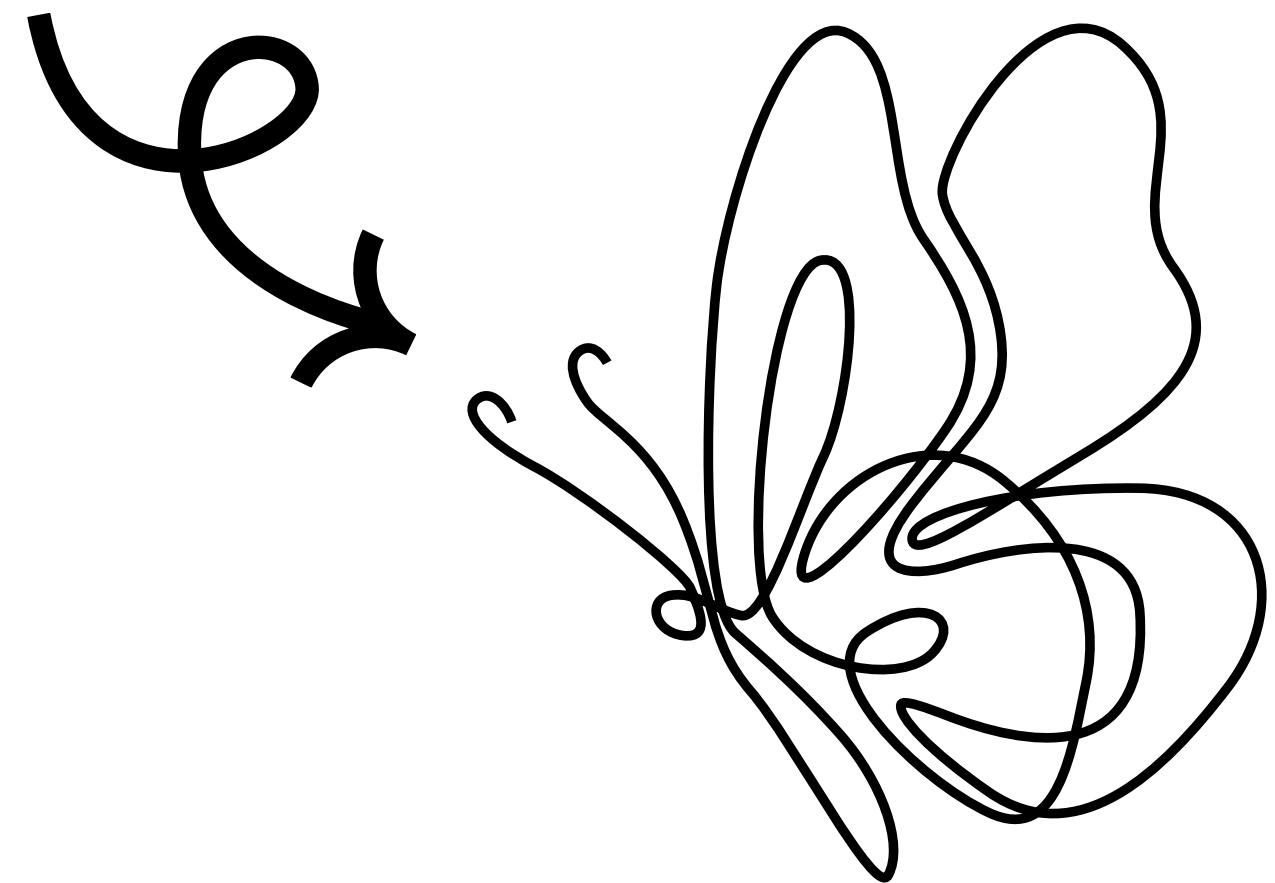
Coin flip: what is the meaning probability?



- A degree of uncertainty
- Randomness in nature
- Initial state
- Model's perspective
- Evaluation of statements
- Information availability

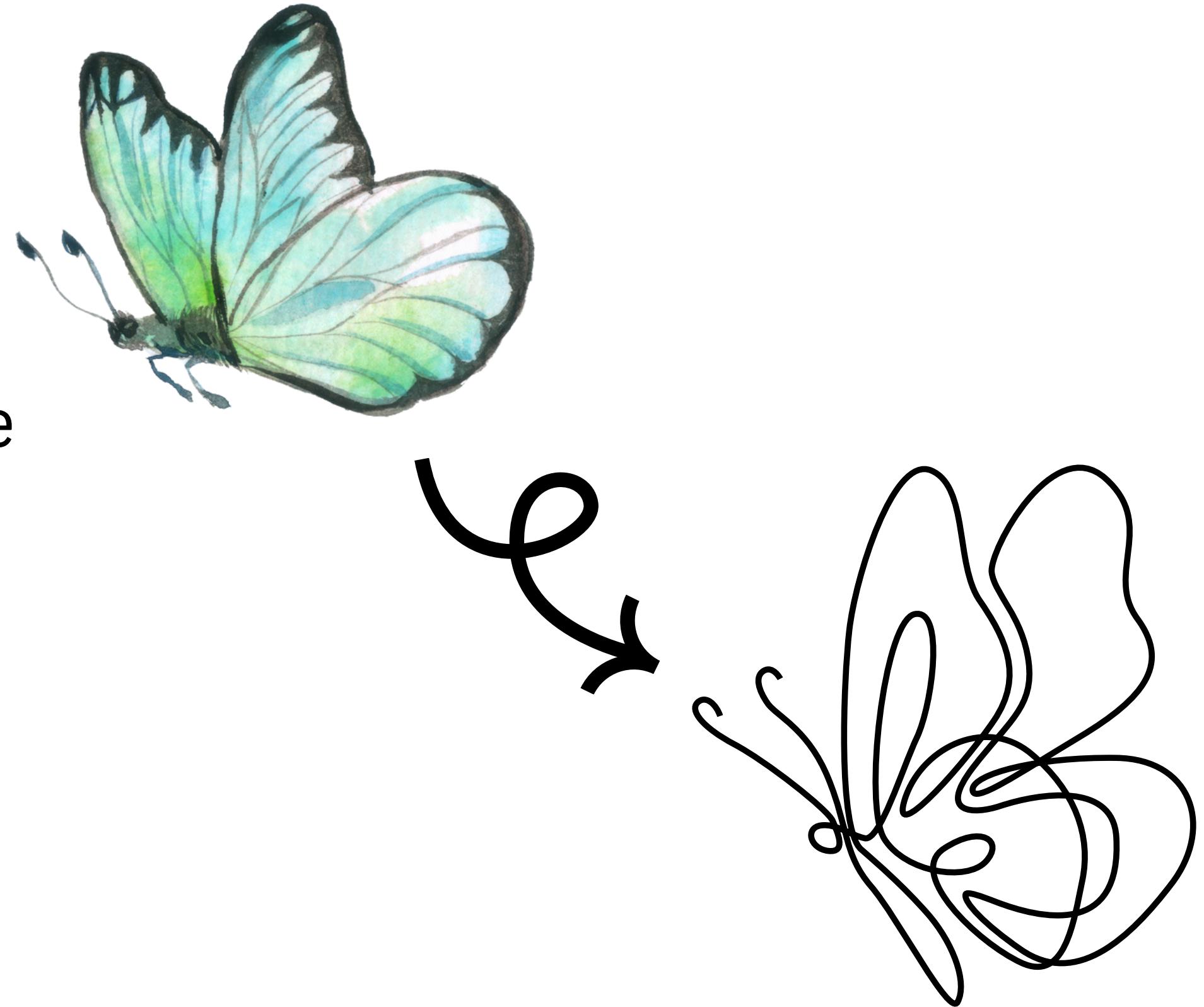
But what is a model?

- To behave the same, to imitate



But what is a model?

- To behave the same, to imitate
- Theory: ‘meta-model’



But what is a model?

- To behave the same, to imitate
- Theory: ‘meta-model’
- Generative space for statements (even descriptions, explanations)



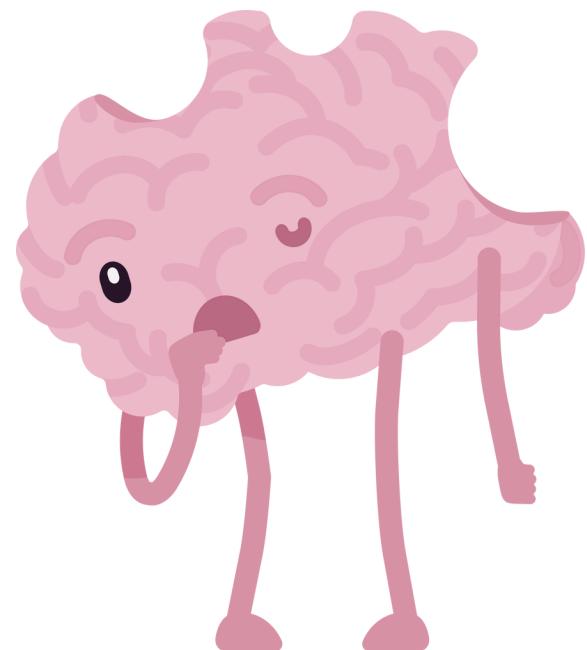
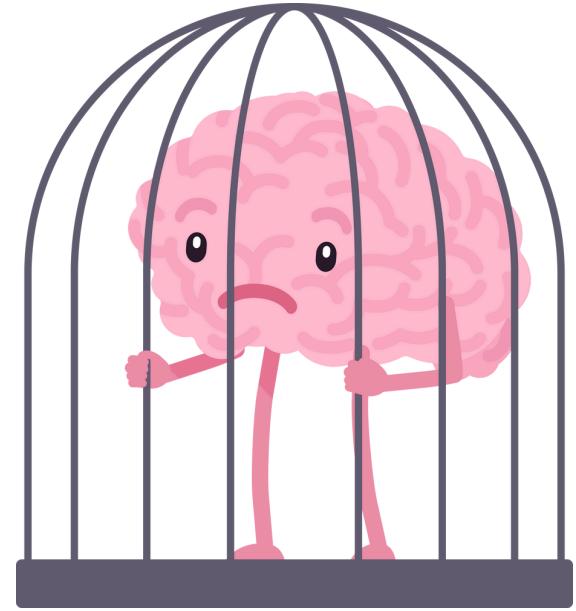
But what is a model?

- To behave the same, to imitate
- Theory: ‘meta-model’
- Generative space for statements (even descriptions, explanations)
- In this sense, mathematics is a language, one of **abstraction**



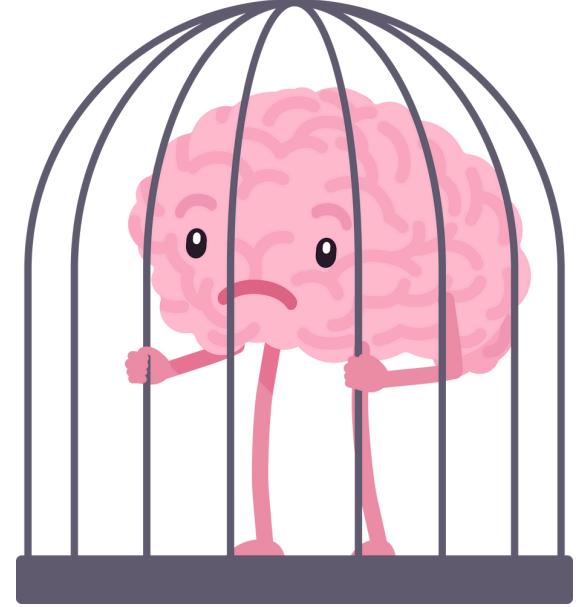
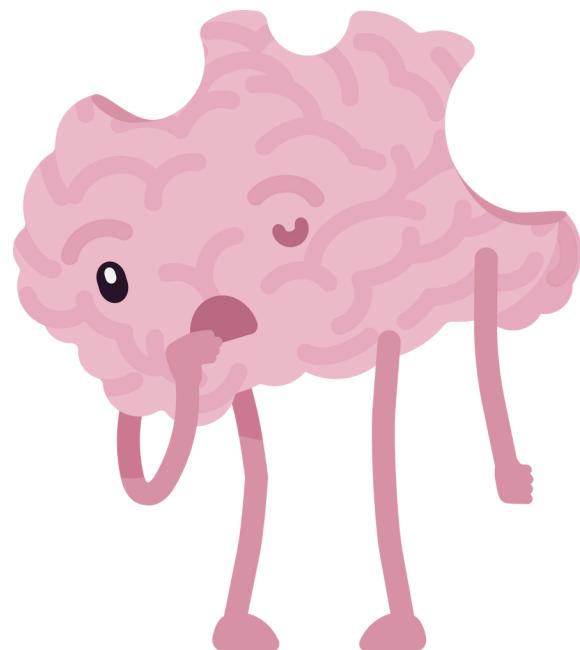
Uncertainty is inescapable

- Our knowledge about the world is always limited, noisy, and interpreted



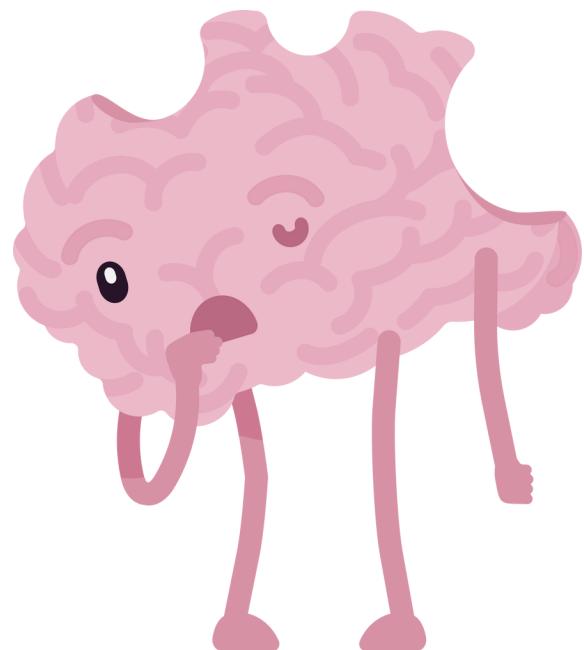
Uncertainty is inescapable

- Our knowledge about the world is always limited, noisy, and interpreted
- Observation of ‘truth’ (real world) is possible only within a model (truth is not independent)

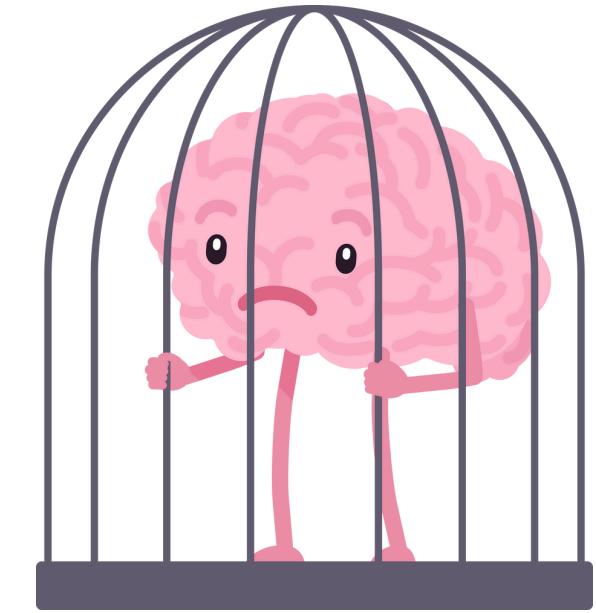


Uncertainty is inescapable

- Our knowledge about the world is always limited, noisy, and interpreted
- Observation of ‘truth’ (real world) is possible only within a model (truth is not independent)

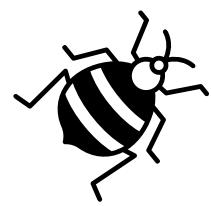
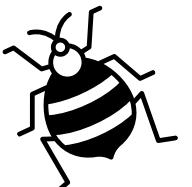
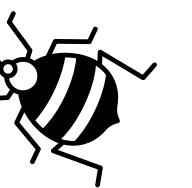
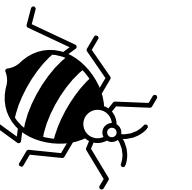


- Representation: Brains, computers, etc, can render/run models as physical processes



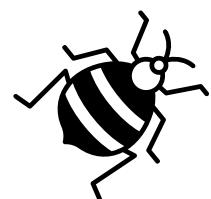
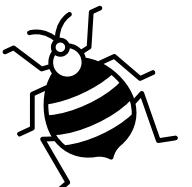
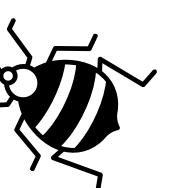
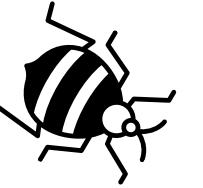
Data collection

- The process of experimentation is necessarily a subjective artifact. Theories are even more so.



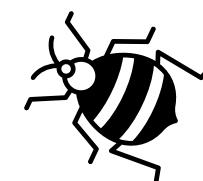
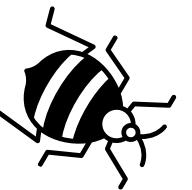
Data collection

- The process of experimentation is necessarily a subjective artifact. Theories are even more so.
- Why this specific approach? Why this question? Why this procedure? Why this variable? Why this selection of facts?



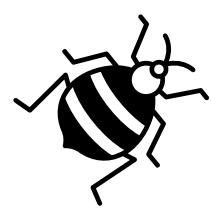
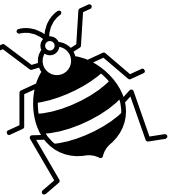
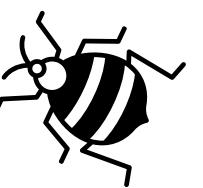
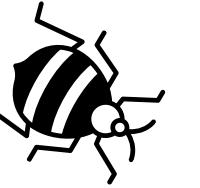
Data collection

- The process of experimentation is necessarily a subjective artifact. Theories are even more so.
- Why this specific approach? Why this question? Why this procedure? Why this variable? Why this selection of facts?
 - e.g.: I want to know how the number of aphids on a plant is related to the size of the plant.



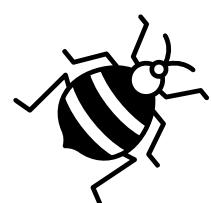
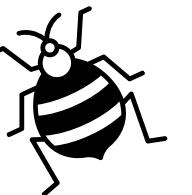
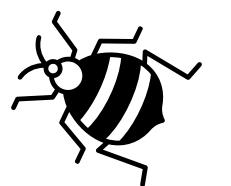
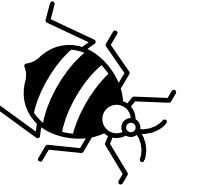
Counting Aphids

- I counted a total of aphids in each of the 5 petals = {5, 10, 2, 8, 5}

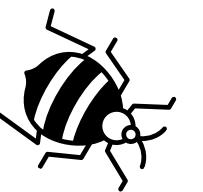


Counting Aphids

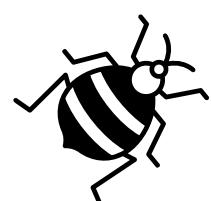
- I counted a total of aphids in each of the 5 petals = {5, 10, 2, 8, 5}
- Binary logic/Mathematics: “I counted a total of 30 aphids”. True or false?



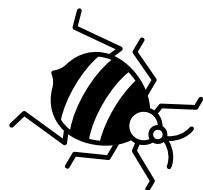
Counting Aphids



- I counted a total of aphids in each of the 5 petals = {5, 10, 2, 8, 5}
- Binary logic/Mathematics: “I counted a total of 30 aphids”. True or false?
- Fuzzy nature: “There are a total of 30 aphids on the plant”. True or false?



Counting Aphids

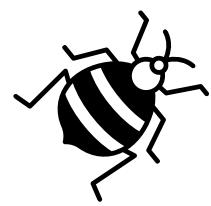
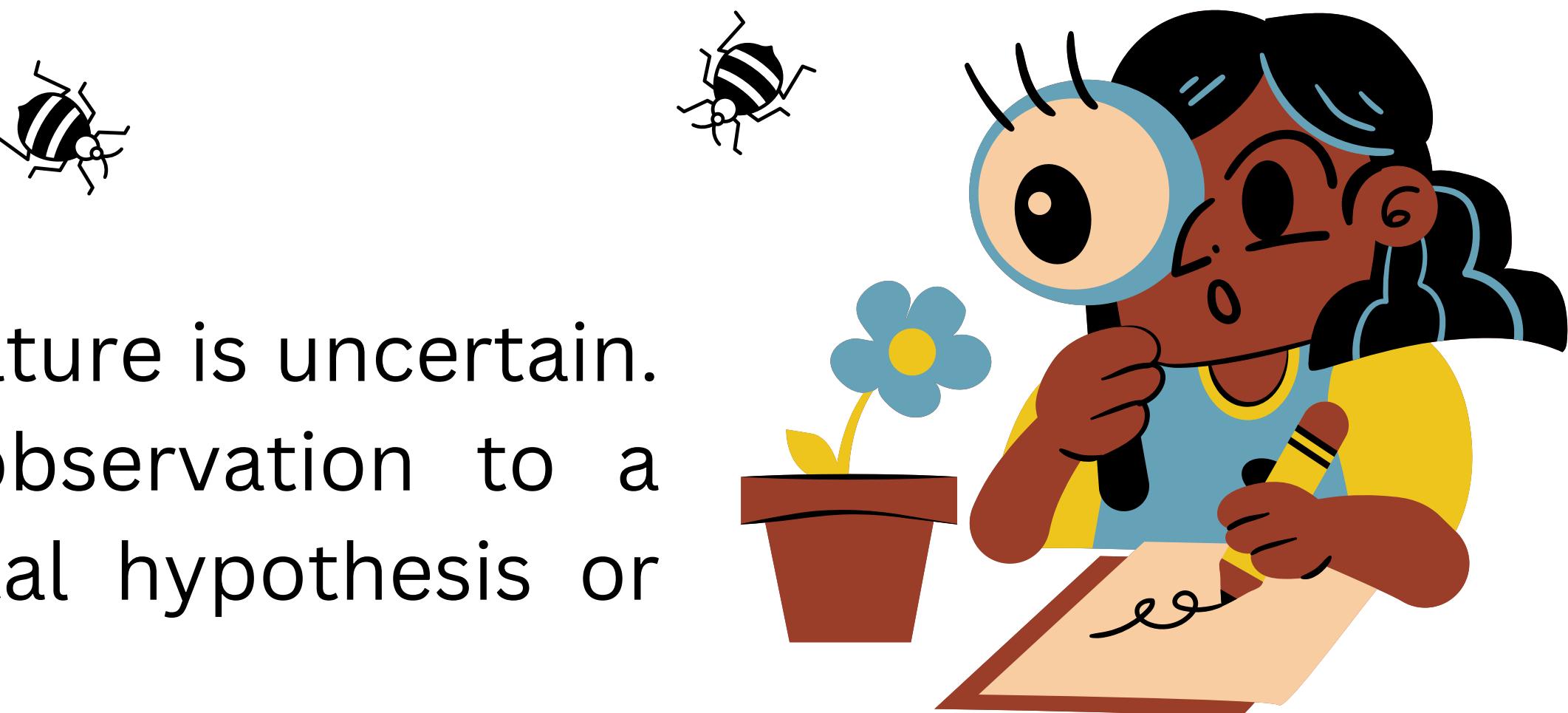


- I counted a total of aphids in each of the 5 petals = {5, 10, 2, 8, 5}
- Binary logic/Mathematics: “I counted a total of 30 aphids”. True or false?
- Fuzzy nature: “There are a total of 30 aphids on the plant”. True or false?
 - What if I made a counting mistake? Or what if some aphids were really hidden? How can I be certain?

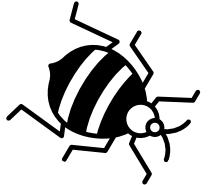


Plausibility

- Every statement about nature is uncertain.
From the most basic observation to a sophisticated experimental hypothesis or scientific theory



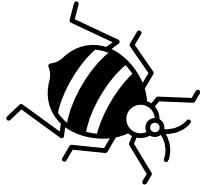
Plausibility



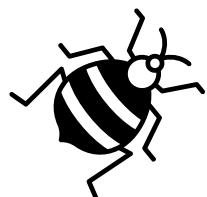
- Every statement about nature is uncertain. From the most basic observation to a sophisticated experimental hypothesis or scientific theory
- The true/false binary is not adequate for science. It must be extended to a continuum representing a degree of uncertainty



Plausibility



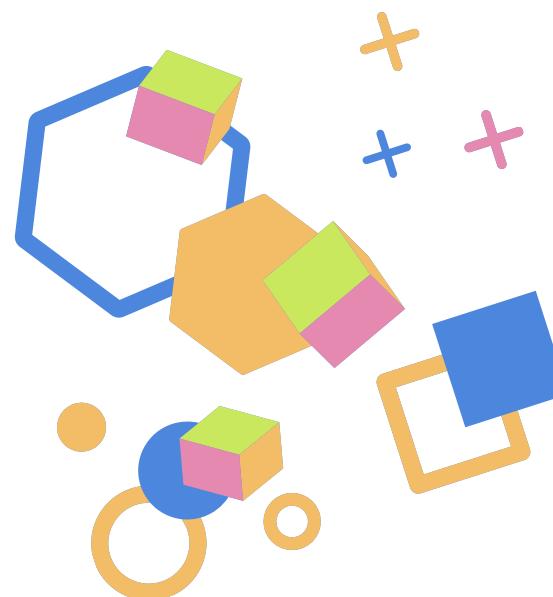
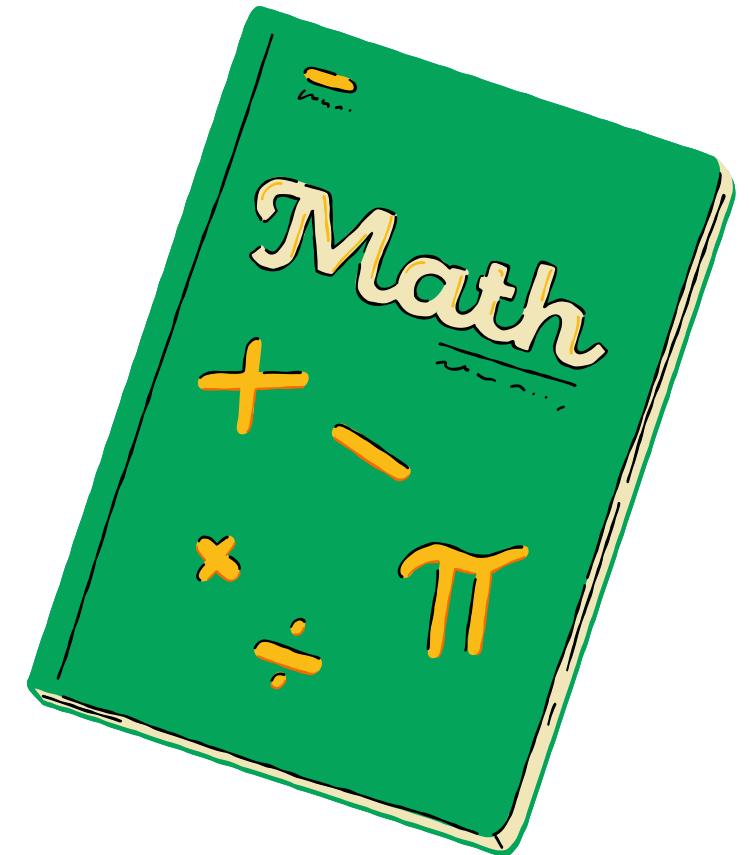
- Every statement about nature is uncertain. From the most basic observation to a sophisticated experimental hypothesis or scientific theory
- The true/false binary is not adequate for science. It must be extended to a continuum representing a degree of uncertainty
- We evaluate statements according to their plausibility!



Mathematical Probability

f(x)

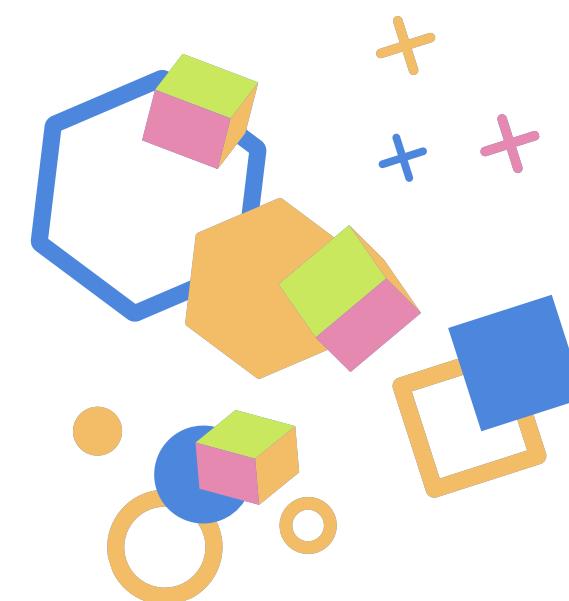
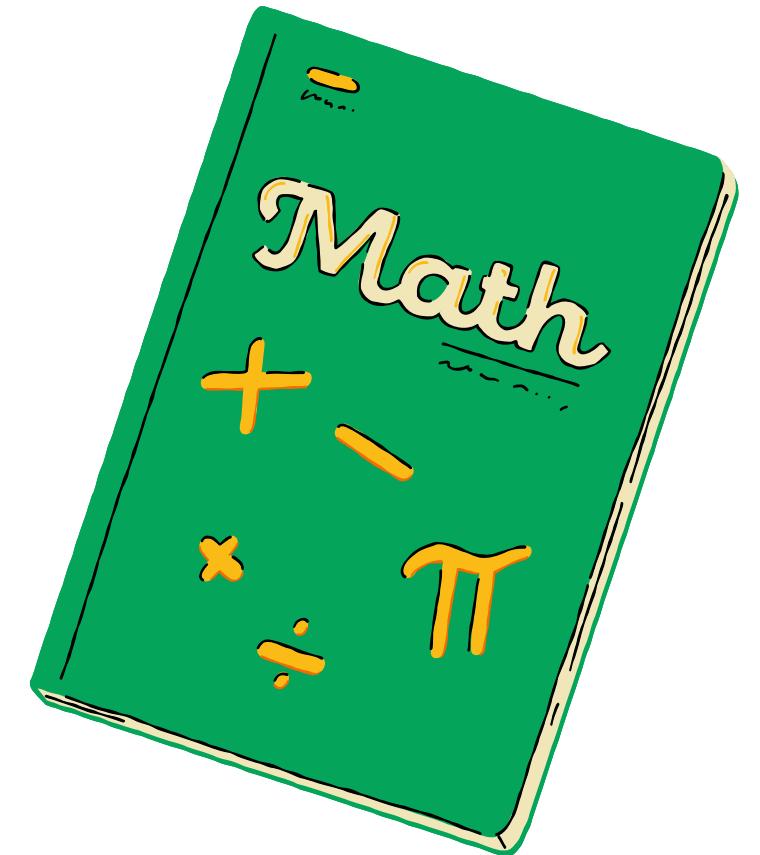
- It turns out that the technical problem of quantifying plausibility is the same as the task of measuring sizes of sets.



Mathematical Probability

f(x)

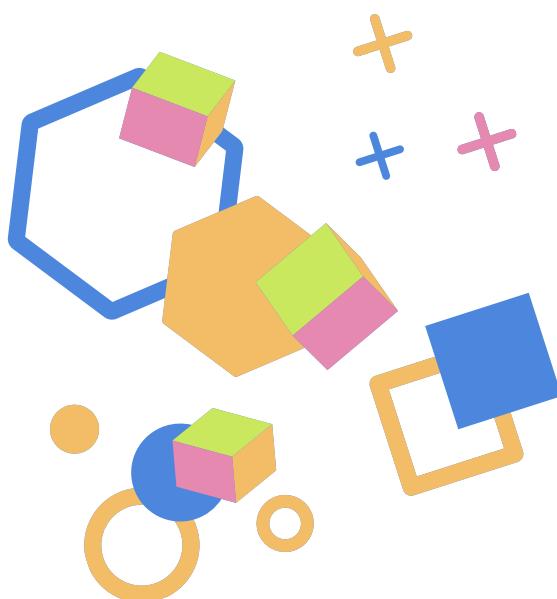
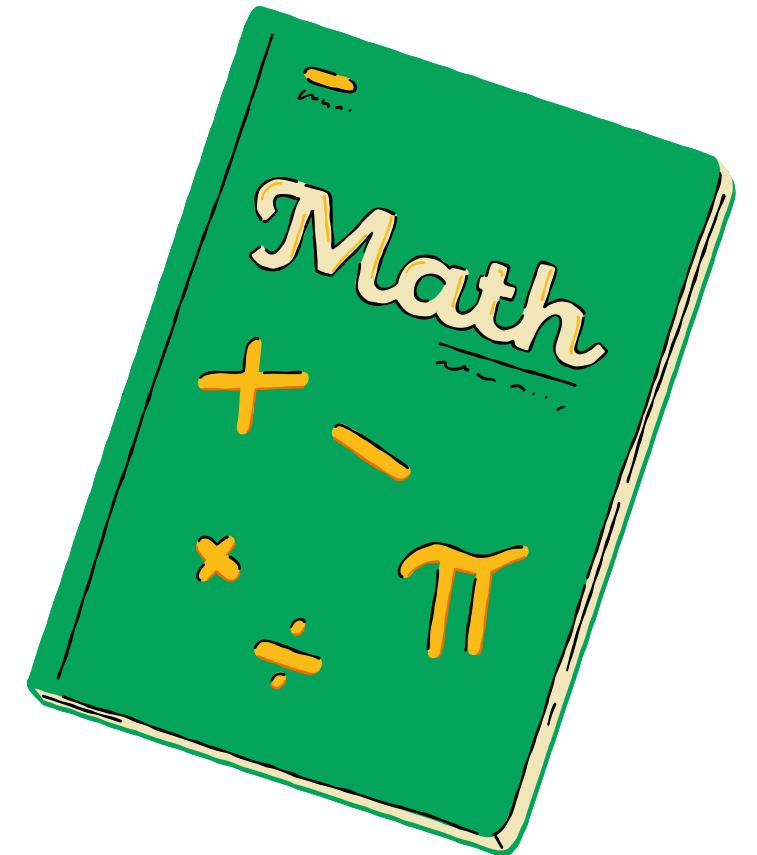
- It turns out that the technical problem of quantifying plausibility is the same as the task of measuring sizes of sets.
- The answer: the plausibility of a statement is quantified by a mathematical structure named probability space, which evaluates with a number in $[0,1]$ all the way between false and true.



A little about ‘the way of math’

f(x)

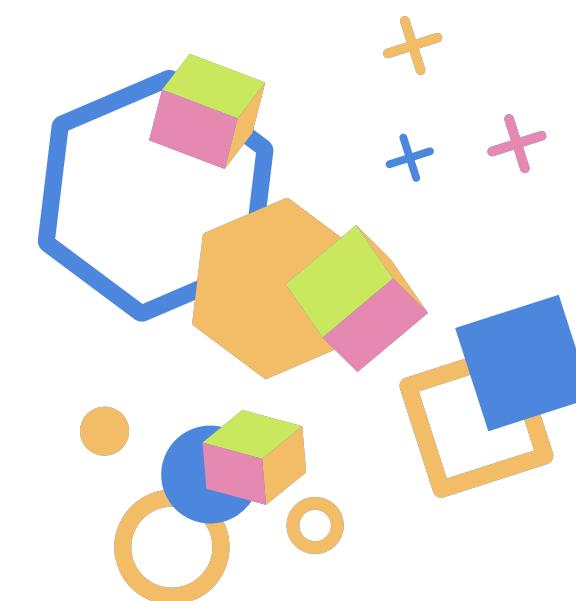
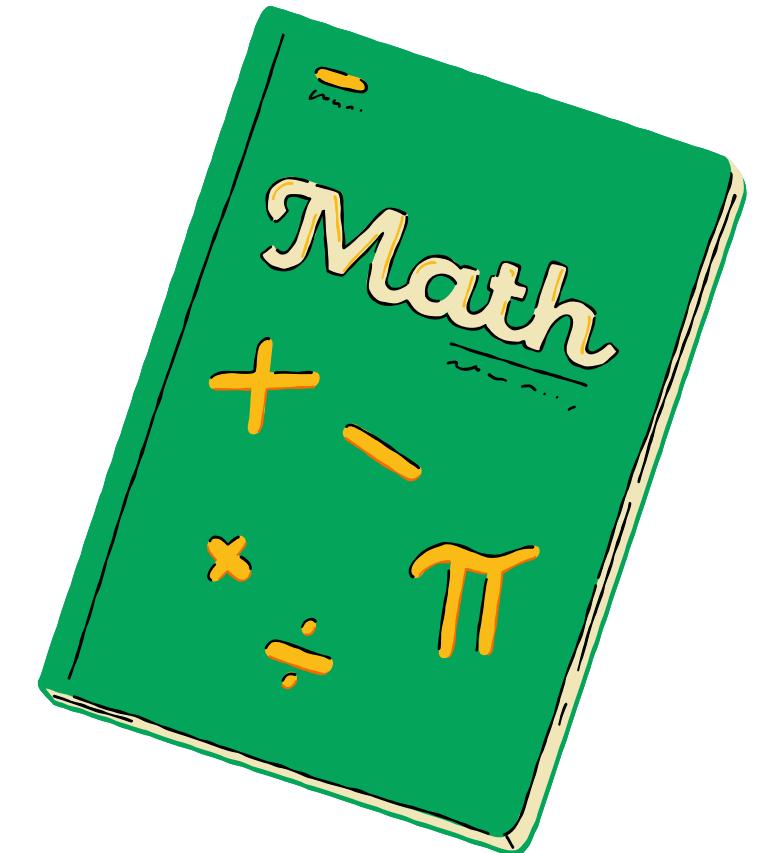
- One of the main goals of mathematics is to define structures and act on them, transform them, and promote equivalencies.



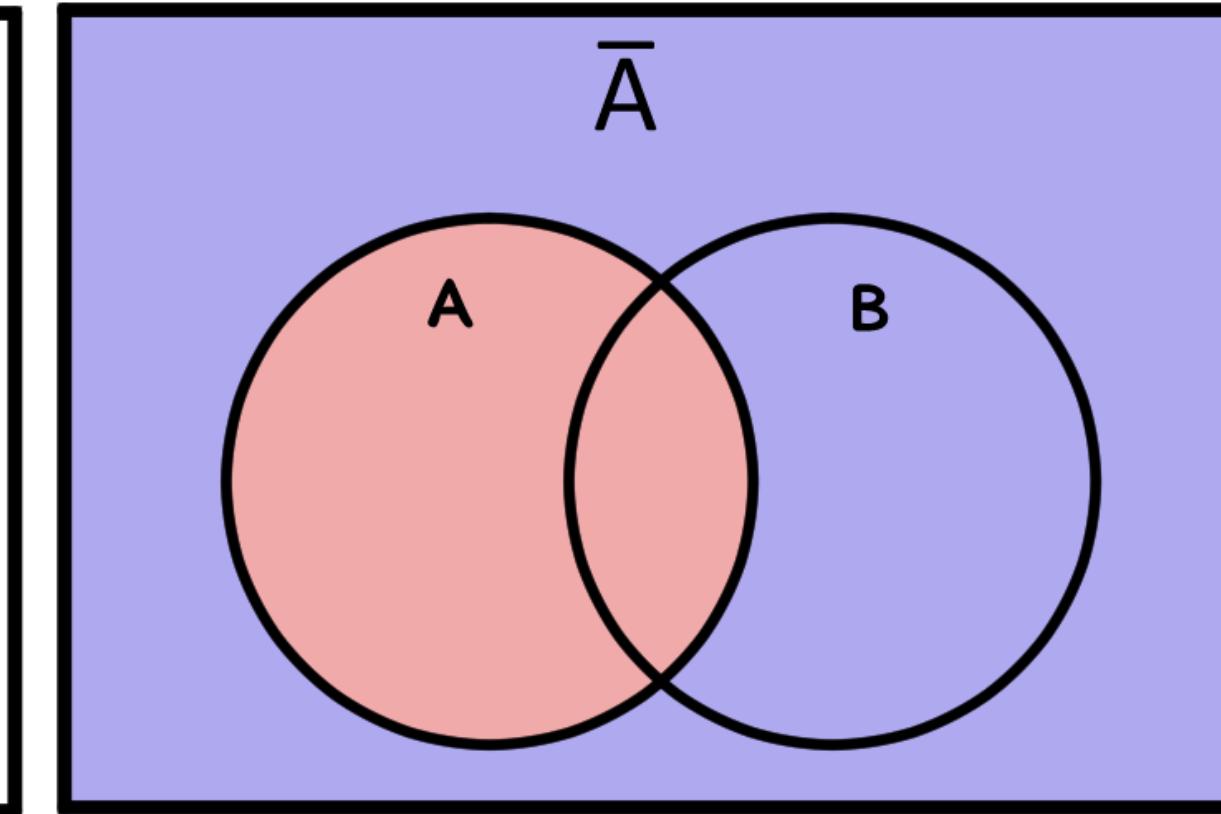
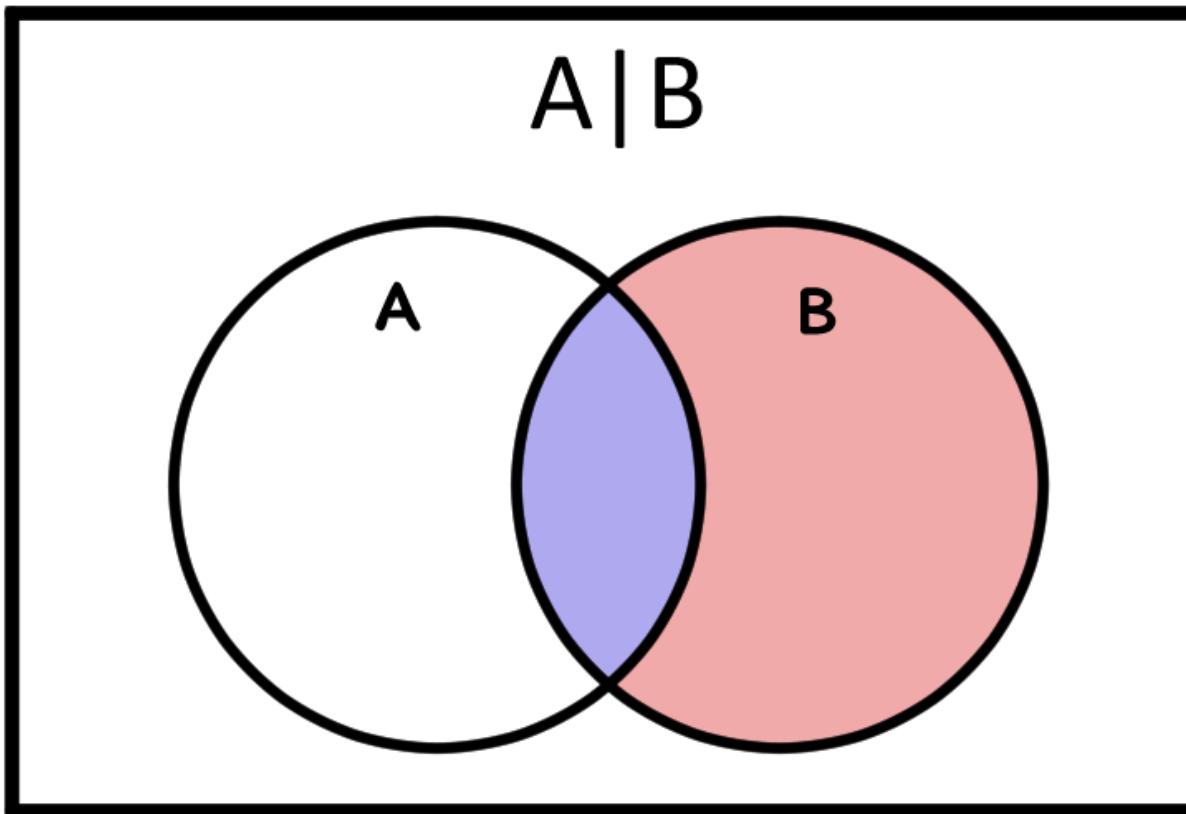
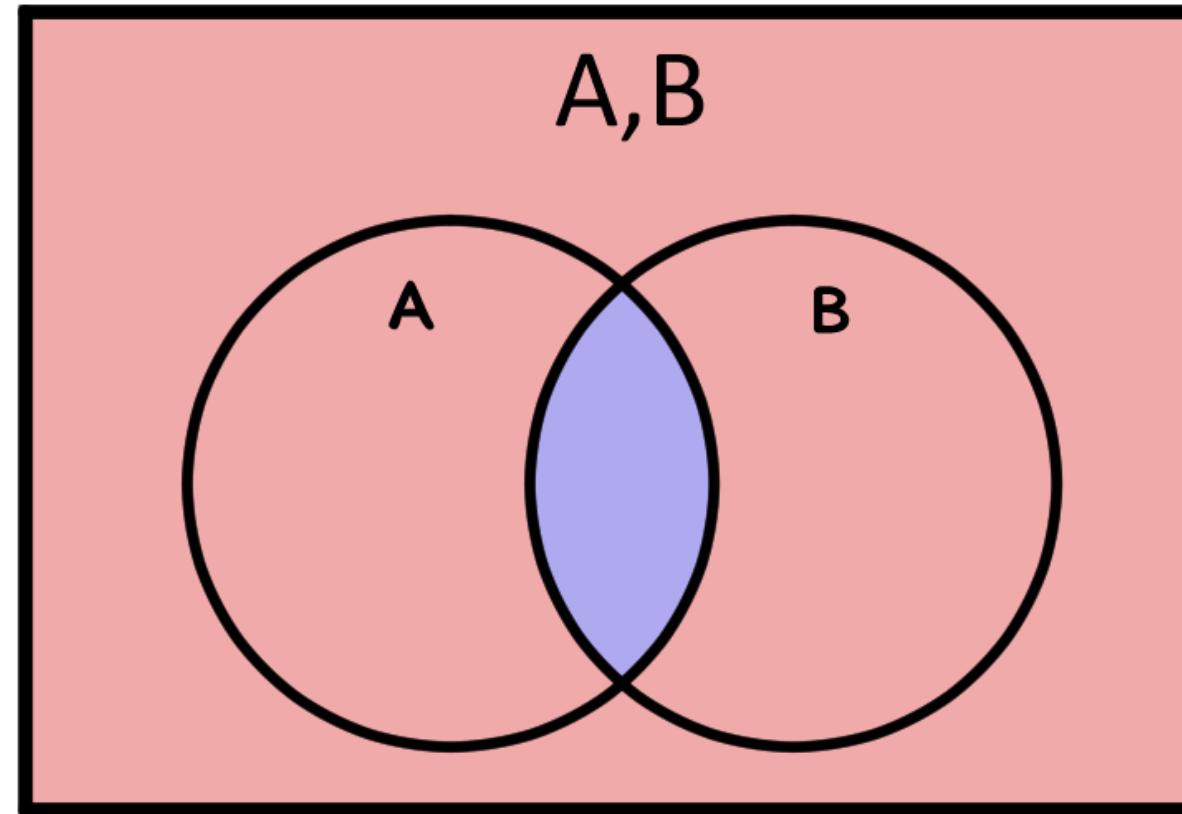
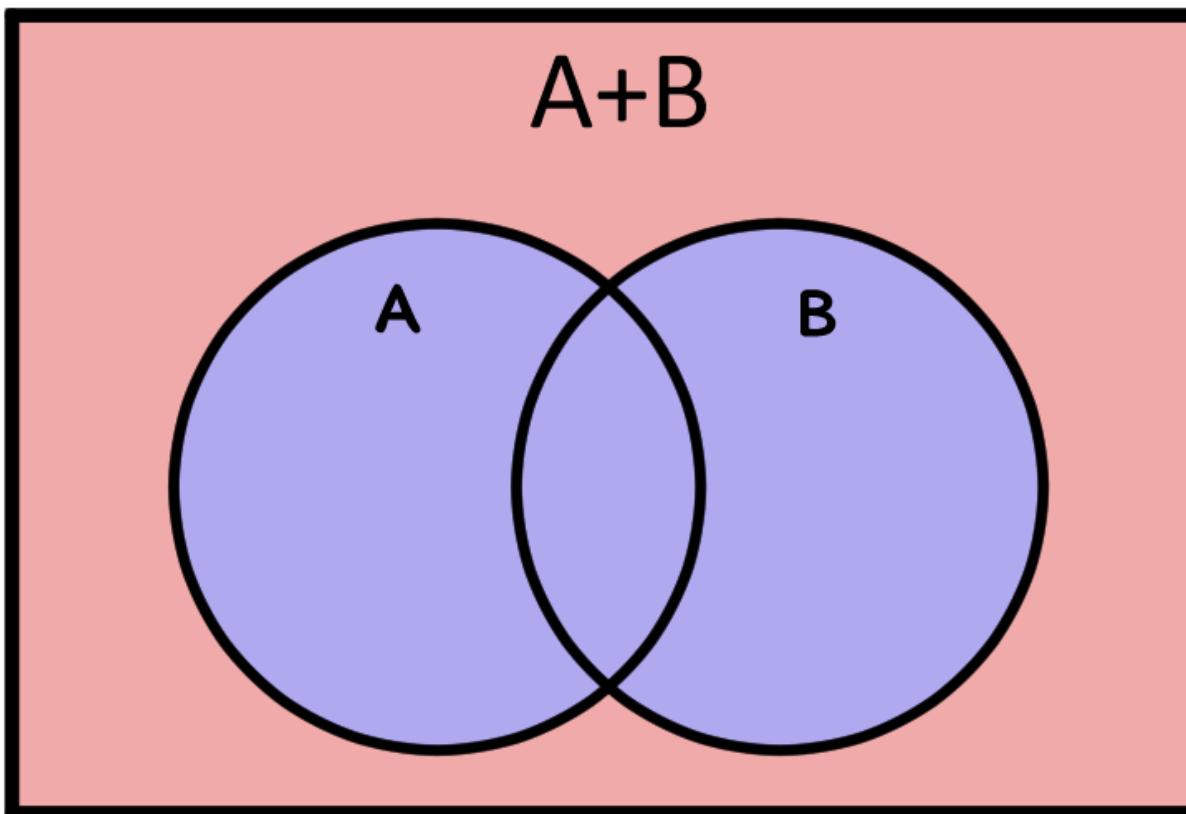
A little about ‘the way of math’

f(x)

- One of the main goals of mathematics is to define structures and act on them, transform them, and promote equivalencies.
- e.g., a function $y=F(x)$ is an action performed on the object x that transforms it into object y , and the same is true for $y=(d/dt)(x)$, which can equate to (encode) the factors changing x

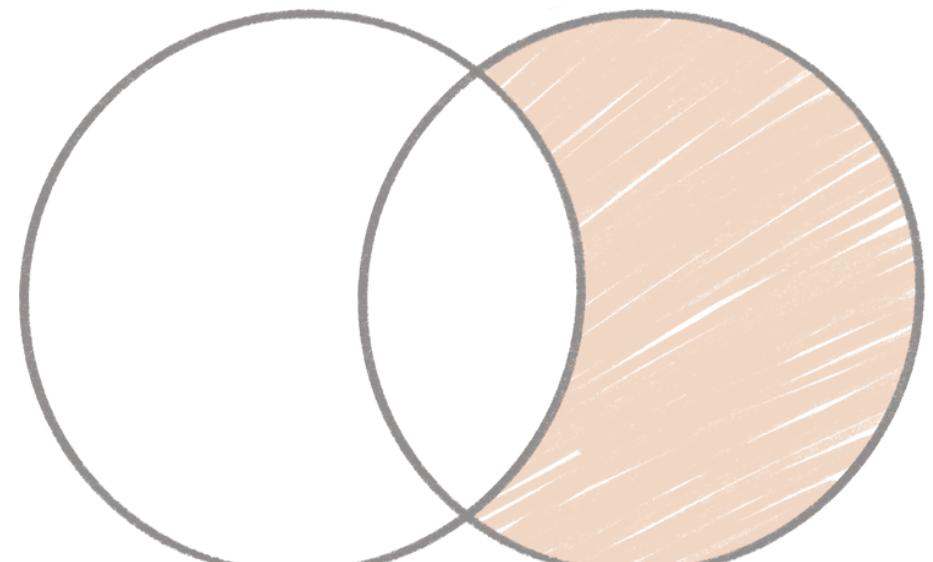


Set/Statement operations



Probability rules

- Sum: $A+B$ (meaning A or B)
- Product: $A.B$ (meaning A and B)
- Statements are recursive, closed: $C=A+B$, $D= (C=A+B)$
- Sum rule: $P(A+B) = P(A) + P(B) - P(A.B)$
- Product rule: $P(A.B) = P(A|B)P(B) = P(B|A)P(A)$
- Conditioning: $A|B$ is a special logical operation



Bayes equation

- Directly from product rule:

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

Bayes equation

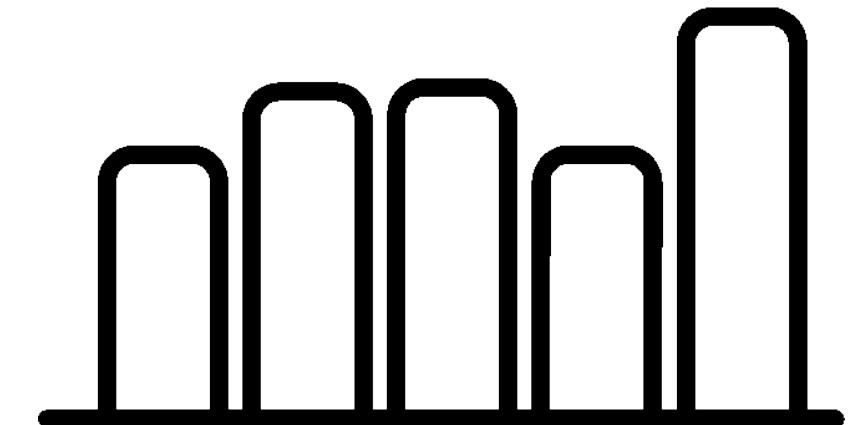
- Directly from product rule:

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

- The fundamental inversion of science:

$$P(M|D) \propto P(D|M)$$

Statistics

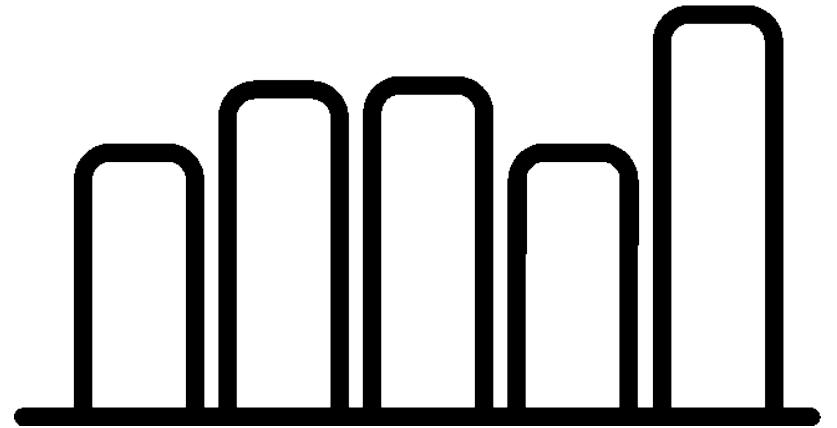


Provides mathematical methods to connect models with data. Since the connection is always uncertain, it's heavily based on the use of probability

Statistics

- Only two general goals:

Model selection:
comparison between
different models,
hypothesis testing

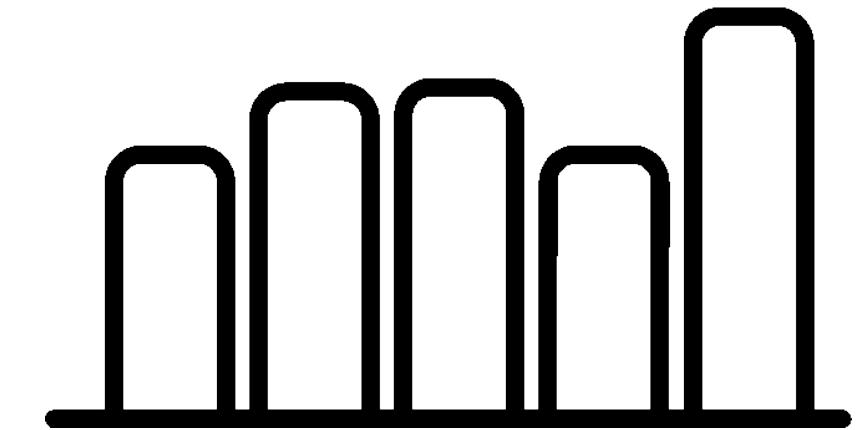


Model determination:
given a model, estimate
parameters, fit the model
to data

Statistics

- Only two general goals:

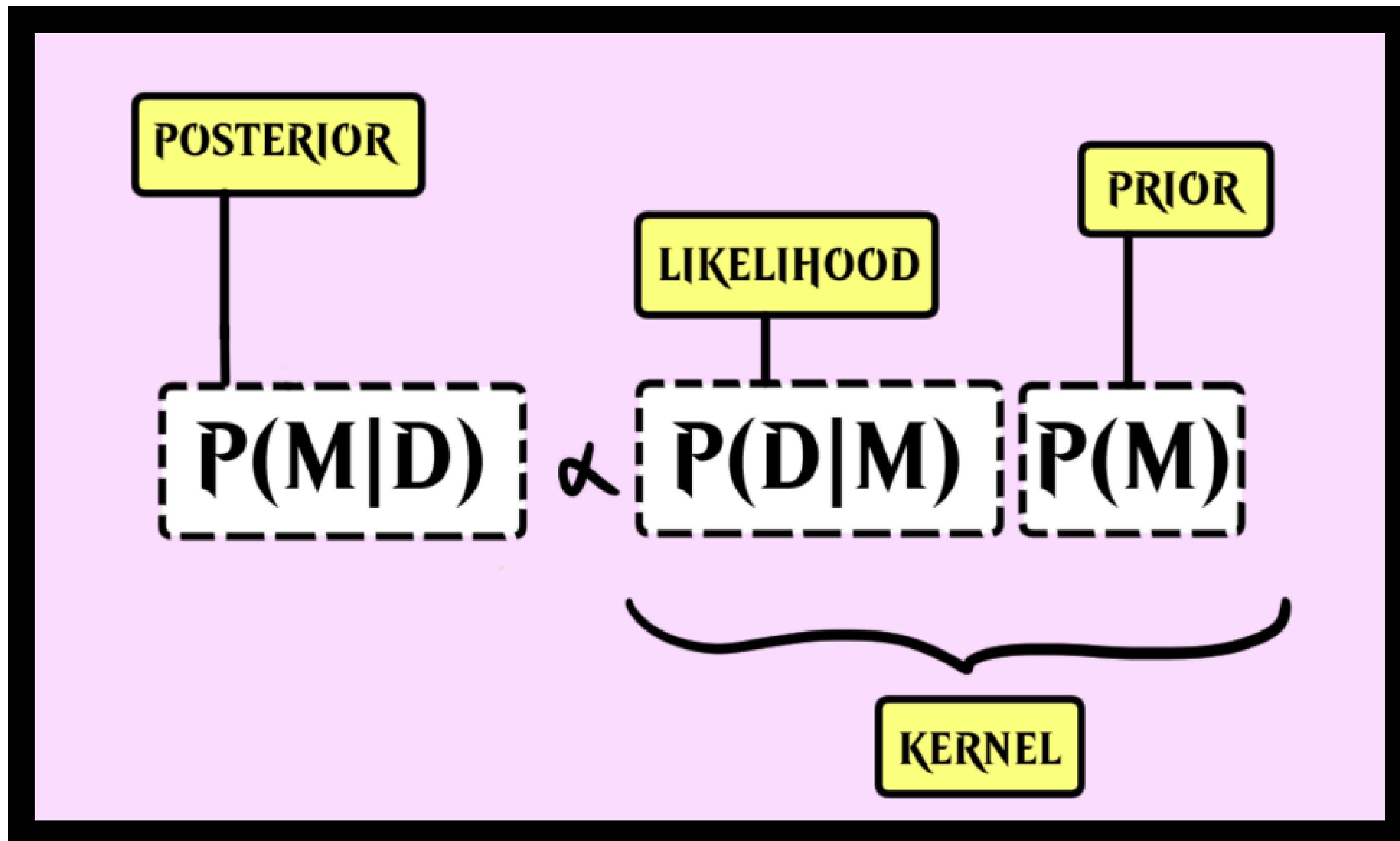
Model selection:
COMPARISON between
different models,
hypothesis testing



Model determination:
given a model, estimate
parameters, **FIT** the
model to data

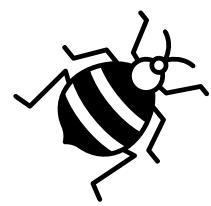
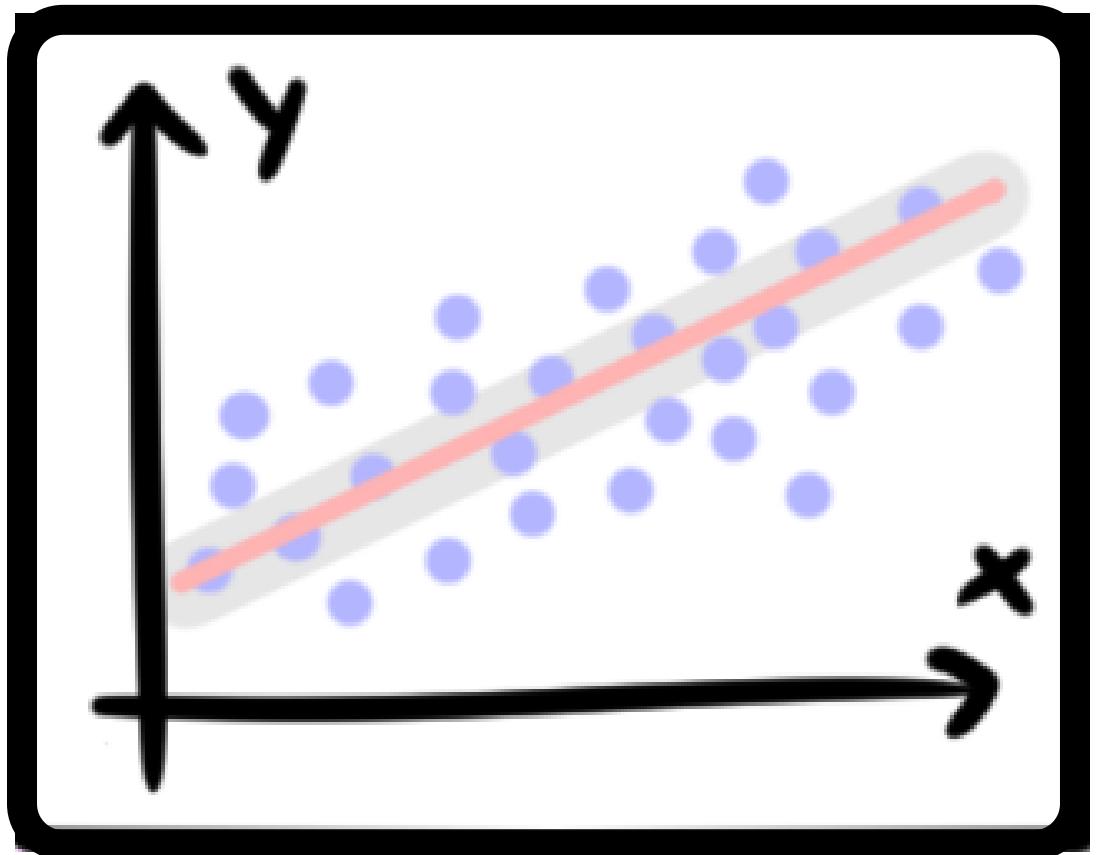
- Two assumptions are generally required: the model (relating the variables/parameters), and the data-generating distribution (the shape of uncertainty around the model). Why is this required?

Bayes equation



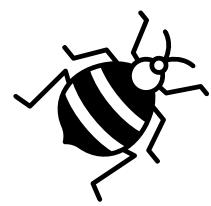
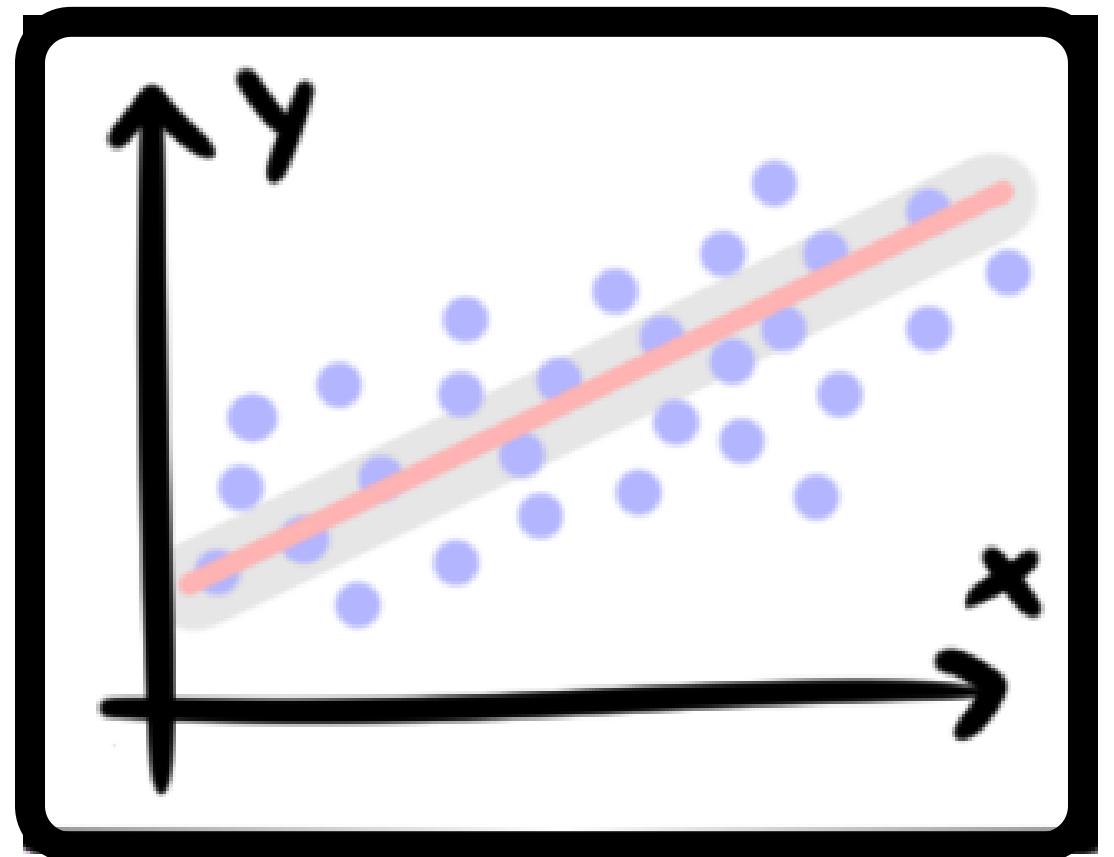
Example: Linear Regression

- Model:
 - $y = ax + b$



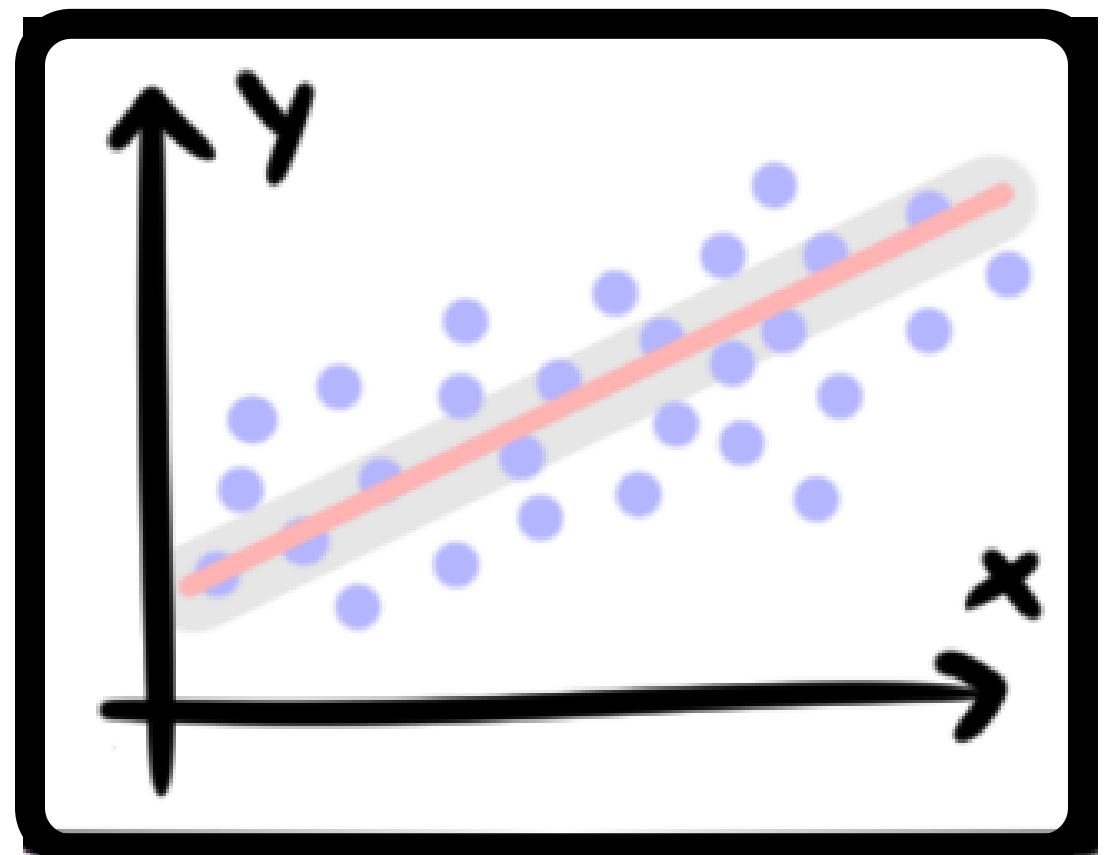
Example: Linear Regression

- Model:
 - $y = ax + b$
- Data-generating distribution (what if knowledge were precise?)
 - $N(y | ax+b, s^2)$



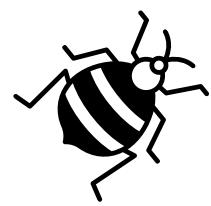
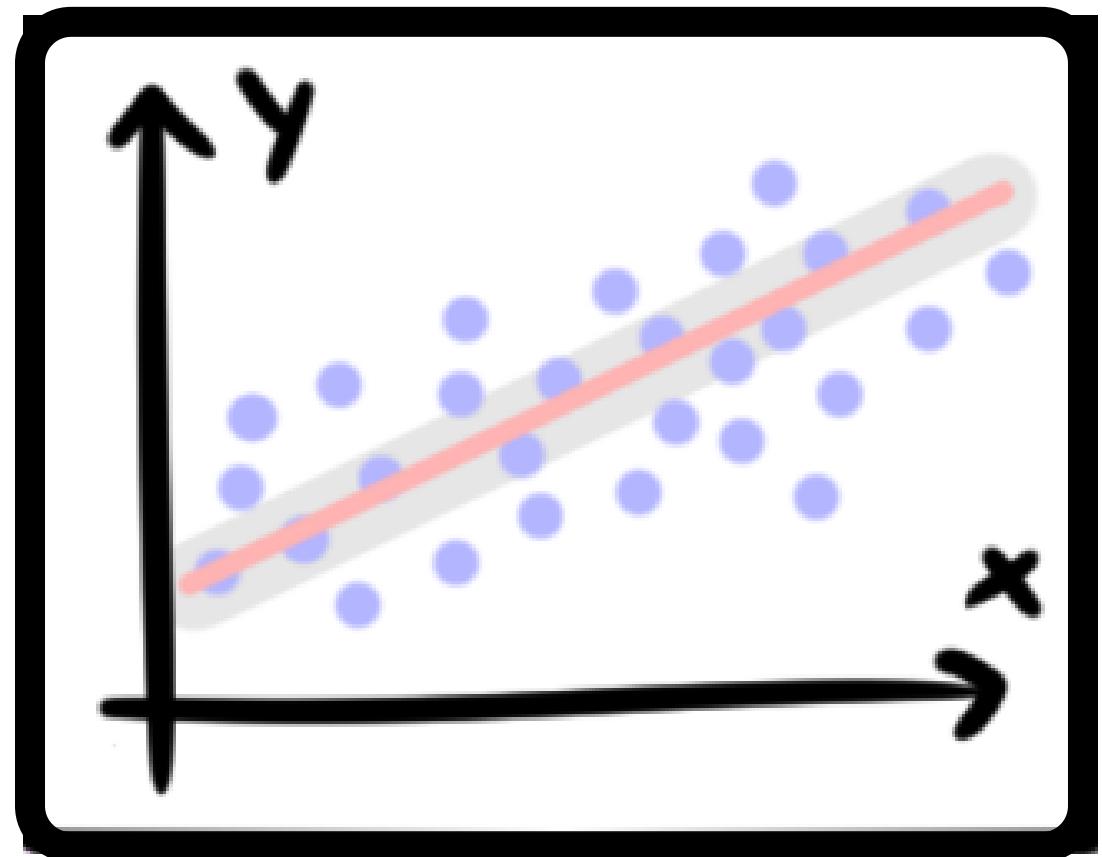
Example: Linear Regression

- Model:
 - $y = ax + b$
- Data-generating distribution (what if knowledge were precise?)
 - $N(y | ax+b, s^2)$
- Likelihood:
 - $P(D|M) = \prod_i \mathcal{N}(y_i | ax_i + b, s^2)$



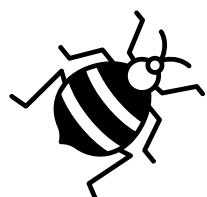
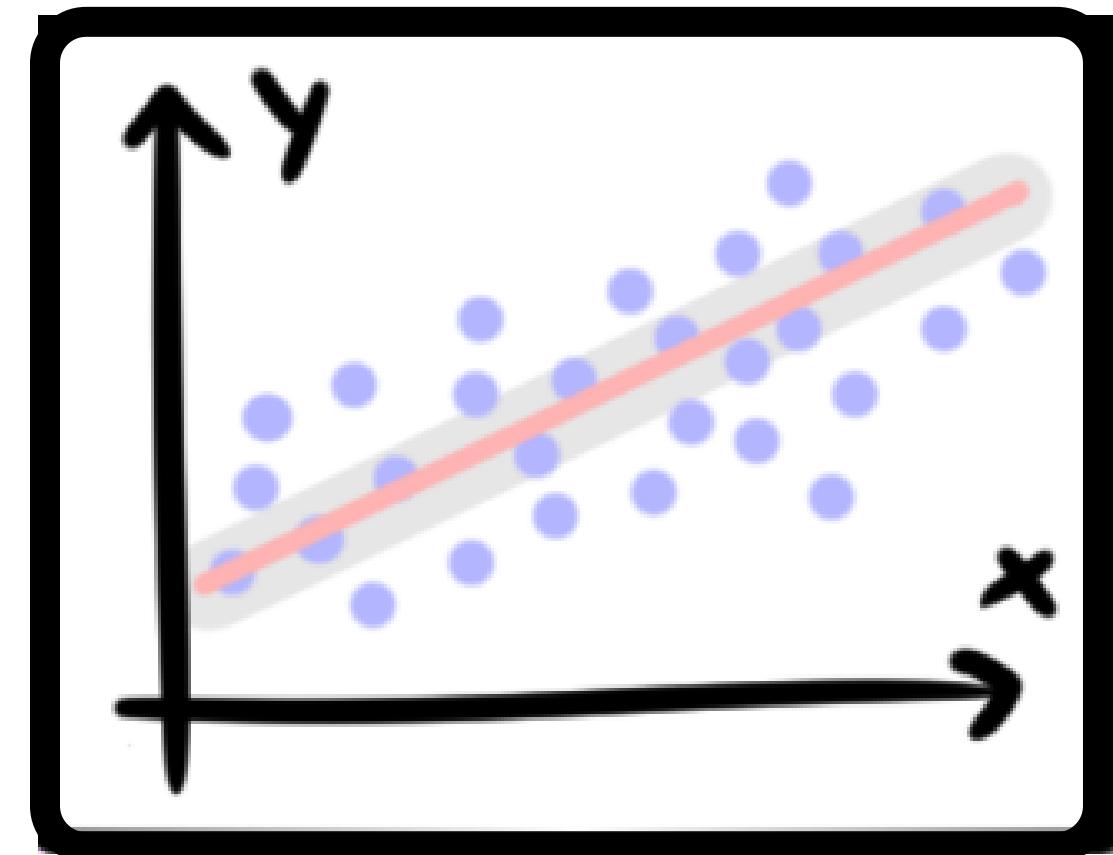
Example: Linear Regression

- Bayesian way: also inform the prior, estimate the posterior for a , b , and s .

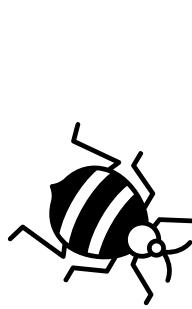


Example: Linear Regression

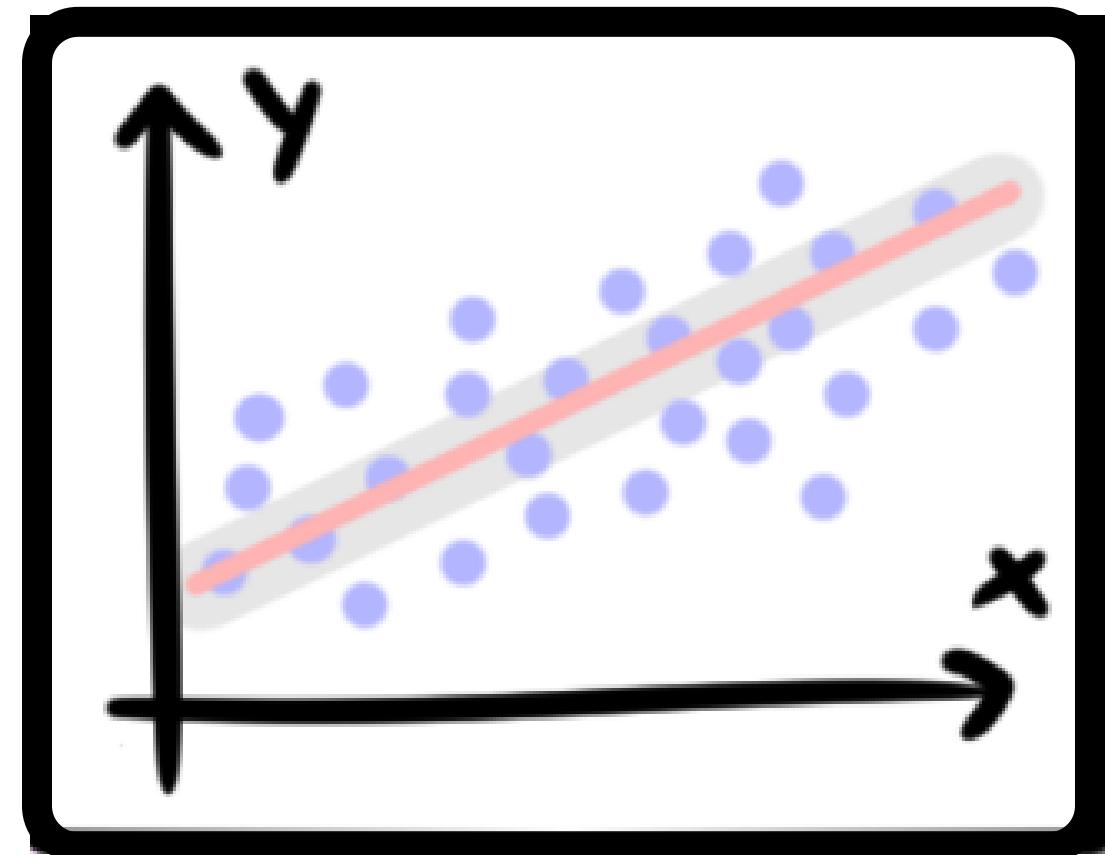
- Bayesian way: also inform the prior, estimate the posterior for a , b , and s .
- Frequentist way: assume flat prior, point-estimate a , b , and s as the values providing maximum likelihood (if you just derive the likelihood and make it zero, you arrive at the standard (a,b,s) for linear regression)



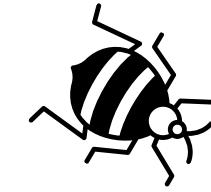
Example: Linear Regression



- This is also equivalent to the optimization procedure of linear least squares regression, where you minimize the loss function given by the sum of the squares of the “errors”
- Adding a regularization to the loss function is the same as recovering a non-flat prior



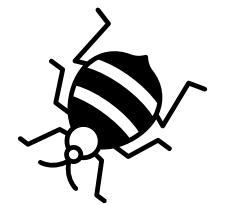
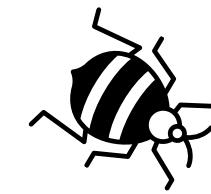
Example: Linear Regression



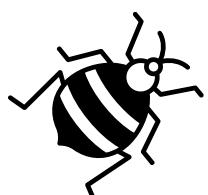
- We can interpret a linear regression as:
 - the proposal of a model to explain the relationship between x and y (linear)



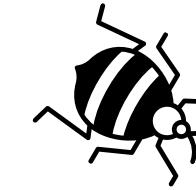
Example: Linear Regression



- We can interpret a linear regression as:
 - the proposal of a model to explain the relationship between x and y (linear)
 - the proposal of a model for uncertainty around the deterministic model (normal distribution)



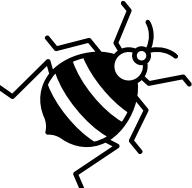
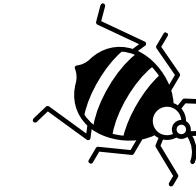
Example: Linear Regression



- We can interpret a linear regression as:
 - the proposal of a model to explain the relationship between x and y (linear)
 - the proposal of a model for uncertainty around the deterministic model (normal distribution)
 - = the statistical assessment to produce the evaluation of the statement saying that nature is significantly explained by the model (model determination giving a, b, s), by using an observation of nature (data)



Example: Linear Regression



- We can interpret a linear regression as:
 - the proposal of a model to explain the relationship between x and y (linear)
 - the proposal of a model for uncertainty around the deterministic model (normal distribution)
 - = the statistical assessment to produce the evaluation of the statement saying that nature is significantly explained by the model (model determination giving a, b, s), by using an observation of nature (data)
 - Also, GLMs are the same, but accept different models