Today

- Inference algorithms intro
- Frequentist inference algorithms
  - sampling distribution algorithm
  - coverage algorithm (confidence intervals)

Inference algorithms

★ ★ ★ ★ Crucial (consider later) ★ ★ ★ ★

Scope and veracity of inference depends on study design

Statistical inference

- Judge the accuracy of an estimation or prediction algorithm
  - Efron & Hastie 2016
- Reliability
- Uncertainty

ISO definition of accuracy: the closeness of a measurement to the true value
Two components: bias, variance

Different inference problems

Estimation
Infer a property of a population (e.g. mean) from a sample

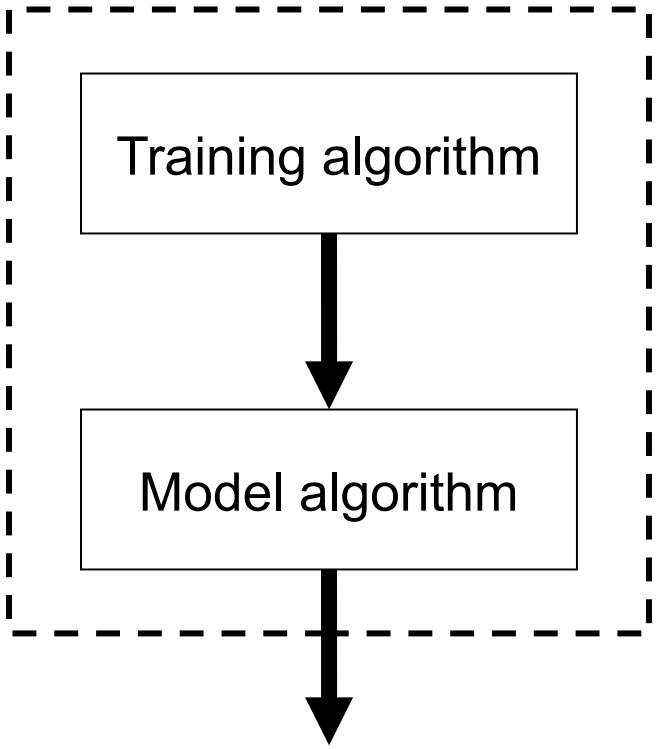
Model comparison (weigh evidence)
Infer the data generating process from among a set of candidate data-generating processes

Hypothesis test (association)
Infer that y is associated with x

Causation (by experiment or observation)
Infer that x causes y
Infer the size of an effect due to an intervention (estimation)
Infer that an intervention had an effect (H-test)

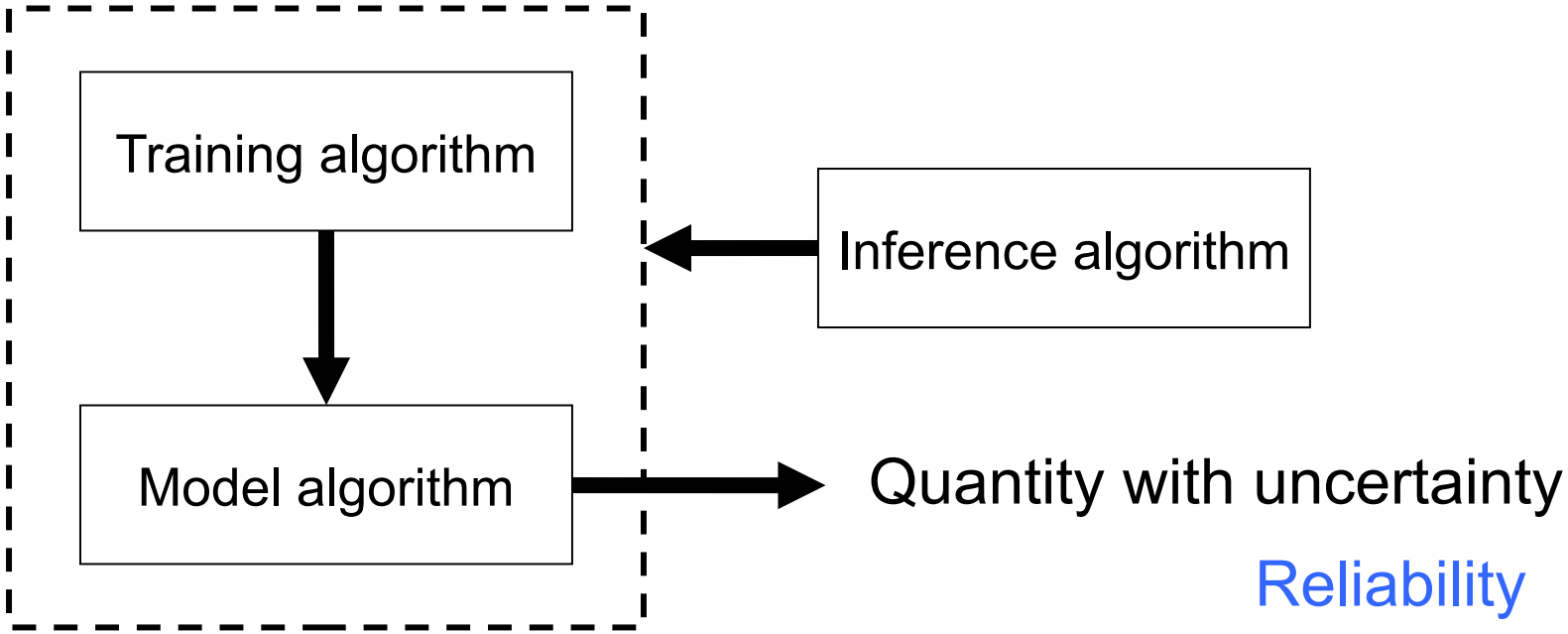
Prediction
Predict the value of a new observation or population state (extrapolation or interpolation)
Predict the population state in the future (forecast/extrapolation)

Algorithms in data science

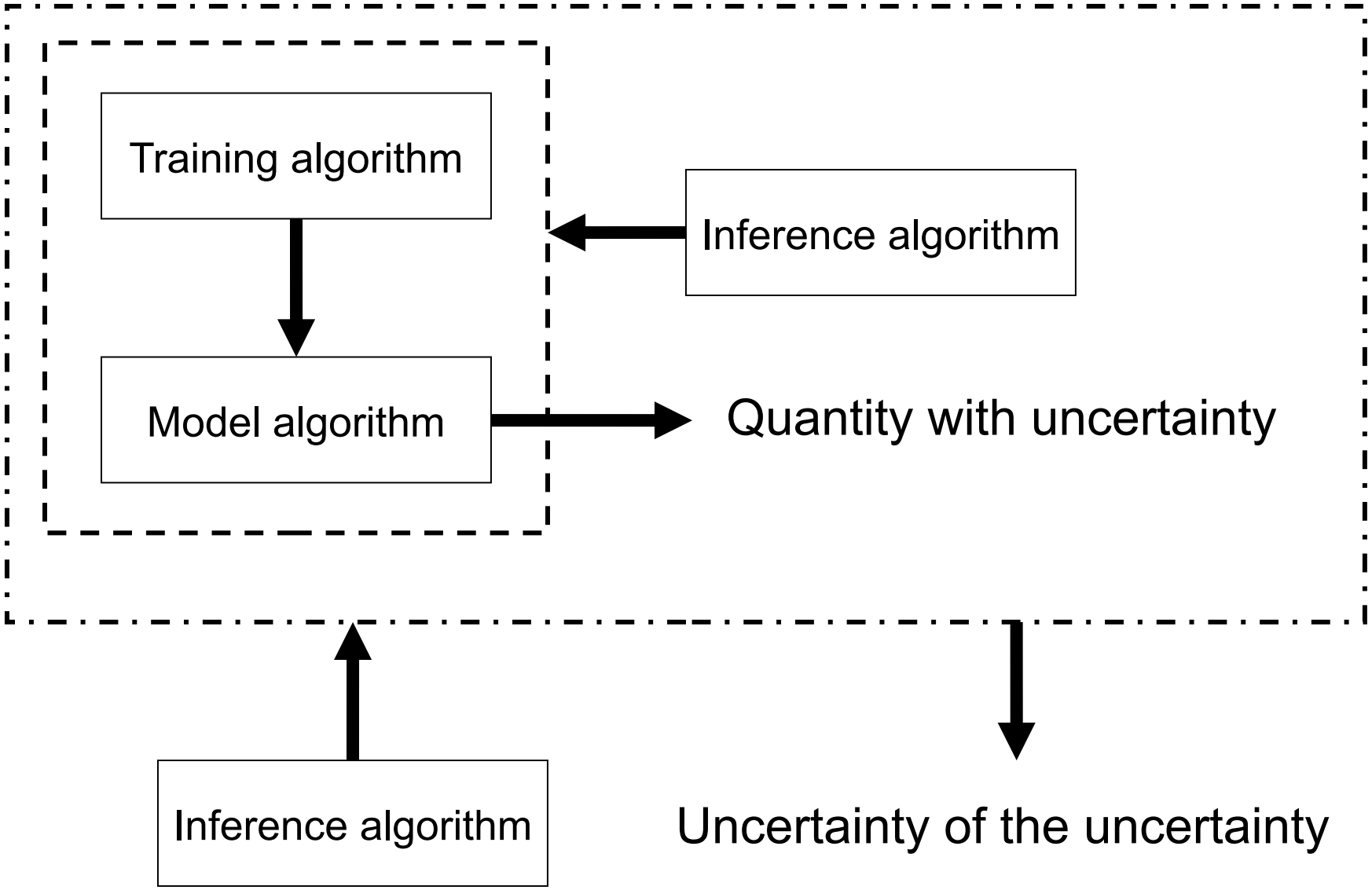


"Dumb" - doesn't say about reliability

Algorithms in data science



Algorithms in data science



Algorithms in data science

- Inference algorithm
  - looking back: considering all the ways data could have happened
    -
    -
    -
  - looking forward: predicting new data and testing against them
    -

These are two big ideas in data science

Algorithms in data science

- Inference algorithm
  - looking back: considering all the ways data could have happened
    - frequentist (sampling distribution)
    - likelihood (probability accounting)
    - Bayesian (likelihood + belief updating)
  - looking forward: predicting new data and testing against them
    -

These are two big ideas in data science

Algorithms in data science

- Inference algorithm
  - looking back: considering all the ways data could have happened
    - frequentist (sampling distribution)
    - likelihood (probability accounting)
    - Bayesian (likelihood + belief updating)
  - looking forward: predicting new data and testing against them
    - cross validation, AIC, machine learning

These are two big ideas in data science

Frequentist inference

- Frequentist probability = long-run frequency
  - e.g. tossing a coin

$$P(\text{Heads}) = \lim_{n \to \infty} \frac{\sum \text{Heads}}{n}$$

Sample vs population statistic

- Population statistic
  - e.g. mean weight
  - there is a true value
  - “fixed” not random
- Sample statistic
  - e.g. mean of sample
  - random variable

Sampling distribution

- Frequentist notion of looking back: considering all the ways data could have happened
- Imaginary repeated sampling from the data generating process

Sampling distribution algorithm

- Data generating process repeated many times
- Each time calc sample statistic

repeat very many times
    sample n units from the population
    calculate the sample statistic
plot sampling distribution (histogram) of the sample statistic

Make the algorithm

What is the mean weight of an individual of this species?

repeat very many times

    sample n units from the population

    calculate the sample statistic

plot sampling distribution (histogram) of the sample statistic

Important: population statistic, sample statistic