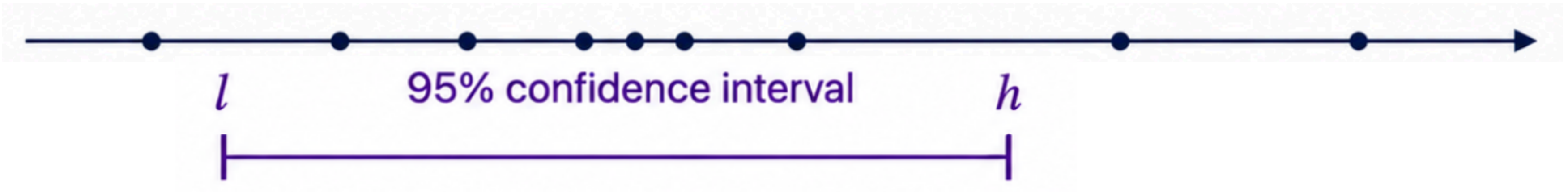
Bootstrap estimation

Natasha Jaques

UNIVERSITY of WASHINGTON

Confidence interval

Say I ran an experiment and got some data that looks like this:

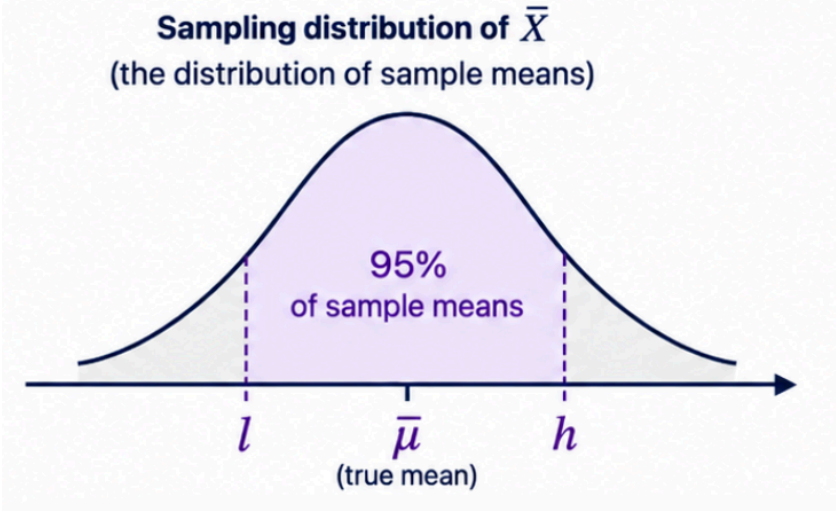


I can estimate the sample mean $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

What if I want a 95% **confidence interval [l,h]** for the true mean μ?

How could we get?

Run the experiment 10,000 times!??



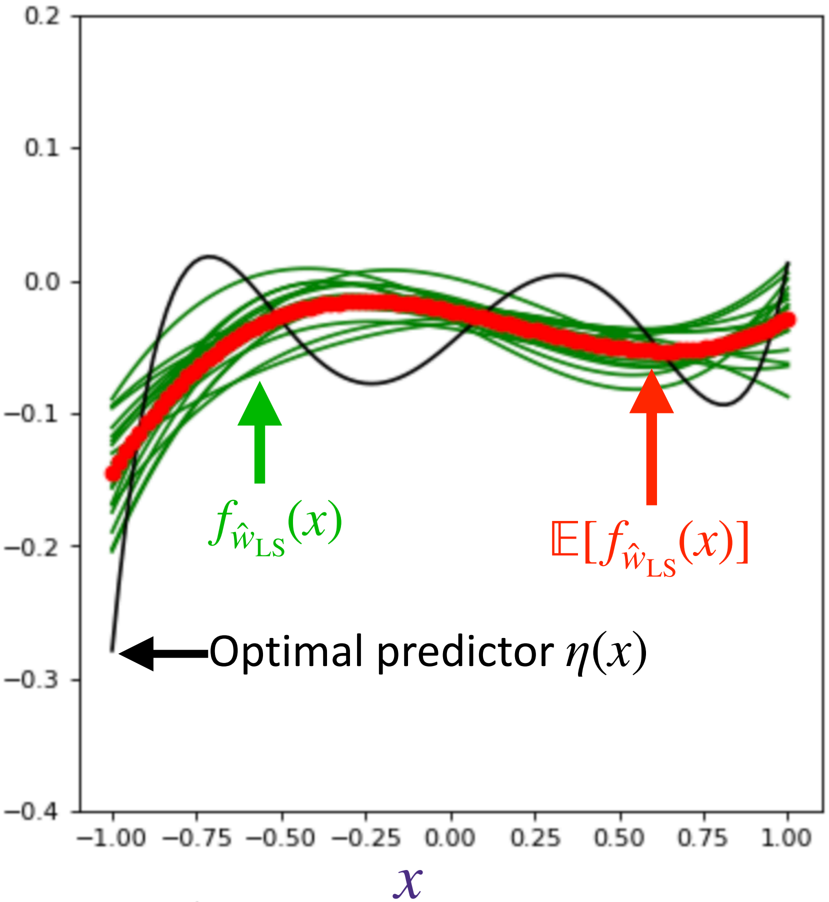
Bootstrapping: doing something seemingly impossible

- pull oneself up by one's bootstraps — improve one's position by one's own efforts without help from others.
- In our context: A general method of calculating uncertainty estimates without any additional data. (Efron, 1979)

Bootstrapping: doing something seemingly impossible

to pull oneself up by one's bootstraps

Remember bias-variance trade-off?



current train error = 0.0036791644380554187
current test error = 0.0037962529988410953

It is seemingly impossible to compute **Variance**, for example, with a single dataset.

Confidence interval

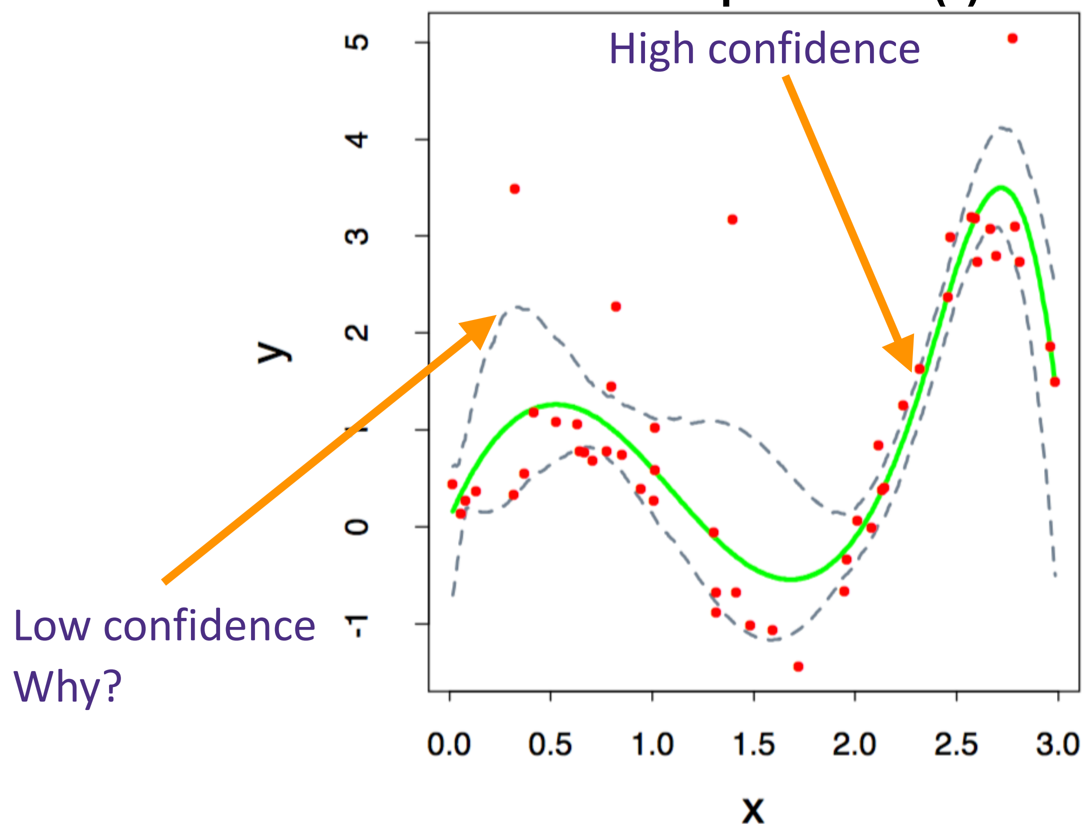
- suppose you have training data $\{(x_i, y_i)\}_{i=1}^n$ drawn i.i.d. from some true distribution $P_{x,y}$
- we train a ridge regressor with some polynomial feature degree

- we wish to build a confidence interval
  for our predictor $f(x)$, using
  5% and 95% percentiles

Confidence interval

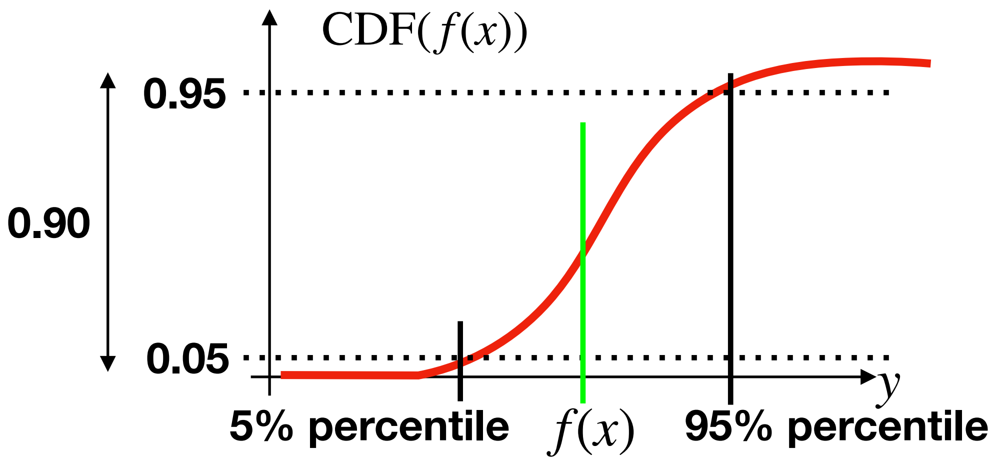
- suppose you have training data $\{(x_i, y_i)\}_{i=1}^n$ drawn i.i.d. from some true distribution $P_{x,y}$
- we train a ridge regressor with some polynomial feature degree
- we wish to build a confidence interval for our predictor $f(x)$, using 5% and 95% percentiles

Example of 5% and 95% percentile curves for predictor f(x)



Confidence interval

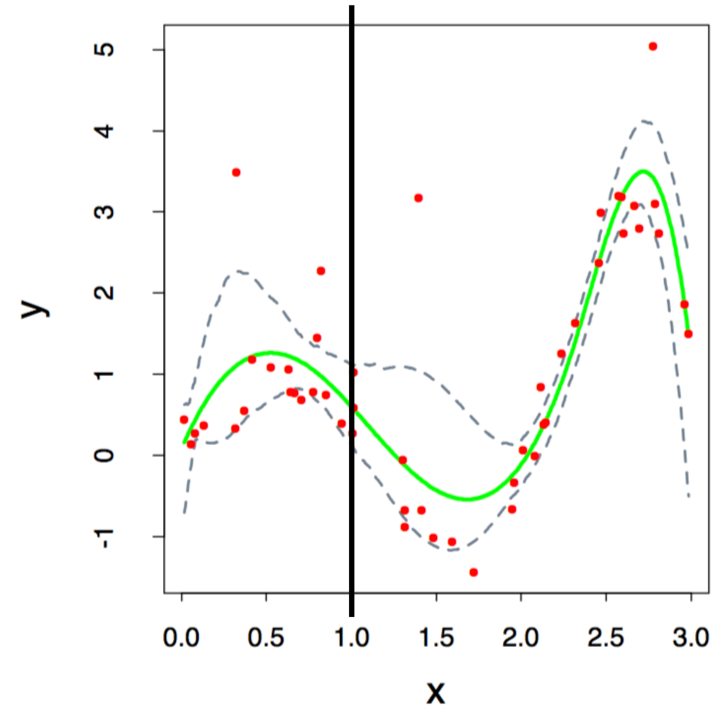
- let's focus on a single $x \in \mathbb{R}^d$
- note that our predictor $f(x)$ is a random variable, whose randomness comes from the training data $S_{\text{train}} = \{(x_i, y_i)\}_{i=1}^n$
- if we know the statistics (in particular the CDF of the random variable $f(x)$) of the predictor, then the **confidence interval** with **confidence level 90%** is defined as

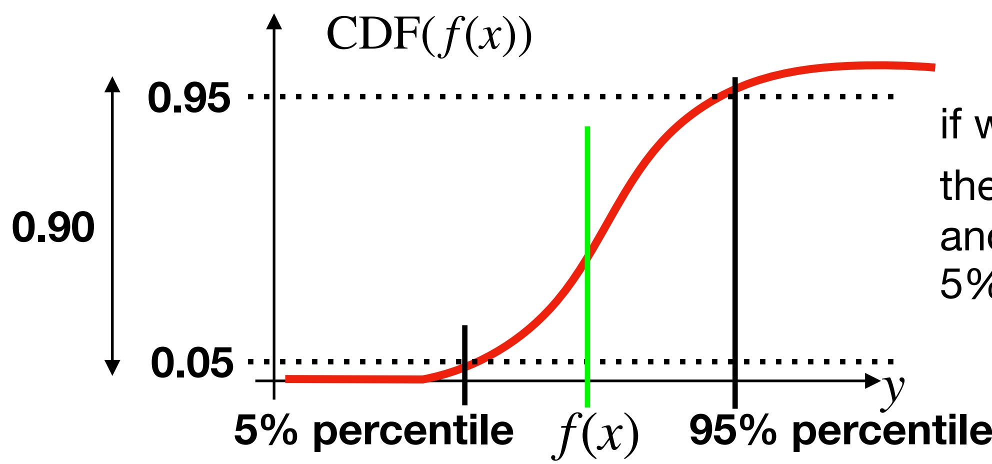


- as we do not have the cumulative distribution function (CDF), we need to approximate them

Confidence interval

- let's focus on a single $x \in \mathbb{R}^d$
- note that our predictor $f(x)$ is a random variable, whose randomness comes from the training data $S_{\text{train}} = \{(x_i, y_i)\}_{i=1}^n$
- if we know the statistics (in particular the CDF of the random variable $f(x)$) of the predictor, then the **confidence interval** with **confidence level 90%** is defined as





if we know the distribution of our predictor $f(x)$, the green line is the expectation $\mathbb{E}[f(x)]$ and the black dashed lines are the 5% and 95% percentiles in the figure above

- as we do not have the cumulative distribution function (CDF), we need to approximate them

Bootstrap

- as we cannot sample repeatedly (in typical cases), we use **bootstrap samples** instead
- bootstrap is a general tool for assessing statistical accuracy
- we learn it in the context of confidence interval for trained models

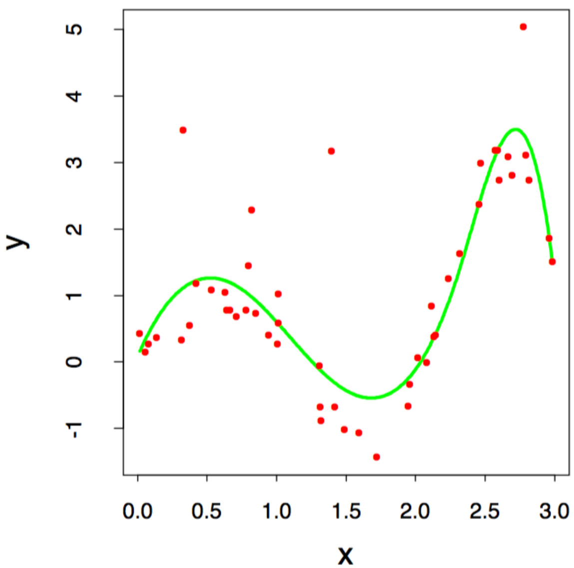
- a **bootstrap dataset** is created from the training dataset by taking n (the same size as the training data) examples uniformly at random **with replacement** from the training data $\{(x_i, y_i)\}_{i=1}^{n}$

- for b=1,…,B
  - create a bootstrap dataset $S^{(b)}_{\text{bootstrap}}$
  - train our model f(x) on $S^{(b)}_{\text{bootstrap}}$
  - Predict $\hat{y} = f(x)$

- compute the empirical CDF from the bootstrap datasets, and compute the confidence interval

bootstrap

training a single predictor



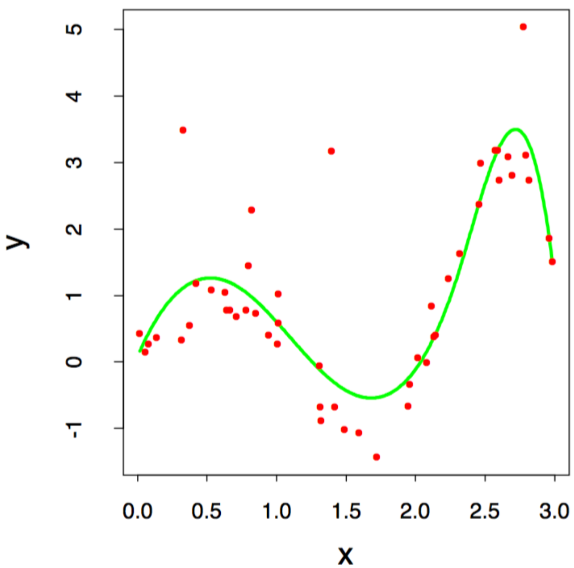
multiple bootstrapped predictors

90% confidence interval

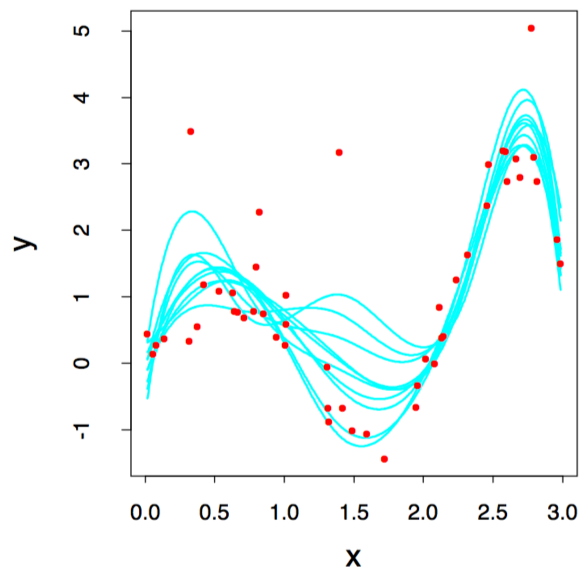
Figures from Hastie et al

bootstrap

training a single predictor



multiple bootstrapped predictors

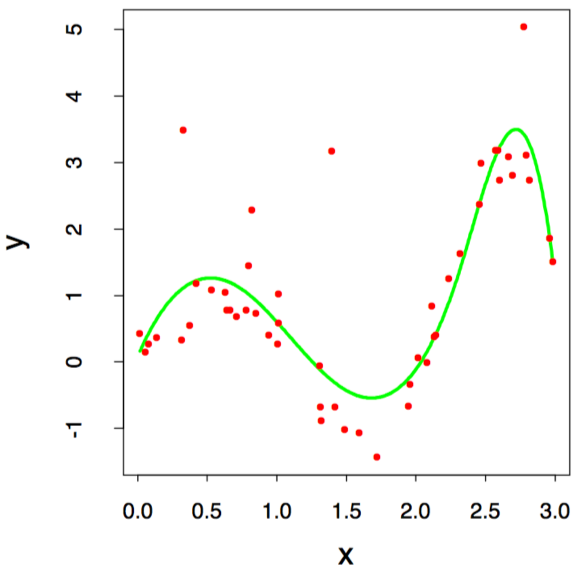


90% confidence interval

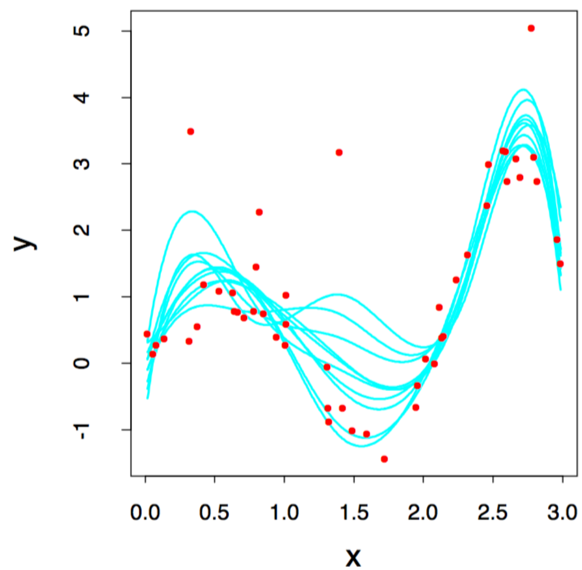
Figures from Hastie et al

bootstrap

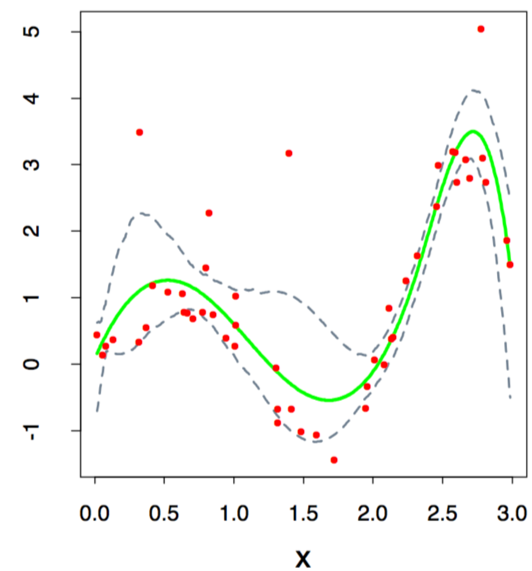
training a single predictor



multiple bootstrapped predictors

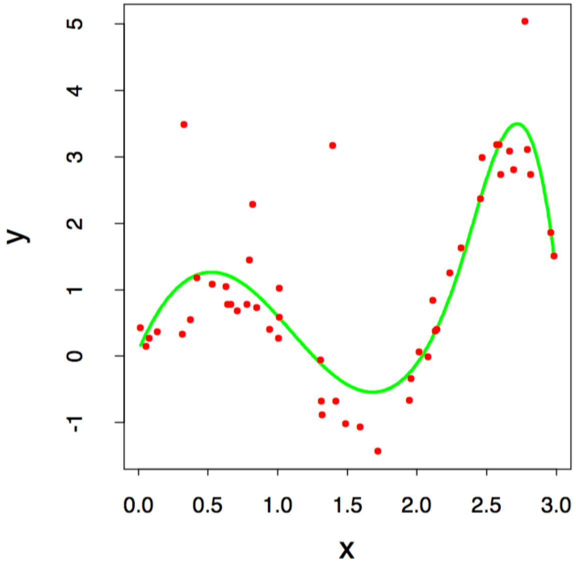


90% confidence interval

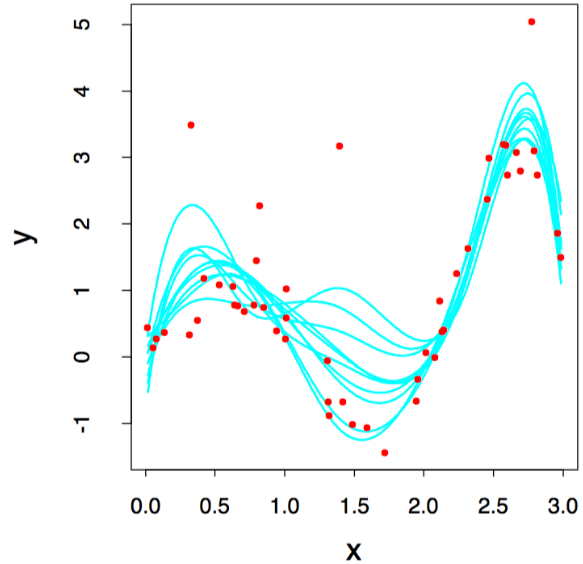


Figures from Hastie et al

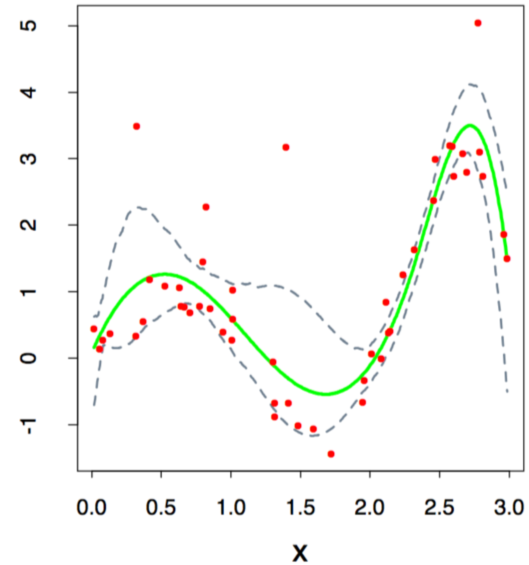
training a single predictor

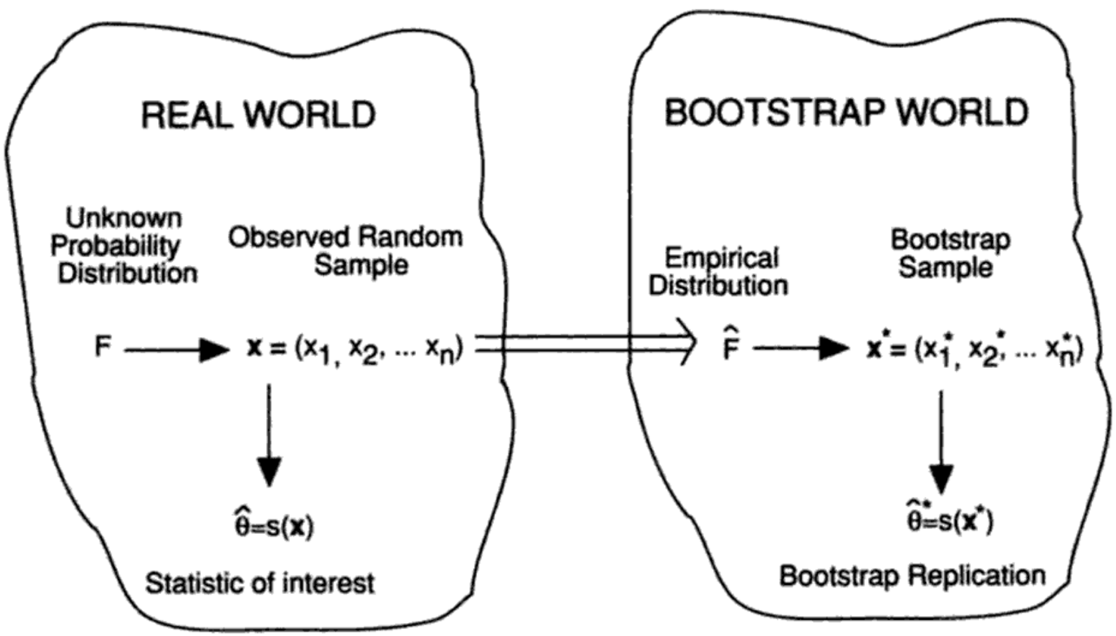


multiple bootstrapped predictors



90% confidence interval





Figures from Efron et al. 1993

Bootstrap Takeaways

Advantages:

- Very simple to use and generally applicable – build a confidence interval around anything
- Appears to give meaningful results even when the amount of data is very small
- Very strong asymptotic theory (as number of examples goes to infinity)

Disadvantages:

- Very few meaningful finite-sample guarantees
- Potentially computationally intensive
- Reliability hinges on test statistic and rate of convergence of empirical CDF to true CDF, which is unknown
- Poor performance on "extreme statistics" (e.g., the max)