Bootstrap

Limitations of CV

- An 80/20 split throws out a relatively large amount of data if only have, say, 20 examples.
- Test error is informative, but how accurate is this number? (e.g., 3/5 heads vs. 30/50)
- How do I get confidence intervals on statistics like the median or variance of a distribution?
- Instead of the error for the entire dataset, what if I want to study the error for a *particular example* x?

The Bootstrap: Developed by Efron in 1979.

Bootstrap: basic idea

cumulative distribution function

Given dataset drawn iid samples with $\boxed{\text{CDF}}$ $\underbrace{F_Z}$: $F_Z(x) = P(Z \leq x)$

$$\mathcal{D} = \underbrace{\{z_1, \ldots, z_n\}} \overset{i.i.d.}{\sim} F_Z$$

We compute a *statistic* of the data to get: $\widehat{\theta} = t(\mathcal{D})$

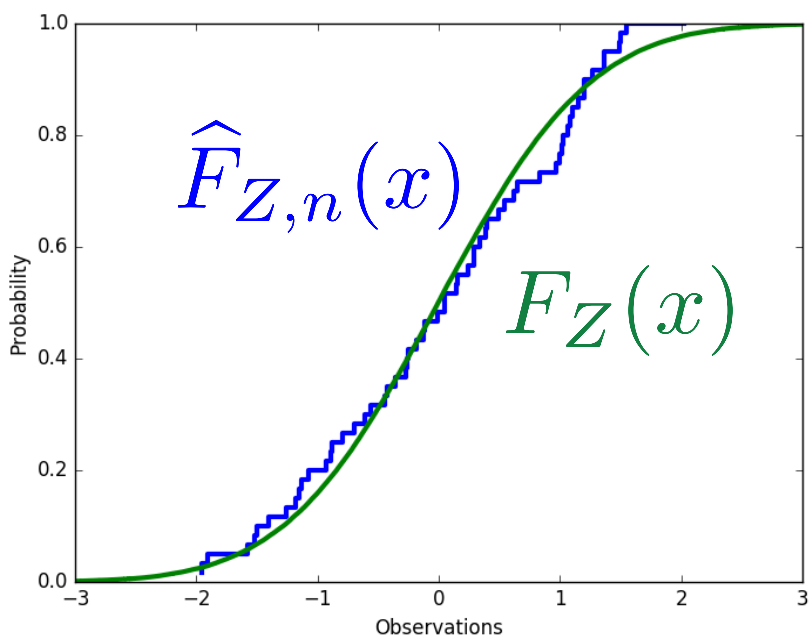
What is the distribution of $\widehat{\theta} = t(\mathcal{D})$

Bootstrap: basic idea

Given dataset drawn iid samples with CDF F_Z:

$$\mathcal{D} = \{z_1, \ldots, z_n\} \overset{i.i.d.}{\sim} F_Z$$

We compute a *statistic* of the data to get: $\widehat{\theta} = t(\mathcal{D})$



$$F_Z(x) = \mathbb{P}(Z \leq x)$$

$$\widehat{F}_{Z,n}(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{z_i \leq x\}$$

unbiased for $F_Z(x)$

$$\boxed{|\widehat{F}_{Z,n}(x) - F_Z(x)| \overset{n\to\infty}{\to} 0 \quad \text{a.s.}}$$

Bootstrap: basic idea

Given dataset drawn iid samples with CDF F_Z:

$$\mathcal{D} = \{z_1, \ldots, z_n\} \overset{i.i.d.}{\sim} F_Z$$

We compute a *statistic* of the data to get: $\widehat{\theta} = t(\mathcal{D})$

For b=1,…,B define the *b*th ***bootstrapped*** dataset as drawing *n* samples **with replacement** from *D*

$$\mathcal{D}^{*b} = \{z_1^{*b}, \ldots, z_n^{*b}\} \overset{i.i.d.}{\sim} \widehat{F}_{Z,n}$$

and the *b*th bootstrapped statistic as: $\theta^{*b} = t(\mathcal{D}^{*b})$

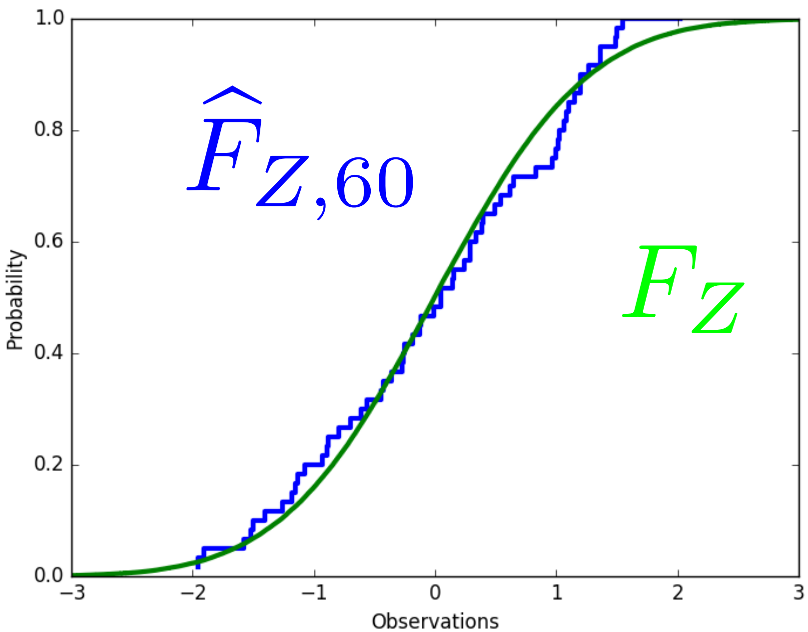
Bootstrap: basic idea

Given dataset drawn iid samples with CDF F_Z:

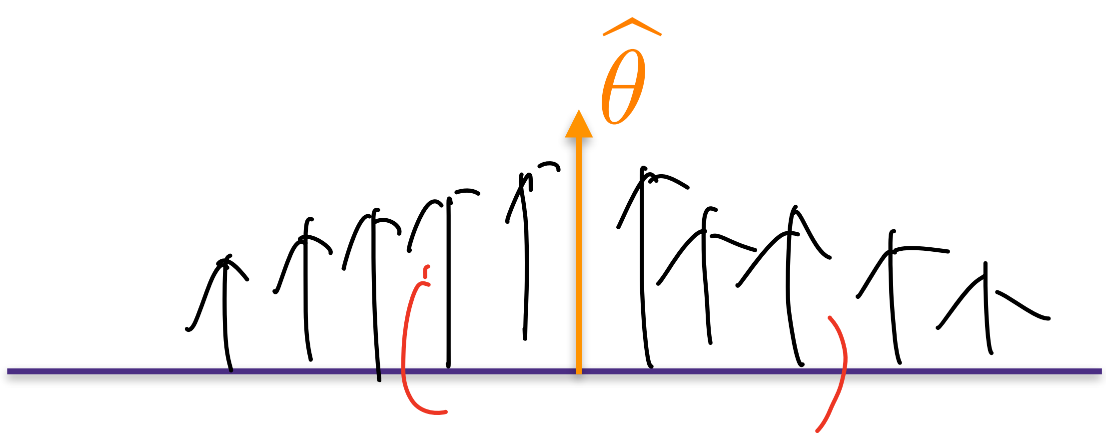
$$\mathcal{D} = \{z_1, \ldots, z_n\} \overset{i.i.d.}{\sim} F_Z$$

We compute a *statistic* of the data to get: $\widehat{\theta} = t(\mathcal{D})$

$$\mathcal{D}^{*b} = \{z_1^{*b}, \ldots, z_n^{*b}\} \overset{i.i.d.}{\sim} \widehat{F}_{Z,n} \qquad \theta^{*b} = t(\mathcal{D}^{*b})$$



$$\boxed{|\widehat{F}_{Z,n}(x) - F_Z(x)| \overset{n\to\infty}{\to} 0 \quad \text{a.s.}}$$

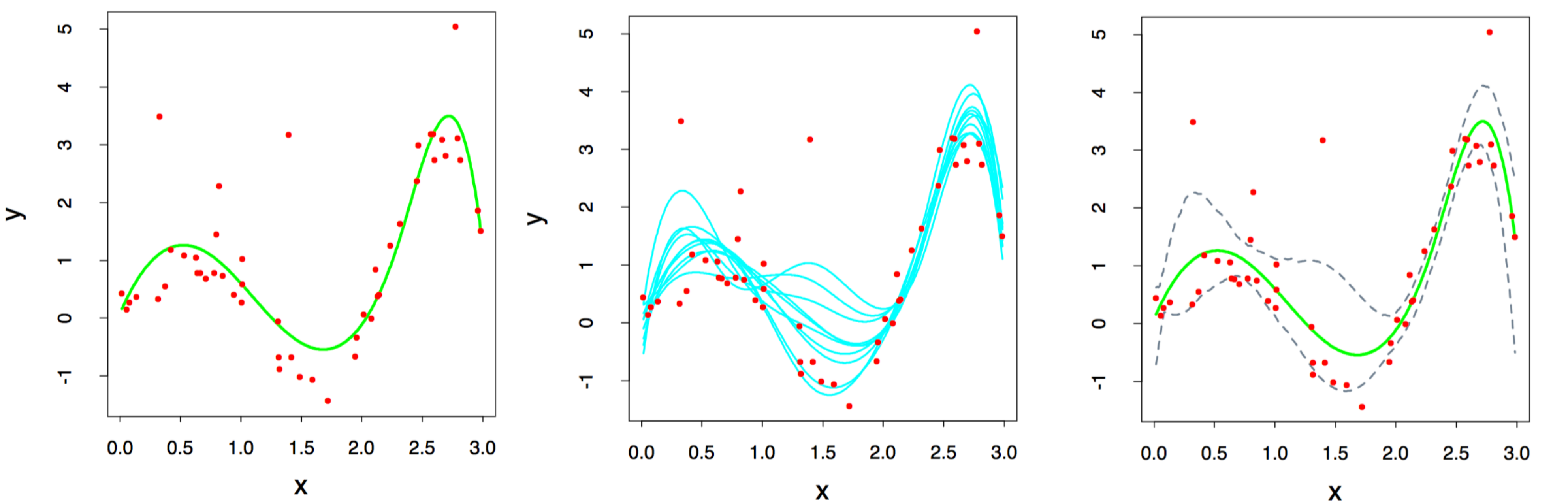


Applications

Common applications of the bootstrap:

- **Estimate parameters that escape simple analysis like the variance or median of an estimate**
- **Confidence intervals**
- **Estimates of error for a particular example:**

$\mathcal{D}$ $\widehat{\theta}$ θ^{*b} for $b = 1, \ldots, 10$ 95% confidence interval



Figures from Hastie et al

Takeaways

Advantages:

- **Bootstrap is very generally applicable. Build a confidence interval around *anything***
- **Very simple to use**
- **Appears to give meaningful results even when the amount of data is very small**
- **Very strong asymptotic theory (as num. examples goes to infinity)**

$B \to \infty$

Takeaways

Advantages:

- **Bootstrap is very generally applicable. Build a confidence interval around *anything***
- **Very simple to use**
- **Appears to give meaningful results even when the amount of data is very small**
- **Very strong asymptotic theory (as num. examples goes to infinity)**

Disadvantages

- **Very few meaningful finite-sample guarantees**
- **Potentially computationally intensive**
- **Reliability relies 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)**

$\hat{F}_Z \to F_Z$

Not perfect, but better than nothing.