

# Bootstrap Confidence Intervals

David S. Rosenberg

NYU: CDS

October 6, 2021

# Contents

1 Bias and Variance

2 The Bootstrap

## Bias and Variance

- Suppose we have a probability distribution  $P$ .
- Often we want to estimate some characteristic of  $P$ .
  - e.g. expected value, variance, kurtosis, median, etc...
- These things are called **parameters** of  $P$ .
- A **parameter**  $\mu = \mu(P)$  is any function of the distribution  $P$ .

- Question: Is  $\mu$  random?
- Answer: Nope. For example if  $P$  has density  $f(x)$  on  $\mathbb{R}$ , then mean is

$$\mu = \int_{-\infty}^{\infty} xf(x) dx,$$

which is just an integral - nothing random.

- Suppose  $\mathcal{D}_n = (X_1, X_2, \dots, X_n)$  is an i.i.d. sample from  $P$ .
- A **statistic**  $S = s(\mathcal{D}_n)$  is any function of the data.
- A statistic  $\hat{\mu} = \hat{\mu}(\mathcal{D}_n)$  is a **point estimator** of  $\mu$  if  $\hat{\mu} \approx \mu$ .

- Question: Are statistics and/or point estimators random?
- Answer: Yes, since we're considering the data to be random.
  - The function  $s(\cdot)$  isn't random, but we're plugging in random inputs.

# Examples of Statistics

- Mean:  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ .
- Median:  $m = \text{median}(X_1, \dots, X_n)$
- Sample variance:  $\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$

Fancier:

- A data histogram is a statistic.
- Empirical distribution function.
- A confidence interval.



# Statistics are Random

- Statistics are random, so they have probability distributions.
- The distribution of a statistic is called a **sampling distribution**.
- We often want to know some **parameters** of the sampling distribution.
  - Most commonly the mean and the standard deviation.
- The standard deviation of the sampling distribution is called the **standard error**.

- Question: Is standard error random?
- Answer: Nope. It's a parameter of a distribution.

# Bias and Variance for Real-Valued Estimators

- Let  $\mu = \mu(P)$  be a real-valued parameter of distribution  $P$ .
- Let  $\hat{\mu} = \hat{\mu}(\mathcal{D}_n)$  be a point estimator of  $\mu$ .
- We define the **bias** of  $\hat{\mu}$  to be  $\text{Bias}(\hat{\mu}) = \mathbb{E}\hat{\mu} - \mu$ .
- An estimator is **unbiased** if  $\text{Bias}(\hat{\mu}) = \mathbb{E}\hat{\mu} - \mu = 0$ .
- We define the **variance** of  $\hat{\mu}$  to be  $\text{Var}(\hat{\mu}) = \mathbb{E}\hat{\mu}^2 - (\mathbb{E}\hat{\mu})^2$ .

Neither bias nor variance depend on a specific sample  $\mathcal{D}_n$ . We are taking expectation over  $\mathcal{D}_n$ .

- Why might we care about the bias and variance of an estimator?

# Putting “Error Bars” on Estimators

- Why do we even care about estimating variance?
- May want to report a confidence interval for our point estimate, e.g.

$$\hat{\mu} \pm \sqrt{\widehat{\text{Var}}(\hat{\mu})}$$

- Where  $\sqrt{\widehat{\text{Var}}(\hat{\mu})}$  is our **estimate of the standard error** of  $\hat{\mu}$ .

## Estimating Variance of an Estimator

- To estimate  $\text{Var}(\hat{\mu})$ , we can use estimates of  $\mathbb{E}\hat{\mu}$  and  $\mathbb{E}\hat{\mu}^2$ .
- Instead of a single sample  $\mathcal{D}_n$  of size  $n$ , suppose we had
  - $B$  independent samples of size  $n$ :  $\mathcal{D}_n^1, \mathcal{D}_n^2, \dots, \mathcal{D}_n^B$
- Can then estimate

$$\mathbb{E}\hat{\mu} \approx \frac{1}{B} \sum_{i=1}^B \hat{\mu}(\mathcal{D}_n^i)$$

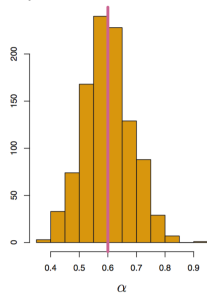
$$\mathbb{E}\hat{\mu}^2 \approx \frac{1}{B} \sum_{i=1}^B [\hat{\mu}(\mathcal{D}_n^i)]^2$$

and

$$\text{Var}(\hat{\mu}) \approx \frac{1}{B} \sum_{i=1}^B [\hat{\mu}(\mathcal{D}_n^i)]^2 - \left[ \frac{1}{B} \sum_{i=1}^B \hat{\mu}(\mathcal{D}_n^i) \right]^2.$$

# Histogram of Estimator

- Want to estimate  $\alpha = \alpha(P) \in \mathbb{R}$  for some unknown  $P$ , and some complicated  $\alpha$ .
- Point estimator  $\hat{\alpha} = \hat{\alpha}(\mathcal{D}_{100})$  for samples of size 100.
- How to get error bars on  $\hat{\alpha}$ ?
- Histogram of  $\hat{\alpha}$  for 1000 random datasets of size 100 (estimates sampling distribution of  $\hat{\alpha}$ ):



Pink line indicates true value of  $\alpha$ . This is Figure 5.10 from *An Introduction to Statistical Learning, with applications in R* (Springer, 2013) with permission from the authors: G. James, D. Witten, T. Hastie and R. Tibshirani.

- If magically we could get 1000 random datasets of size 100 from  $P$ , then we could make this histogram of  $\hat{\alpha}$ , which estimates the sampling distribution of  $\hat{\alpha}$ .
- If we are trying to estimate the standard error of  $\hat{\alpha}$  (i.e. the standard deviation of the sampling distribution of  $\hat{\alpha}$ ) to create a confidence interval, we can use the empirical standard deviation  $\hat{\sigma} = \text{SD}(\hat{\alpha}_1, \dots, \hat{\alpha}_{1000})$ . Then a 95% confidence interval would be something like  $\hat{\alpha} \pm 1.96\hat{\sigma}$ .
- Of course, in reality we only get one sample  $\mathcal{D}_{100}$  from  $P$ , not 1000 draws. Enter the magical bootstrap...

# The Bootstrap

---



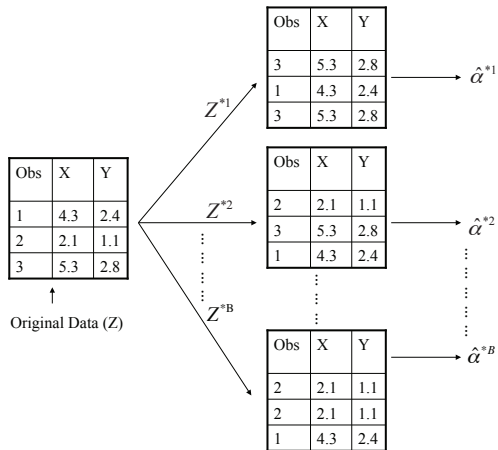
# The Bootstrap Sample

- A **bootstrap sample** from  $\mathcal{D}_n = (x_1, \dots, x_n)$  is a sample of size  $n$  drawn *with replacement* from  $\mathcal{D}_n$ .
- In a bootstrap sample, some elements of  $\mathcal{D}_n$ 
  - will show up multiple times, and
  - some won't show up at all.
- Each  $x_i$  has a probability of  $(1 - 1/n)^n$  of not being selected.
- Recall from analysis that for large  $n$ ,

$$\left(1 - \frac{1}{n}\right)^n \approx \frac{1}{e} \approx .368.$$

- So we expect  $\sim 63.2\%$  of elements of  $\mathcal{D}$  will show up at least once.

# The Bootstrap Sample



From *An Introduction to Statistical Learning, with applications in R* (Springer, 2013) with permission from the authors: G. James, D. Witten, T. Hastie and R. Tibshirani.

# The Bootstrap Method

## Definition

A **bootstrap method** is when you *simulate* having  $B$  independent samples from  $P$  by taking  $B$  bootstrap samples from the sample  $\mathcal{D}_n$ .

- Given original data  $\mathcal{D}_n$ , compute  $B$  bootstrap samples  $D_n^1, \dots, D_n^B$ .
- For each bootstrap sample, compute some function

$$\phi(D_n^1), \dots, \phi(D_n^B)$$

- Work with these values as though  $D_n^1, \dots, D_n^B$  were i.i.d.  $P$ .
- **Amazing fact:** Things often come out very close to what we'd get with independent samples from  $P$ .

# Independent vs Bootstrap Samples

- Want to estimate  $\alpha = \alpha(P)$  for some unknown  $P$  and some complicated  $\alpha$ .
- Point estimator  $\hat{\alpha} = \hat{\alpha}(\mathcal{D}_{100})$  for sample of size 100.
- Histogram of  $\hat{\alpha}$  based on
  - 1000 independent samples of size 100, vs
  - 1000 bootstrap samples of size 100 from the original sample  $\mathcal{D}_{100}$

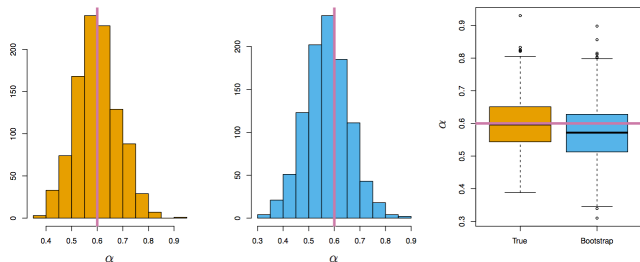


Figure 5.10 from *ISLR* (Springer, 2013) with permission from the authors: G. James, D. Witten, T. Hastie and R. Tibshirani.

## Normal approximated confidence intervals

- Suppose we have a point estimator  $\hat{\mu} = \hat{\mu}(\mathcal{D}_n)$ .
- To get error bars, we can compute the “bootstrap variance”.
  - Draw  $B$  bootstrap samples from  $\mathcal{D}_n$ .
  - Compute sample SD:  $\hat{\sigma} = \text{SD}(\hat{\mu}(D_n^1), \dots, \hat{\mu}(D_n^B))$ .
- The normal approximated bootstrap confidence interval is

$$\hat{\mu} \pm q_{\alpha/2} \hat{\sigma},$$

where  $q_{\alpha}$  is the  $\alpha$ -quantile of a standard normal.

- For a typical 95% confidence interval, we'd take  $\alpha = 0.05$  and find  $q_{\alpha/2} \approx -1.96$ .