# The Bootstrap

Alexander Quispe

November 5, 2021

# The Bootstrap

These notes are based on the book *An introduction to statistical learning (Witten, Hastie, Tibshirani 2013)*

- Used to quantify the uncertainty associated with a given estimator or statistical learning method.
- For instance, to obtain an estimate of the standard error of a coefficient, or a confidence interval for that coefficient.
- The use of the term bootstrap derives from the phrase **to pull oneself up by one's bootstraps.**

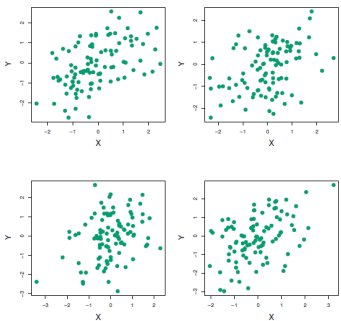
# A simple example

- Suppose that we wish to invest a fixed sum of money in two financial assets that yield returns of $X$ and $Y$ , respectively, where $X$ and $Y$ are random quantities.
- We will invest a fraction $\alpha$ of our money in $X$, and will invest the remaining $1 - \alpha$.
- Find the $\alpha$ that minimizes the total risk. In other words, we want to minimize $Var(\alpha X + (1 - \alpha) Y)$

# A simple example

Figure: Figure 5.9 of ISL

## Example continued




*Each panel displays* 100 *simulated returns for investments* $X$ *and* $Y$. *From left to right and top to bottom, the resulting estimates for* $\alpha$ *are* 0.576, 0.532, 0.657, *and* 0.651.

# A simple example

Figure: Figure 5.9 of ISL

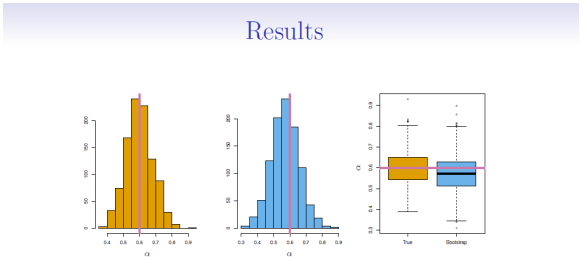
## Example continued

- But the values of $\sigma_X^2$, $\sigma_Y^2$, and $\sigma_{XY}$ are unknown.
- We can compute estimates for these quantities, $\hat{\sigma}_X^2$, $\hat{\sigma}_Y^2$, and $\hat{\sigma}_{XY}$, using a data set that contains measurements for $X$ and $Y$.
- We can then estimate the value of $\alpha$ that minimizes the variance of our investment using

$$\hat{\alpha} = \frac{\hat{\sigma}_Y^2 - \hat{\sigma}_{XY}}{\hat{\sigma}_X^2 + \hat{\sigma}_Y^2 - 2\hat{\sigma}_{XY}}.$$

# A simple example

Figure: Figure 5.9 of ISL




*Left:* A histogram of the estimates of $\alpha$ obtained by generating 1,000 simulated data sets from the true population. *Center:* A histogram of the estimates of $\alpha$ obtained from 1,000 bootstrap samples from a single data set. *Right:* The estimates of $\alpha$ displayed in the left and center panels are shown as boxplots. In each panel, the pink line indicates the true value of $\alpha$.