5.2 The Bootstrap

Zongyi Liu

2023-05-24

5.2 The Bootstrap

Bootstrap

The bootstrap is a widely applicable and useful statistical tool that can be used to quantify the uncertainty associated with a given estimator or statistical learning method.

It also has advantage that it can be easily applied to a wide range of statistical learning methods, including some for which a measure of variability is otherwise difficult to obtain and is not automatically output by statistical software.

Example

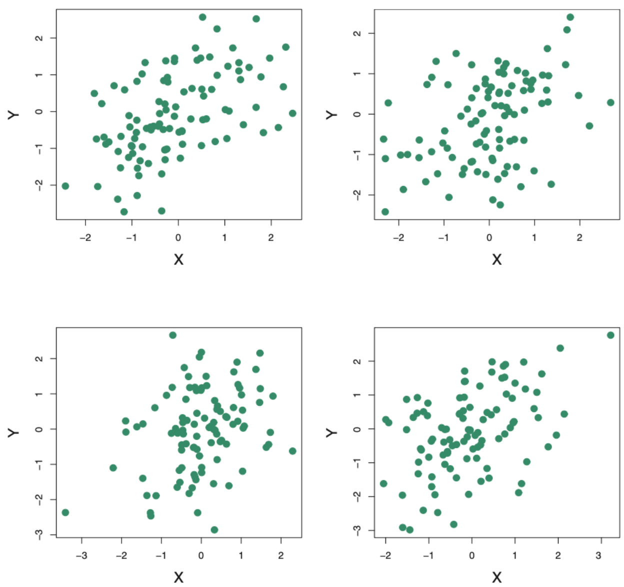
Suppose we want to invest a fixed sum of money in two financial assets that yield returns of X and Y, where X and Y are random quantities. We will need an α for X and $(1-\alpha)$ for Y, which could minimize the total risk, or variance of the investment. In other words, we need to minimize $Var(\alpha X + (1-\alpha)Y$:

$$\alpha = \frac{\sigma_Y^2 - \sigma_{XY}}{\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}},$$

Using the past data set, we can estimate the value of α using

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

We simulated 100 pairs of returns for the investments X and Y to get estimates for σ_X^2, σ_Y^2, and σ_{XY}, then we can get the estimate value of $\hat{\alpha}$ using equations above.



The results are ranging from 0.532 to 0.657.

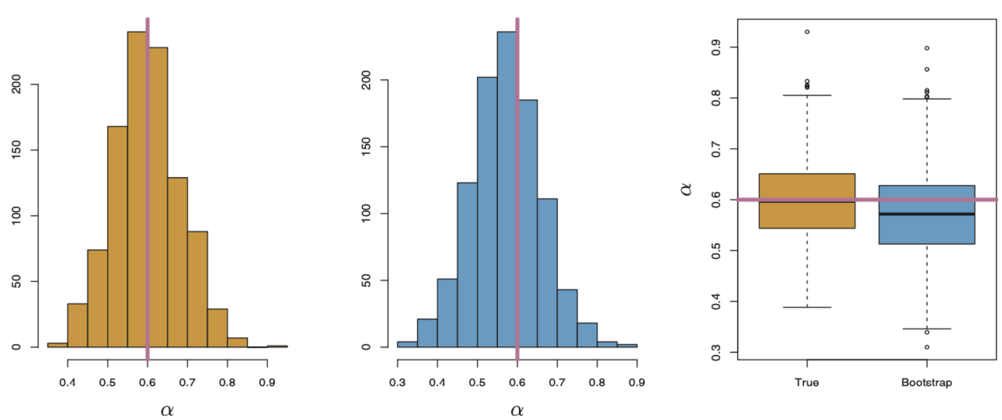
We can do the bootstrap 1000 times, obtaining 1000 estimates for α. Then we can compute the mean and the standard deviation:

$$\bar{\alpha} = \frac{1}{1,000} \sum_{r=1}^{1,000} \hat{\alpha}_r = 0.5996,$$

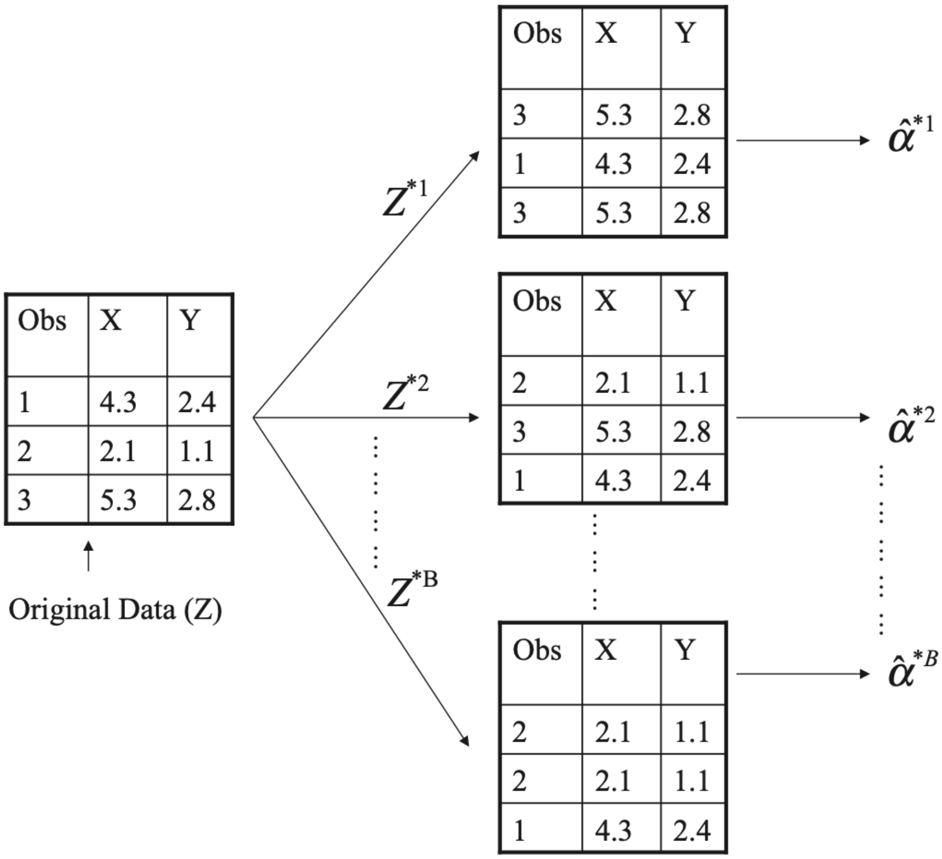
and

$$\sqrt{\frac{1}{1,000-1} \sum_{r=1}^{1,000} \left(\hat{\alpha}_r - \bar{\alpha}\right)^2} = 0.083.$$

We can see that the result we obtained from bootstrap is very close to the true value.



The rationale of this approach is that rather than repeatedly obtaining independent data sets from the population, we instead obtain distinct data sets by repeatedly sampling observations from the original data set. This can be shown as below:



The sampling is performed with **replacement**, which means that the same observation can occur more than once in the bootstrap data set. It is repeated B times in order to produce the bootstrap data set. We can then compute the standard error of these bootstrap estimates using the formula:

$$\mathrm{SE}_B(\hat{\alpha}) = \sqrt{\frac{1}{B-1}\sum_{r=1}^{B}\left(\hat{\alpha}^{*r} - \frac{1}{B}\sum_{r'=1}^{B}\hat{\alpha}^{*r'}\right)^2}.$$