Bootstrapping and Confidence Intervals

Bookmark this page

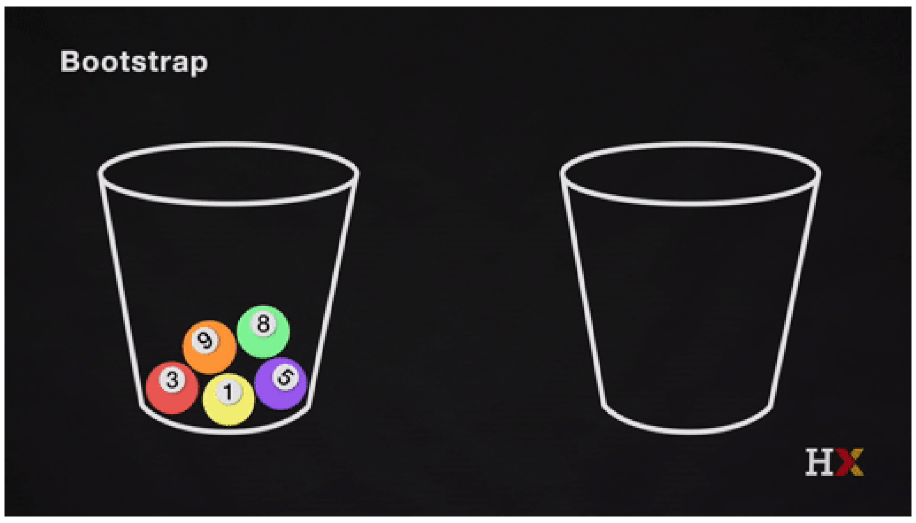
Producing Alternative Data Sets

In the absence of active imagination, magic, parallel universes, and the like, we need an alternative way of producing fake data sets that resemble parallel universes.

Bootstrapping is the practice of sampling observed data (X, Y) to estimate statistical properties.

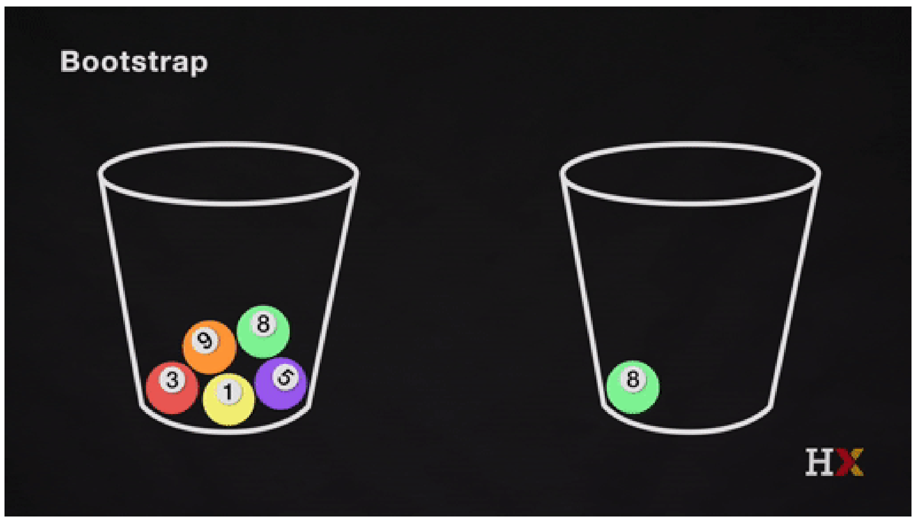
Bootstrap Example

We first randomly pick a ball and replicate it. We then move the replicated ball to another bucket.

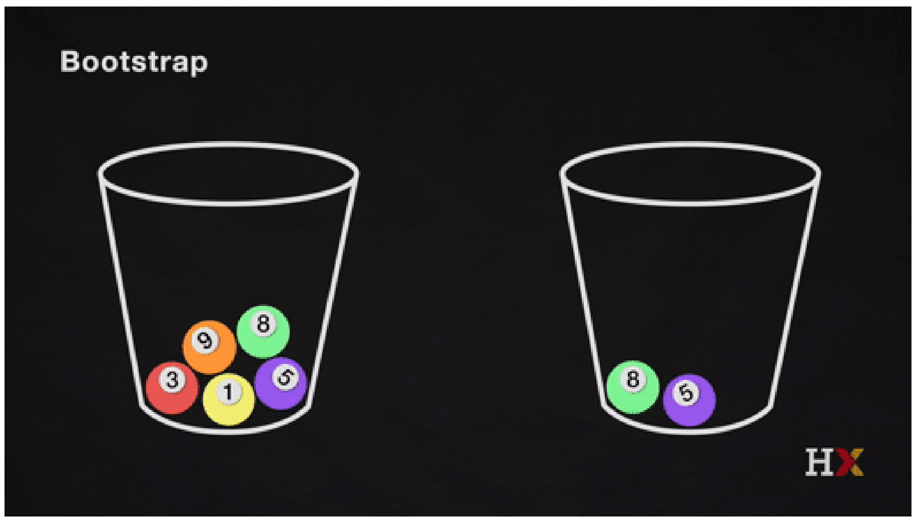


This is called **sampling with replacement**.

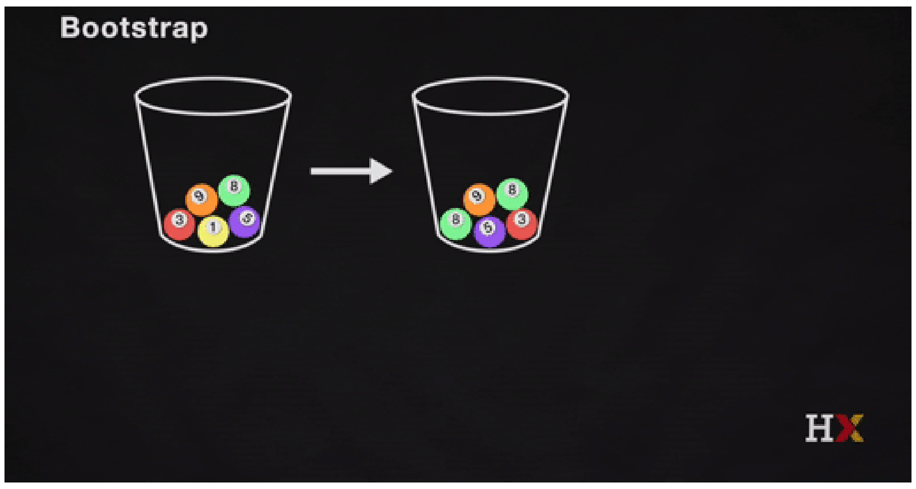
We then randomly pick another ball and replicate it again. As before, we move the replicated ball to the other bucket.



We repeat this process. We continue until the 'other' bucket has the **same number of balls** as the original.

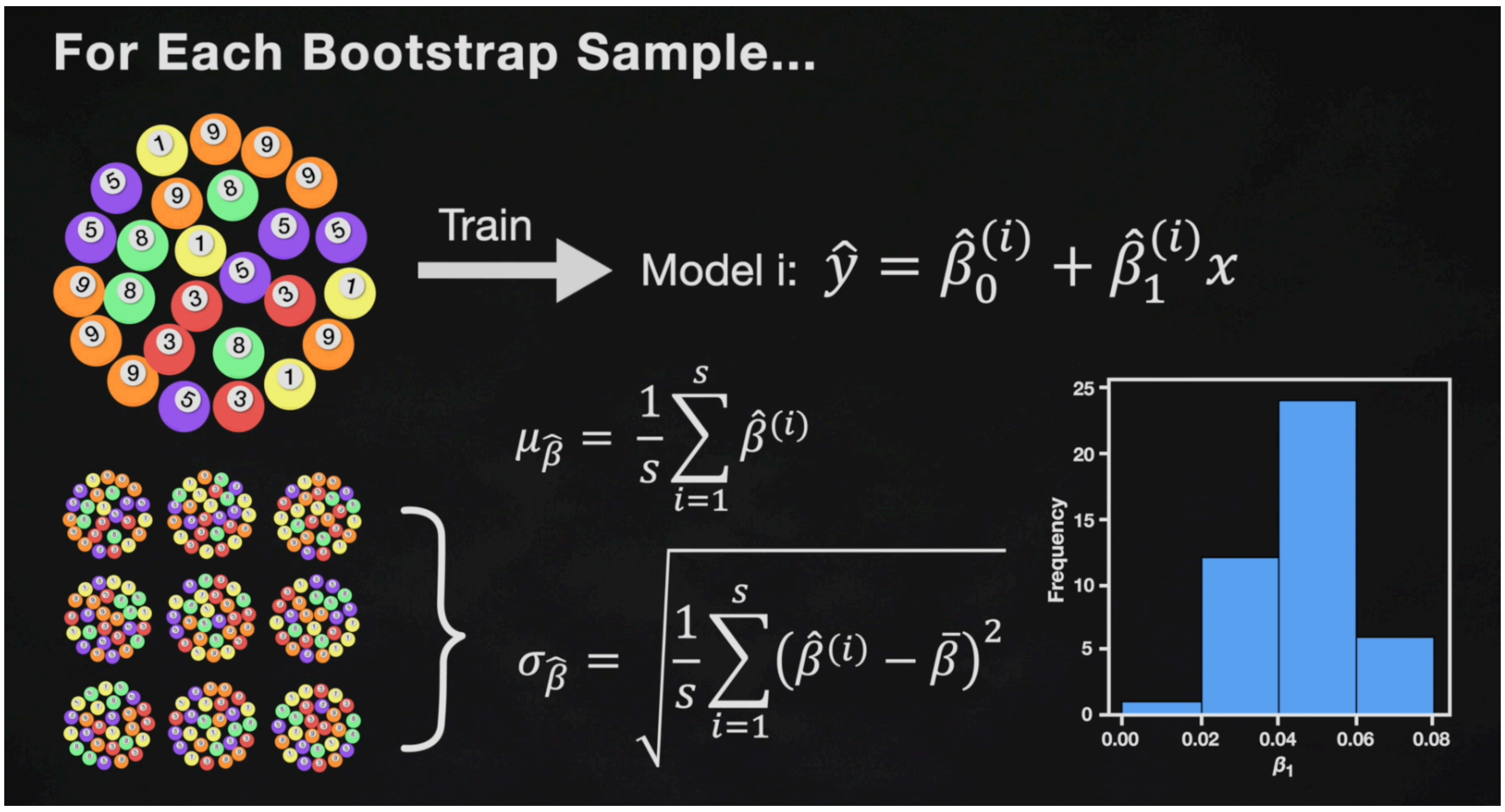


We repeat the same process and acquire many sets of bootstrapped observations.



Bootstrapping for Estimating Sampling Error

Bootstrapping is the practice of estimating the properties of an estimator by measuring those properties by, for example, sampling from the observed data. For example, we can compute $\hat{\beta}_0$ and $\hat{\beta}_1$ multiple times by randomly sampling our data set. We then use the variance of our multiple estimates to approximate the true variance of $\hat{\beta}_0$ and $\hat{\beta}_1$.



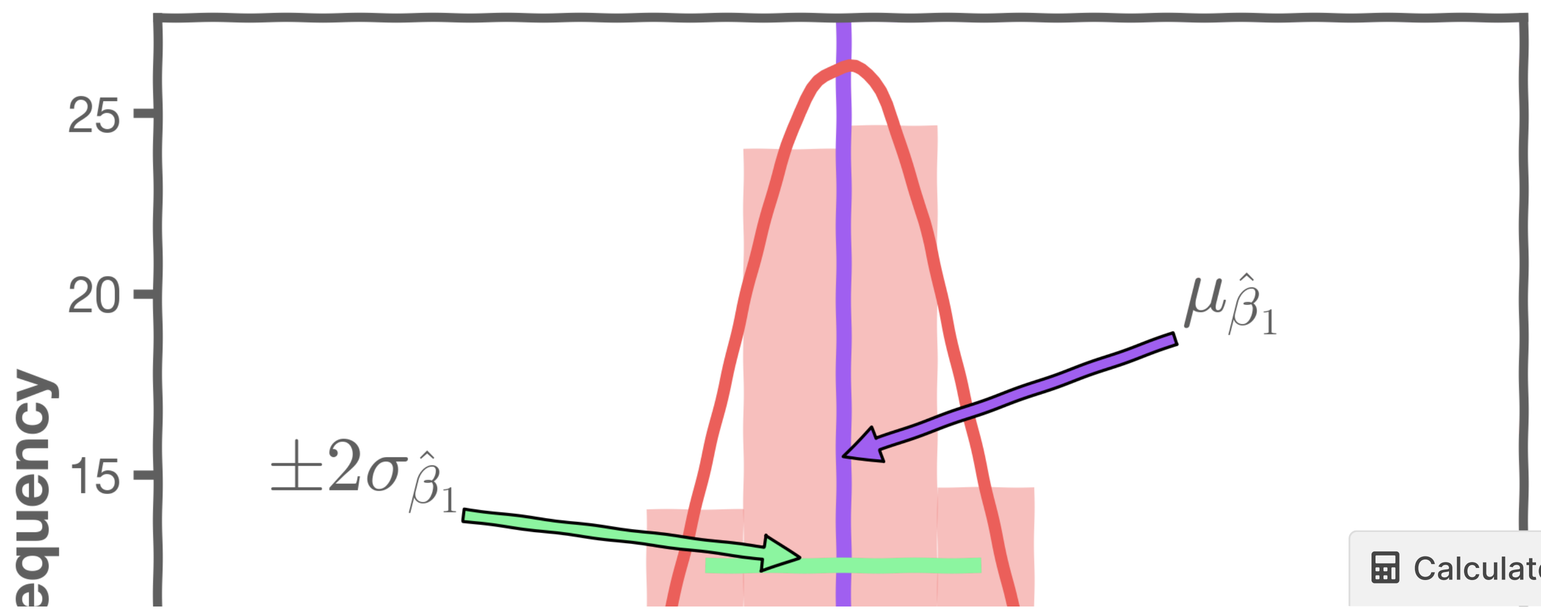
We can now estimate the mean and standard deviation of the estimates $\hat{\beta}_0$ and $\hat{\beta}_1$.

$$\hat{y} = \hat{\beta}_0^{(i)} + \hat{\beta}_1^{(i)} x$$

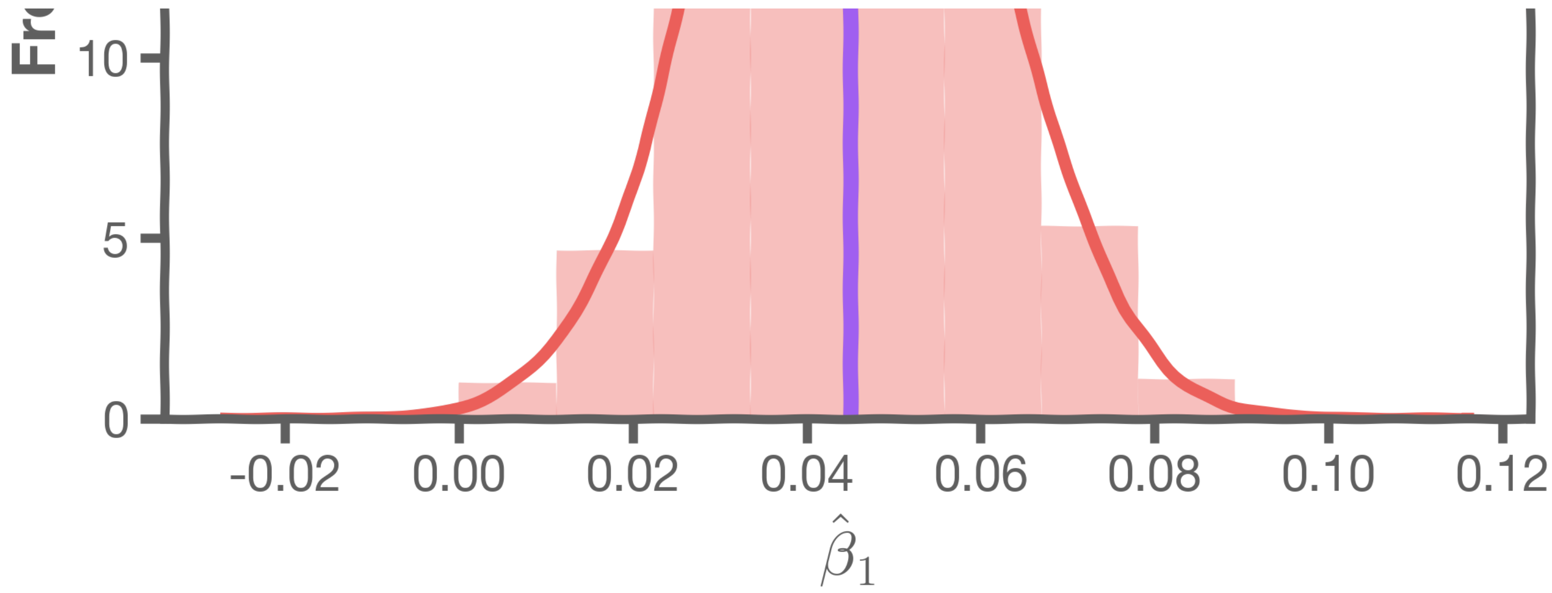
$$\mu_{\hat{\beta}} = \frac{1}{s}\sum_{i=1}^{s} \hat{\beta}^{(i)}$$

$$\sigma_{\hat{\beta}} = \sqrt{\frac{1}{s}\sum_{i=1}^{s} \left(\hat{\beta}^{(i)} - \overline{\beta}\right)}$$

The standard errors give us a sense of our uncertainty over our estimates. Typically, we express this uncertainty as a 95% confidence interval, which is the range of values such that the true value of β_1 is contained in this interval with 95% percent probability.



Calculator



If we assume normality, then:

$$CI_{\hat{\beta}}\,(95\%) \;=\; \left(\hat{\beta} \,-\, 2\sigma_{\hat{\beta}},\, \hat{\beta} \,+ 2\sigma_{\hat{\beta}}\right)$$

Discussion Board (External resource)

Haga clic en Aceptar para que su nombre de usuario y dirección de correo electrónico se envíen a una aplicación de terceros.

Aceptar

Calculator