16) Bootstrap

Vitor Kamada

February 2018

Reference

Tables, Graphics, and Figures from

An Introduction to Statistical Learning

James et al. (2017): Chapters: 5.2, 5.3.4

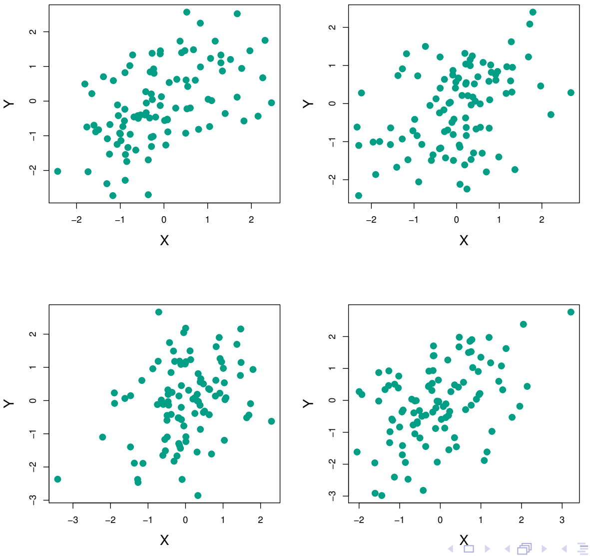
100 Simulated Returns for Investments

$$Var[\alpha X + (1-\alpha)Y]$$

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

$$\hat{\alpha} \in [53\% \text{ to } 65\%]$$

57.6%, 53.2%, 65.7%, and 65.1%



1000 Estimates for α

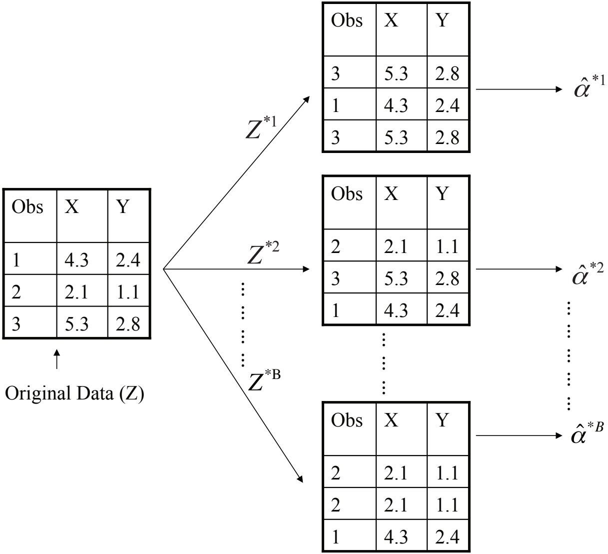
$$\sigma_X^2 = 1,\ \sigma_Y^2 = 1.25,\ \sigma_{XY} = 0.5$$

$$\therefore \alpha = 0.6$$

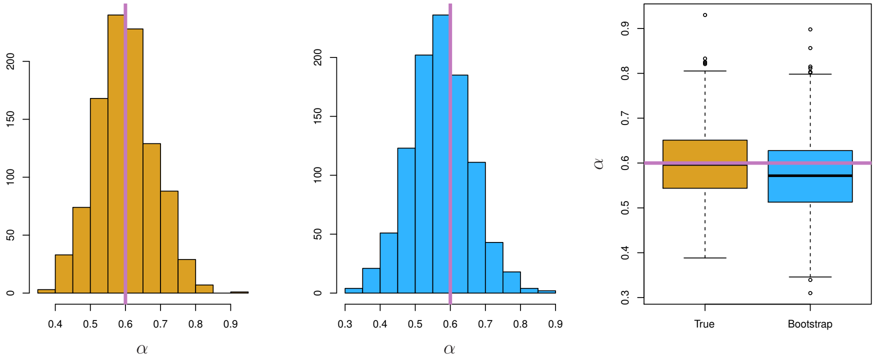
$$\bar{\alpha} = \frac{1}{1000} \sum_{r=1}^{1000} \hat{\alpha}_r = 0.5996$$

$$SE(\hat{\alpha}) = \sqrt{\frac{1}{999} \sum_{r=1}^{1000} (\hat{\alpha}_r - \bar{\alpha})^2} = 0.083$$

Bootstrap Approach (Sampling from Data)



1000 Simulated Data Sets from the True Population vs 1000 Bootstrap Samples from a Single Data Set



1000 Simulated Data Sets

$$SE(\hat{\alpha}) = 0.087$$

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

$$SE_B(\hat{\alpha}) = 0.083$$

library(ISLR)

```
alpha.fn=function(data,index){
 X=data$X[index]
 Y=data$Y[index]
 return((var(Y)-cov(X,Y))/(var(X)
 +var(Y)-2*cov(X,Y)))
}
alpha.fn(Portfolio,1:100)
```

0.576

alpha.fn(Portfolio,sample(100,100, replace=T))

0.596

boot(Portfolio,alpha.fn,R=1000)

```
Bootstrap Statistics :
     original       bias     std. error
t1* 0.5758321 0.002705445  0.09197062
```

Bootstrap and OLS

boot.fn=function(data,index)

return(coef(lm(mpg~horsepower,data=data,subset=index)))

set.seed(1); boot(Auto,boot.fn,1000)

```
Bootstrap Statistics :
       original          bias     std. error
t1* 39.9358610  0.0269563085 0.859851825
t2* -0.1578447 -0.0002906457 0.007402954
```

summary(lm(mpg~horsepower,data=Auto))$coef

```
              Estimate  Std. Error   t value      Pr(>|t|)
(Intercept) 39.9358610 0.717498656  55.65984 1.220362e-187
horsepower  -0.1578447 0.006445501 -24.48914  7.031989e-81
```

set.seed(1)

boot.fn=function(data,index)

coefficients(lm(mpg~horsepower+I(horsepower^2),
data=data,subset=index))

boot(Auto,boot.fn,1000)

```
        original          bias     std. error
t1*  56.900099702  6.098115e-03 2.0944855842
t2*  -0.466189630 -1.777108e-04 0.0334123802
t3*   0.001230536  1.324315e-06 0.0001208339
```

summary(lm(mpg~horsepower+I(horsepower^2),data=Auto))$coef

```
                    Estimate   Std. Error   t value      Pr(>|t|)
(Intercept)     56.900099702 1.8004268063  31.60367 1.740911e-109
horsepower      -0.466189630 0.0311246171 -14.97816  2.289429e-40
I(horsepower^2)  0.001230536 0.0001220759  10.08009  2.196340e-21
```