

# Bootstrap

Wednesday, September 22, 2021 6:07 PM

- Overview

• Resampling with Replacement!

\* Original data is IID  $\prod_{i=1}^n P(x_i)$

- Population Dist

- Sample Dist

- Statistics  $\Rightarrow$  Measuring fun Population using Sample

Ex.

Observation	$x_i$	$y_i$
1	4.1	10
2	4.2	12
3	-3.9	-3
$\vdots$	$\vdots$	$\vdots$
$n$	-2.1	-4

  
 $D = \begin{matrix} \text{Observation} & x_i & y_i \\ 1 & 4.1 & 10 \\ 2 & 4.2 & 12 \\ 3 & -3.9 & -3 \\ \vdots & \vdots & \vdots \\ n & -2.1 & -4 \end{matrix}$ 
  
 $x_i \rightarrow$ 
  
 $D^{*1} = \begin{matrix} 3 & -3.9 & -3 \\ 2 & 4.2 & 12 \\ 1 & 4.1 & 10 \\ \vdots & \vdots & \vdots \end{matrix}$ 
  
 $D^{*2} = \begin{matrix} 2 & 4.2 & 12 \\ 2 & 4.2 & 12 \\ 1 & 4.1 & 10 \end{matrix}$ 
  
 $D^{*B} = \begin{matrix} 3 & -3.9 & -3 \\ 1 & 4.1 & 10 \\ 3 & -3.9 & -3 \end{matrix}$ 

• Point no. Distributional Estimator

$\hookrightarrow E[e_{00s}] \leftarrow$  Test-Train Split

**Point**

$\hat{\mu}$  single number

$\hookrightarrow E[e_{00s}] \leftarrow$  permutations from k-fold CV

**Dist**

$\hat{\mu}$  list of MSE or error from each CV fold


—

$x \sim$

Ex.  $n=100$   $p=55$   $x_1 \dots x_p$

$x_1$   
 $\vdots$   
 $x_{100}$

$x_i$



$k=5$  fold CV

for each fold

training set is 80  $x_i$

test set is 20  $x_i$

MSE

$R^2$

$\vdots$

MSE:  $MSE_1, MSE_2, \dots, MSE_{k=5}$

• Bootstrap statistic

- Metric to calculate:  $\hat{\alpha}$

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

$B \rightarrow$  # of bootstrap resampling instances

Ex.  $\hat{\alpha}$  is  $\hat{\mu}$

for each  $D^{*r} \rightarrow$  resample of  $D$

calculate  $\hat{\mu}^{*r}$

$\rightarrow$  calculate average  $(E[\hat{\mu}])$

of  $\hat{\mu}^{*r}$  across all  $D^{*r}$