

# The Bootstrap

confidence interval for regression coefficient

$$(-\alpha) \times 100\% \quad CI = \hat{\beta} \pm t(n-k, 1-\frac{\alpha}{2}) \cdot SE(\hat{\beta})$$

$$95\% \quad CI = \hat{\beta} \pm 2 \cdot SE(\hat{\beta})$$

Assuming iid Normal errors  $\varepsilon_i$ :

confidence interval for binomial proportion

$$\text{Wald CI} = \hat{p} \pm 2 \sqrt{\hat{p}(1-\hat{p})/n}$$

$$\text{Agrestic CI} = \hat{p}' \pm 2 \sqrt{\hat{p}'(1-\hat{p}')/(n+4)}, \quad \hat{p}' = \frac{s_n + 2}{n+4}$$

Assuming the binomial sum is approximately Normal.

Both of these CI assume Normality.

What if we don't want to assume normality?  
Or what if it's too complicated to analytically  
compute CI?

The Bootstrap is a technique to obtain standard errors or confidence intervals without assuming a distribution (i.e., it's a nonparametric technique to quantify uncertainty) by Resampling data.

## bootstrap estimated binomial proportion:

input  $T$  = training dataset

for  $b = 1, \dots, B$ :

- $T^{(b)}$  = Re-sample  $m$  observations  $X_i$  with replacement

- $\hat{P}^{(b)}$  = Estimated binomial proportion from  $T^{(b)}$

Sort the  $\hat{P}$ 's so that  $\hat{P}^{(1)} \leq \dots \leq \hat{P}^{(B)}$

Then  $SE(\hat{P}) = \text{sd}(\hat{P}^{(1)}, \dots, \hat{P}^{(B)})$ , 95% CI =  $[\hat{P}^{(2.5^{\text{th}} \text{ quantile})}, \hat{P}^{(97.5^{\text{th}} \text{ quantile})}]$

## bootstrap a regression parameter:

input  $T$  = training dataset

for  $b = 1, \dots, B$ :

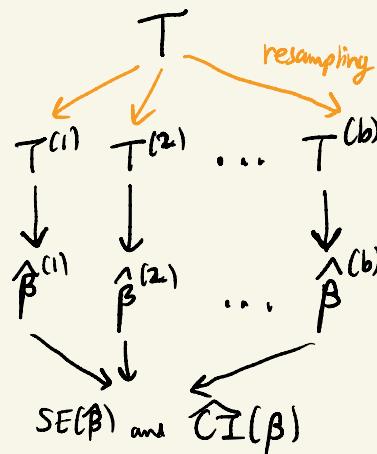
- $T^{(b)}$  = Re-sample  $m$  observations  $(X_i, Y_i)$  with replacement

- fit the Regression from  $T^{(b)}$  and record the parameter estimate  $\hat{\beta}^{(b)}$

Sort the  $\hat{\beta}$ 's so that  $\hat{\beta}^{(1)} \leq \dots \leq \hat{\beta}^{(B)}$

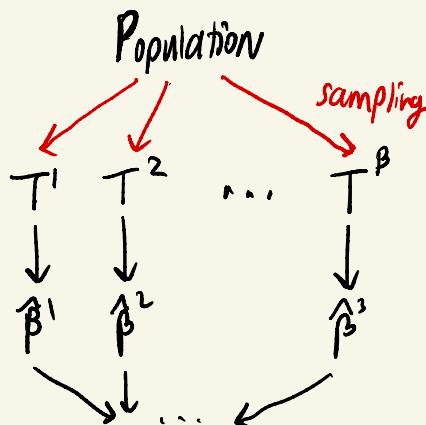
Then  $SE(\hat{\beta}) = \text{sd}(\hat{\beta}^{(1)}, \dots, \hat{\beta}^{(B)})$ , 95% CI =  $[\hat{\beta}^{(2.5^{\text{th}} \text{ quantile})}, \hat{\beta}^{(97.5^{\text{th}} \text{ quantile})}]$

# Visualizing the bootstrap resampling scheme:



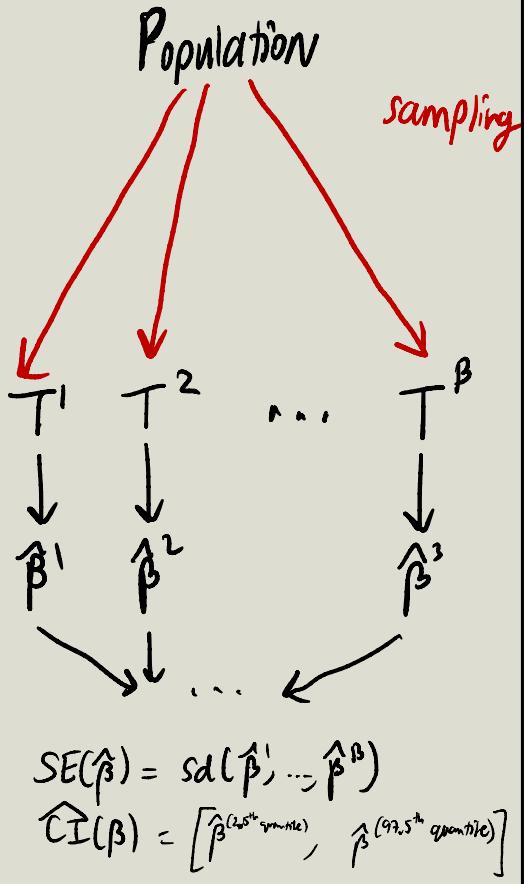
Why does this resampling scheme work?

How we would ideally obtain  
Standard Errors and Confidence Intervals  
(which is impossible):

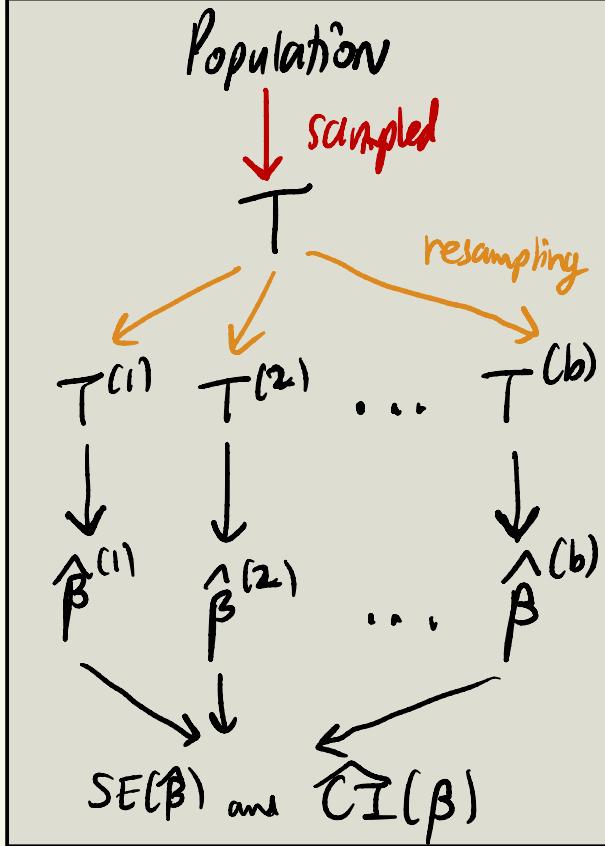


$$SE(\hat{\beta}) = \text{sd}(\hat{\beta}^1, \dots, \hat{\beta}^b), \quad CI(\hat{\beta}) = [\hat{\beta}^{(2.5^{\text{th}} \text{ quantile})}, \hat{\beta}^{(97.5^{\text{th}} \text{ quantile})}]$$

Ideal uncertainty quantification:



Bootstrapped uncertainty quantification:



Bootstrapping works because the training dataset  $T$  is itself a sample from the population; thus resampling from  $T$  mimics sampling from the population!

# Drawbacks of the bootstrap

- doesn't always work
  - e.g. works well for SE of means but not extrema (like max or min)
- computationally intensive
- can underestimate uncertainty:
  - CI often not wide enough
  - (see Brill "Analytics Have Some Humility")
    - ↳ can calibrate the bootstrap to achieve desired coverage

## M&M Demo:

You have an M&M bag with  $N$  M&M's and  $S$  blue ones. Supposing the color of each M&M is randomly drawn from some iid distribution, what is a 95% confidence interval of the probability the company makes an M&M blue? Use the bootstrap and compare to the Wald and Agresti intervals.

Class field trip to WaWa to buy M&Ms.  
Grab a bag of M&M's and do it!  
Does it work?