Statistical Computing with R – MATH 6382[1,*]
Set 7 (Resampling: Bootstrap and Jackknife)

Tamer Oraby
UTRGV
tamer.oraby@utrgv.edu

[1] Based on textbook.

[*] Last updated November 13, 2016

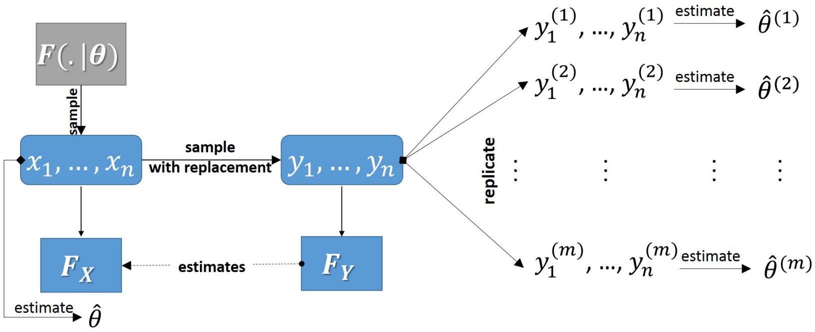
Estimation using Bootstrap

Bootstrap

By Efron (1979, 1981), to estimate

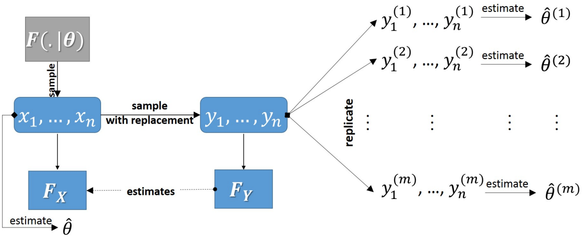
- Bias
- Standard error
- Confidence interval (5 different ways)

Bootstrap



$m = 50$ is good enough.

Bootstrap



Let $\bar{\hat{\theta}}_B = \frac{1}{m}\sum_{i=1}^{m} \hat{\theta}^{(i)}$

- $\hat{Bias} = \bar{\hat{\theta}}_B - \hat{\theta}$ where $\hat{\theta}$ is the estimate of θ using the original sample $x_1, \ldots, x_n$
- Standard error $se(\hat{\theta}_B)$ is the standard deviation of the bootstrap estimates $\hat{\theta}^{(1)}, \ldots, \hat{\theta}^{(m)}$

Bootstrap

$100(1-\alpha)\%$ Bootstrap Confidence Intervals (BCI):

1. Standard Normal BCI

$$\hat{\theta} \pm z_{\alpha/2} se(\hat{\theta})$$

2. Basic BCI

$$\left(2\hat{\theta} - \hat{\theta}_{1-\alpha/2}, 2\hat{\theta} - \hat{\theta}_{\alpha/2}\right)$$

3. Percentile BCI

$$\left(\hat{\theta}_{\alpha/2}, \hat{\theta}_{1-\alpha/2}\right)$$

4. t-type – BCI

$$\left(\hat{\theta} - t^*_{1-\alpha/2} se(\hat{\theta}), \hat{\theta} + t^*_{\alpha/2} se(\hat{\theta})\right)$$

Bootstrap

with t^*_α is the α quantile of $\{t^{(1)}, \dots, t^{(m)}\}$ where $t^{(i)} = \dfrac{\hat{\theta}^{(i)} - \hat{\theta}}{se(\hat{\theta}^{(i)})}$ and estimation of $se(\hat{\theta}^{(i)})$ requires a further bootstrap from the bootstrapped sample $y_1^{(i)}, \dots, y_n^{(i)}$

Bootstrap

$100(1-\alpha)\%$ Bootstrap Confidence Intervals (BCI):

1. Bias Corrected accelerated BCI or BCa – BCI

$$\left(\hat{\theta}^*_{\alpha_1}, \hat{\theta}^*_{\alpha_2}\right)$$

are the α_1 and α_2 quantiles of $\{\hat{\theta}^{(1)}, \ldots, \hat{\theta}^{(m)}\}$ and

$$\alpha_1 = \Phi(\hat{z}_0 + \frac{\hat{z}_0 + z_{\alpha/2}}{1 - \hat{a}(\hat{z}_0 + z_{\alpha/2})})$$

and

$$\alpha_2 = \Phi(\hat{z}_0 + \frac{\hat{z}_0 + z_{1-\alpha/2}}{1 - \hat{a}(\hat{z}_0 + z_{1-\alpha/2})})$$

Bootstrap

where Φ is the cdf of the standard normal, $z_{\alpha/2}$ is the standard normal quantile, and the bias corrector

$$\hat{z}_0 = \Phi^{-1}\left(\frac{1}{m}\sum_{i=1}^{m} I(\hat{\theta}^{(i)} \leq \hat{\theta})\right)$$

where I is the indicator function, and the acceleration factor

$$\hat{a} = \frac{\sum_{i=1}^{m}(\hat{\theta}^{(i)} - \hat{\theta})^3}{6\left(\sum_{i=1}^{m}(\hat{\theta}^{(i)} - \hat{\theta})^2\right)^{3/2}}$$

which measures skewness.

Estimation using Jackknife

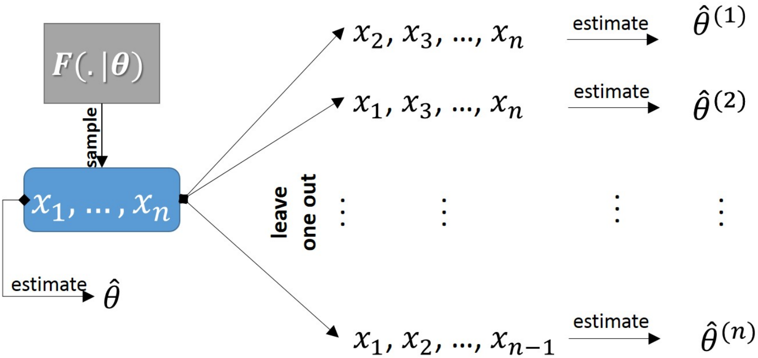
Jackknife

By Quenouille and Tukey, to estimate

- Bias
- Standard error

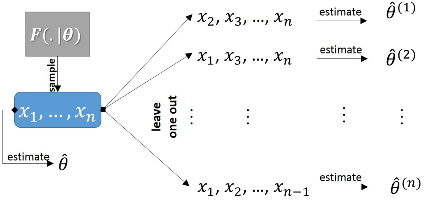
The estimate $\hat{\theta}$ must be a smooth plug-in estimator: small changes in the data results in small changes in the value of the estimate. The sample mean is a smooth plug-in for the population mean while the sample median is not.

Jackknife



$F(.|\theta)$
sample
$x_1, \ldots, x_n$
estimate
$\hat{\theta}$
leave one out
$x_2, x_3, \ldots, x_n$
estimate
$\hat{\theta}^{(1)}$
$x_1, x_3, \ldots, x_n$
estimate
$\hat{\theta}^{(2)}$
$x_1, x_2, \ldots, x_{n-1}$
estimate
$\hat{\theta}^{(n)}$

Jackknife



Let $\bar{\hat{\theta}}_J = \frac{1}{n}\sum_{i=1}^{n} \hat{\theta}^{(i)}$

- $\hat{Bias} = (n-1)(\bar{\hat{\theta}}_J - \hat{\theta})$ where $\hat{\theta}$ is the estimate of θ using the original sample $x_1, \dots, x_n$
- Standard error $se(\hat{\theta}_J)$ is $\sqrt{n-1}$ times the standard deviation of the jackknife estimates $\hat{\theta}^{(1)}, \dots, \hat{\theta}^{(n)}$

End of Set 7