Jackknife Resampling

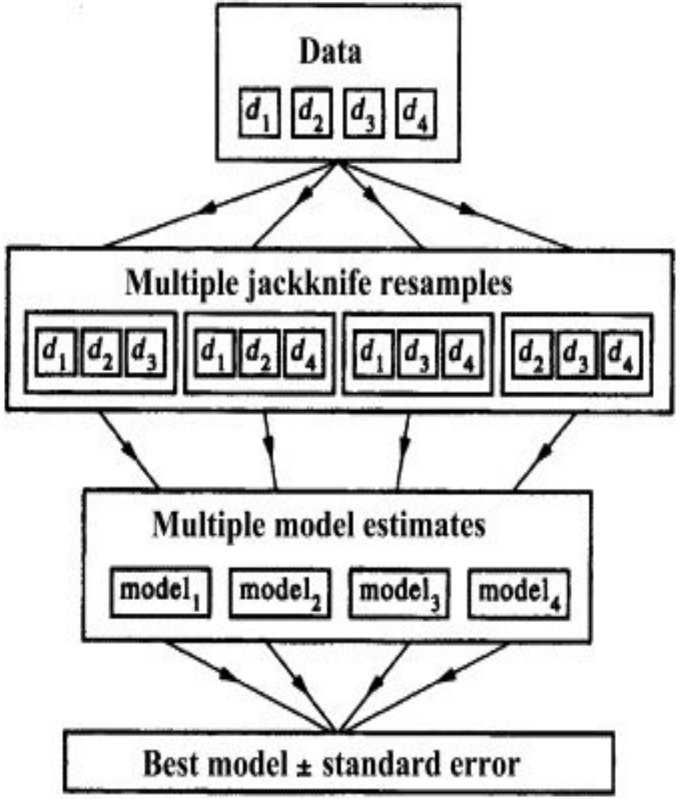
Deepthi Gopinath, Thalia Huynh, Tianyi Jiang, Michelle Ha

STAT 641 Spring ‘21

History

The Jackknife method, invented by Quenouille (1949), is an alternative resampling method to the bootstrap which is useful for variance and bias estimation. It is a linear approximation of the bootstrap method. This technique was refined in 1956 and expanded later by John Tukey in 1958. The Jackknife and bootstrap method both provide a relatively easy way to estimate the precision of an estimator, θ. However, the Jackknife is generally less computationally intensive than the bootstrap.

How It Works



The method is based upon sequentially deleting one observation from the dataset, recomputing the estimator, here , n times. That is, there are exactly n jackknife estimates obtained in a sample of size n-1.

$$\widehat{se}_{jack} = \sqrt{\frac{n-1}{n}\sum_{i=1}^{n}(\hat{\theta}_{(i)} - \bar{\hat{\theta}}_{(.)})^2}$$

Bias-corrected Jackknife Estimator

- To see the relationship with jackknife estimation, we can write the mean with observation removed as $\overline{X}_{(i)}$ as:

$$\overline{X}_{(i)} = \frac{\left(\sum_{j=1}^{n} X_j\right) - X_i}{n-1}$$

- Therefore an individual X_i can be written as

$$X_i = \left(\sum_{j=1}^{n} X_j\right) - (n-1)\overline{X}_{(i)} = n\overline{X} - (n-1)\overline{X}_{(i)}. \tag{4}$$

- We now apply this approach to statistics other than the sample mean. For example, when estimating a parameter θ with estimator $\widehat{\theta}$, if we replace the sample means in (4) with the corresponding estimators of θ, we have

$$\mathrm{PV}(\mathbf{x}_{(i)}) = n\widehat{\theta} - (n-1)\widehat{\theta}_{(i)}$$

- $\mathrm{PV}(\mathbf{x}_{(i)})$ is called the i^{th} **pseudo-value**.

$$\widehat{\mathrm{bias}}(\widehat{\theta}) = (n-1)(\widehat{\theta}_{(\cdot)} - \widehat{\theta}) \tag{3}$$

- The **bias-corrected jackknife estimate** of θ is

$$\widehat{\theta}_{\text{jack}} = \widehat{\theta} - \widehat{\mathrm{bias}}(\widehat{\theta}) = n\widehat{\theta} - (n-1)\widehat{\theta}_{(\cdot)}$$

Demonstration in R (Mean)

library(bootstrap)

x <- c(3.56, 0.69, 0.10, 1.84, 3.93, 1.25, 0.18, 1.13, 0.27, 0.50,

0.67, 0.01, 0.61, 0.82, 1.70, 0.39, 0.11, 1.20, 1.21, 0.72)

 # sample mean

mean(x)

 # jackknife mean

jack_mean <- jackknife(x,mean)

jack_mean

 # bias-corrected jackknife estimate

bc_jmean <- mean(x) - jack_mean$jack.bias

bc_jack

```
[1] 1.0445 # sample mean
$jack.se
[1] 0.2369516

$jack.bias
[1] 0

$jack.values
 [1] 0.9121053 1.0631579 1.0942105
 [4] 1.0026316 0.8926316 1.0336842
 [7] 1.0900000 1.0400000 1.0852632
[10] 1.0731579 1.0642105 1.0989474
[13] 1.0673684 1.0563158 1.0100000
[16] 1.0789474 1.0936842 1.0363158
[19] 1.0357895 1.0615789

$call
jackknife(x = x, theta = mean)

[1] 1.0445 # bias-corrected jackknife estimate
```

Demonstration in R (Standard Deviation)

```
library(bootstrap)

x <- c(3.56, 0.69, 0.10, 1.84, 3.93, 1.25, 0.18, 1.13,
0.27, 0.50, 0.67, 0.01, 0.61, 0.82, 1.70, 0.39, 0.11,
1.20, 1.21, 0.72)

# sample standard deviation

sd(x)

# jackknife standard deviation

jack_sd <- jackknife(x,sd)

jack_sd

# bias-corrected jackknife estimate

bc_jsd <- sd(x) - jack_sd$jack.bias

bc_jsd
```

```
[1] 1.05968 # sample standard deviation
$jack.se
[1] 0.2802791

$jack.bias
[1] -0.03710029

$jack.values
 [1] 0.9029186 1.0853369 1.0644890
 [4] 1.0715868 0.8357022 1.0875825
 [7] 1.0684568 1.0885209 1.0724860
[10] 1.0807253 1.0849440 1.0595853
[13] 1.0836350 1.0873628 1.0771155
[16] 1.0771511 1.0650050 1.0880677
[19] 1.0879814 1.0858855

$call
jackknife(x = x, theta = sd)

[1] 1.09678 # bias-corrected jackknife estimate
```

Jackknife vs Bootstrap

Advantages:

- Useful method for estimating and compensating for bias in an estimator.
- Generally less computationally intensive compared to the bootstrap.
- More orderly

Disadvantages:

- Doesn’t perform as well as Bootstrap in most cases
- More conservative than the bootstrap method
- Performs poorly when the the estimator is not sufficiently smooth
- Require observations to be independent

When to use which:

- The bootstrap handles skewed distribution better
- The Jackknife method works for smaller original data samples

Citation

https://math.montana.edu/jobo/thainp/jack.pdf

https://medium.com/@lymielynn/bootstrapping-vs-jackknife-d5172965207b

https://si.biostat.washington.edu/sites/default/files/modules/2017_sisg_1_9_v3_0.pdf