Resample Methods and Confidence intervals

Travis Xie

August 26, 2021

Abstract

This document is the reference for the project of evaluating resampling methods in population genetics. In the project, we try to implement difference resampling methods and confidence interval methods.

1 Resample methods

In the project, we use four different resampling methods: bootstrap, delete-one jackknife, delete-m jackknife (block jackknife with equal size), delete-mj jackknife (block jackknife with unequal sizes).

1.1 Bootstrap

Chapter 5 Reference: Efron, B., amp; Tibshirani, R. (1993). The bootstrap estimate of standard error. In An introduction to the bootstrap (pp. 45–56). essay, Chapman and Hall.

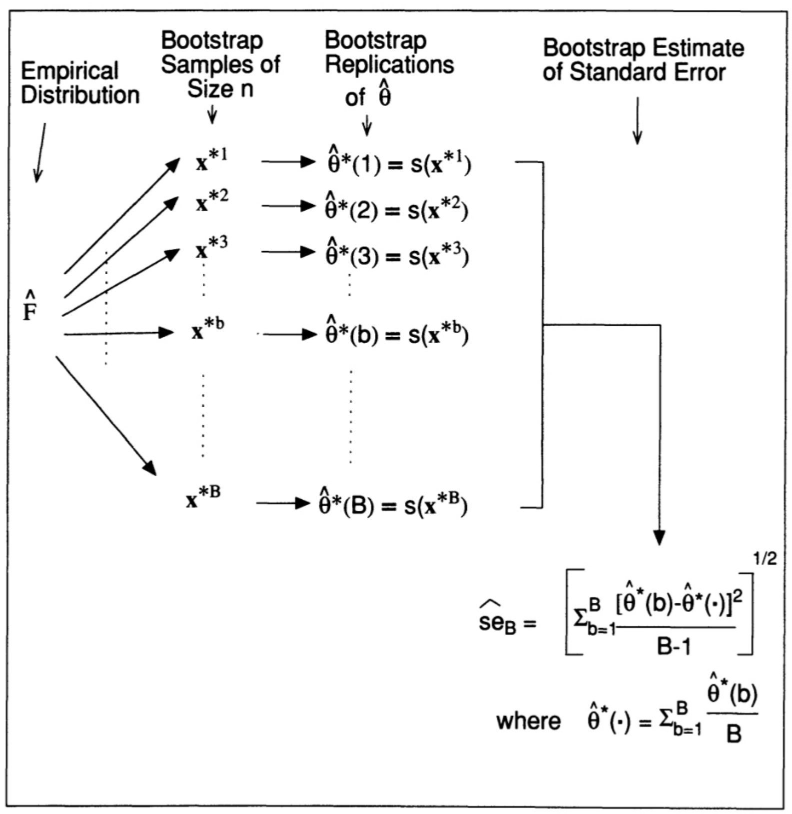

Figure 1: Bootstrap

1.1.1 Standard normal interval

See equation 13.2 in the book

$$\hat{\theta}_{lo} = \hat{\theta} - z^{1-\alpha} \cdot \hat{s}e_B \quad (1)$$

$$\hat{\theta}_{hi} = \hat{\theta} + z^{\alpha} \cdot \hat{s}e_B \quad (2)$$

```
def normal(bt, confidence, obs_value):
    '''
    bt: bootstrap values
    confidence: confidence level
    obs_value: estimate value \hat\theta
    '''

    qnorm = abs(norm.ppf((1 - confidence) / 2))

    upper = obs_value + qnorm * np.std(bt, ddof=1) # change degree of freedom
    lower = obs_value - qnorm * np.std(bt, ddof=1)

    return lower, upper
```

1.1.2 The percentile interval

See equation 13.5 in the book

$$[\hat{\theta}_{\%,lo}, \hat{\theta}_{\%,hi}] \approx [\hat{\theta}_B^{*(\alpha)}, \hat{\theta}_B^{*(1-\alpha)}] \quad (3)$$

```
def quantile(bt, confidence):
    """
    bt: bootstrap values
    confidence: confidence level
    """

    cutoff = (1 - confidence) / 2
    lower = np.quantile(bt, cutoff)
    upper = np.quantile(bt, 1 - cutoff)

    return lower, upper
```

1.1.3 Bias corrected Normal CI

See lecture slides as well as chapter 14 for details. Did not find the exact match of this method in the book.

```
def bias_corrected_normal(bt, confidence, obs_value):
    """
    bt: bootstrap values
    confidence: confidence lvel
    obs_value: estimate value \hat\theta
    """

    # caculate the bias between the resampled mean and the observed sample mean
    mean = np.mean(bt)
    bias = mean - obs_value

    # correct the observed sample mean
    obs_value -= bias
```

```
qnorm = abs(norm.ppf((1 - confidence) / 2))

upper = obs_value + qnorm * np.std(bt, ddof=1)
lower = obs_value - qnorm * np.std(bt, ddof=1)

return lower, upper
```

1.1.4 Results

CI methods	lower	upper
normal	0.146454	0.155567
quantile	0.146643	0.155231
bias corrected	0.146472	0.155586

Table 1: Bootstrap confidence intervals

1.2 Jackknife

Details for jackknife see note: jackknife.

1.2.1 Delete-one and Delete-m jackknife

We put delete-one and delete-m jackknife in one section because delete-on jackknife is just a special case of delete-m jackknife where we split the sample into n blocks and each block only contain one observation. There are two methods of calculating the confidence interval: 1) using pseudo values; 2) not using pseudo values.

$$ps_i(x) = n\hat{\theta} - (n-1)\hat{\theta}_i \tag{4}$$

where n is the number of observations in delete-one jackknife and the number of blocks in delete-m jackknife. $\hat{\theta}$ is the estimator of the observations (mean of the sample in our project). $\hat{\theta}_i$ is the estimator of θ with the i-th observation or group removed.

$$ps(X) = \frac{1}{n}\sum_{i=1}^{n} ps_i(X) \tag{5}$$

$$V_{ps}(X) = \frac{1}{n-1}\sum_{i=1}^{n}(ps_i(X) - ps(X))^2 \tag{6}$$

Then the standard normal confidence interval is:

$$\left(ps(X) - z^{1-\alpha}\sqrt{\frac{1}{n}V_{ps}(X)}, ps(X) + z^{\alpha}\sqrt{\frac{1}{n}V_{ps}(X)}\right) \tag{7}$$

We can also estimate the standard error by not using the pseudo values (delete-m jackknife for unequal M paper)

$$\hat{\sigma}^2 = \frac{g-1}{g}\sum_{i=1}^{g}\left(\hat{\theta}_j - \bar{\theta}_m\right) \tag{8}$$

$$\hat{s}e = \sqrt{\hat{\sigma}^2} \tag{9}$$

where $\hat{\theta}_j$ is the estimator of $\hat{\theta}$ with the m observations of group j removed from the sample and $\bar{\theta}_m = \frac{1}{g}\sum_{i=1}^{g}\hat{\theta}_j$. Hence, the standard normal confidence interval becomes:

$$\left(\hat{\theta} - z^{1-\alpha}\hat{s}e, \hat{\theta} + z^{\alpha}\hat{s}e\right) \tag{10}$$

The difference between 7) and 10) is the mean value we use. In 7), we use the mean of pseudo values $ps(X)$as the estimator , while in 10), we use the observed sample mean $\hat{\theta}$ as the estimator. The difference between two intervals are within 1×10^{-6}, so we can use either way to calculate the confidence interval. Both methods can be applied to delete-one jackknife and delete-m with equal sizes jackknife.

1.2.2 Delete-mj Jackknife

However, when it comes to jackknife with unequal sizes (delete-mj), the difference of two CI methods becomes too large to be ignored. In delete-mj, the pseudo values become:

$$ps_j(X) = h_j\hat{\theta} - (h_j - 1)\hat{\theta}_j \tag{11}$$

where $\hat{\theta}_j$ is still the estimator of θ with the m_j observations of group j removed from the sample, $h_j = \frac{n}{m_j}$, n is the total number of observations. We can still calculate the confidence interval by 7) but the standard error is biased because it does not take consideration of unequal sizes in each block. So we will adjust equation 8) such that it that takes into account of unequal sizes.

$$\hat{\theta}_{J(m_j)} = g\hat{\theta} - \sum_{j=1}^{g}\Big(1 - \frac{m_j}{n}\Big)\hat{\theta}_j \tag{12}$$

$$\hat{\sigma}^2_{J(m_j)} = \frac{1}{g}\sum_{j=1}^{g}\frac{1}{h_j - 1}\Big(h_j\hat{\theta} - (h_j - 1)\hat{\theta}_j - \hat{\theta}_{J(m_j)}\Big)^2 \tag{13}$$

$$\hat{se} = \sqrt{\hat{\sigma}^2_{J(m_j)}} \tag{14}$$

g is the number of blocks. With the se, we can apply 10) to find the confidence interval.

1.2.3 Results

CI methods	lower	upper
delete one pseudo	0.146334	0.155687
delete one se	0.146334	0.155687
delete m pseudo	0.146012	0.156009
delete m se	0.146012	0.156009
delete mj pseudo	0.145564	0.155958
delete mj se	0.146334	0.155687

Table 2: Jackknife confidence intervals

Since the standard error methods works very well for all three jackknife resampling methods, I recommend us to use it in our project instead of using pseudo values.

2 Reference

Book: An Introduction to the Bootstrap. Chapter 6: The bootstrap estimate of standard error. Chapter 11: The jackknife. Chapter 13: Confidence intervals based on bootstrap percentiles.
Paper: Delete-m Jackknife for Unequal m.
Leture notes: Resampling Data: Using a Statistical Jackknife. Washington University. Bootstrap Confidence intervals. University of Washington.