Jackknife

One of the earliest techniques to obtain reliable statistical estimators is the jackknife technique. It requires less computational power than more recent techniques.

Suppose we have a sample $x = (x_1, x_2, ..., x_n)$ and an estimator $\hat{\theta} = s(x)$. The jackknife focuses on the samples that *leave out one observation at a time*:

$$x_{(i)} = (x_1, x_2, ..., x_{i-1}, x_{i+1}, ..., x_n)$$

for $i = 1, 2, ..., n$, called *jackknife samples*. The *i*th jackknife sample consists of the data set with the *i*th observation removed. Let $\hat{\theta}_{(i)} = s(x_{(i)})$ be the *i*th jackknife replication of $\hat{\theta}$. The jackknife estimate of standard error defined by

$$\widehat{SE}_{jack} = [\frac{n-1}{n} \sum (\hat{\theta}_{(i)} - \hat{\theta}_{(.)})^2]^{1/2} \tag{3}$$

where $\hat{\theta}_{(.)} = \sum_{i=1}^{n} \hat{\theta}_{(i)} \big/ n$.

The jackknife only works well for linear statistics (e.g., mean). It fails to give accurate estimation for non-smooth (e.g., median) and nonlinear (e.g., correlation coefficient) cases. Thus improvements to this technique were developed.

Delete-*d* jackknife

Instead of leaving out one observation at a time, we leave out *d* observations. Therefore, the size of a delete-*d* jackknife sample is (*n*-*d*), and there are $\binom{n}{d}$ jackknife samples.

Let $\hat{\theta}_{(Z)}$ denote $\hat{\theta}$ applied to the data set with subset *s* removed. The formula for the delete-*d* jackknife estimate of standard error is

$$\{\frac{n-d}{d\binom{n}{d}} \sum (\hat{\theta}_{(Z)} - \hat{\theta}_{(.)})^2\}^{1/2} \tag{4}$$

where $\hat{\theta}_{(.)} = \sum \hat{\theta}_{(Z)} \big/ \binom{n}{d}$ and the sum is over all subsets *s* of size (*n*-*d*) chosen without replacement for $x_1, x_2, ..., x_n$.

It can be shown that the delete-*d* jackknife is consistent for the median if $\sqrt{n}/d \to 0$ and $(n-d) \to \infty$. Roughly speaking, it is preferable to choose a *d* such that $\sqrt{n} < d < n$ for the delete-*d* jackknife estimation of standard error.

Bootstrap

Bootstrap is the most recently developed method to estimate errors and other statistics. It requires the much greater power that modern computers can provide.

Example. Consider a sample $\boldsymbol{x} = (x_1, x_2, ..., x_n)$, in which x_i is drawn from an empirical distribution $\hat{F}$ (or from a population). Samples of size n are drawn from $\boldsymbol{x}$ with replacement. There are n^n possible samples, called the *ideal bootstrap samples*. Consider a simple case when n = 2. The original sample $\boldsymbol{x} = (x_1, x_2)$ yields 2^2 = 4 ideal bootstrap samples: $\boldsymbol{x}^{*1} = (x_1, x_1), \boldsymbol{x}^{*2} = (x_1, x_2), \boldsymbol{x}^{*3} = (x_2, x_1), \boldsymbol{x}^{*4} = (x_2, x_2)$.

However, getting all ideal bootstrap samples becomes unrealistic as n becomes a large number and the computational tasks are incredibly heavy. The bootstrap estimate of standard error is the standard deviation of the bootstrap replications:

$$\widehat{SE}_{boot} = \left\{ \sum_{b=1}^{B} [s(\boldsymbol{x}^{*b}) - s(.)]^2 \big/ (B-1) \right\}^{1/2} \qquad (5)$$

where $s(.) = \sum_{b=1}^{B} s(\boldsymbol{x}^{*b}) \big/ B$.

Comparing (3) with (5), one can find that the factor in the jackknife's standard error. formula is roughly n times larger. This is called the *inflation factor*. The reason is that, unlike bootstrap samples, jackknife samples are very similar to the original sample and therefore the difference between jackknife replications is small. One can consider the special case when $\hat{\theta} = \bar{x}$ and verify (3).

Suppose $s(\boldsymbol{x})$ is the mean $\bar{x}$. The bootstrap algorithm for estimating standard errors:

1. Select B independent bootstrap samples $\boldsymbol{x}^{*1}, \boldsymbol{x}^{*2}, ..., \boldsymbol{x}^{*B}$, each consisting of n data values drawn with replacement from $\mathbf{x}$.
2. Evaluate the bootstrap replication corresponding to each bootstrap sample

$$\hat{\theta}^*(b) = s(\mathbf{x}^{*b}), \quad b = 1, 2, \cdots, B.$$

3. Estimate the standard error by the sample standard deviation of the B replicates

$$\widehat{SE}_B = \left\{ \sum_{b=1}^{B} [\widehat{\theta}^*(b) - \widehat{\theta}^*(.)]^2 \big/ (B-1) \right\}^{1/2}$$

where $\widehat{\theta}^*(.) = \sum_{b=1}^{B} \widehat{\theta}^*(b) \big/ B$.

Other properties of bootstrap:

- Typical values of B, the number of bootstrap samples, are ≥ 200 for standard error estimation.
- Bootstrap methods can also assess more complicated accuracy measures, like biases, prediction errors, and confidence intervals.
- Bootstrap confidence intervals add another factor of 10 to the computational burden.

The payoff for heavy computation:

- An increase in the statistical problems that can be analyzed.
- A reduction in the assumption of the analysis.
- The elimination of the routine but tedious theoretical calculations usually associated with accuracy assessment.