Introduction to resampling methods

- Definitions and Problems
- Non-Parametric Bootstrap
- Parametric Bootstrap
- **Jackknife**
- Permutation tests
- Cross-validation

Introduction

The bootstrap method is not always the best one. One main reason is that the bootstrap samples are generated from $\hat{f}$ and not from f. Can we find samples/resamples exactly generated from f?

- If we look for samples of size n, then the answer is no!
- If we look for samples of size m ($m < n$), then we can indeed find (re)samples of size m exactly generated from f simply by looking at different subsets of our original sample $\mathbf{x}$!

Looking at different subsets of our original sample amounts to sampling without replacement from observations $x_1, \cdots, x_n$ to get (re)samples (now called subsamples) of size m. This leads us to subsampling and the jackknife.

Jackknife

- The jackknife has been proposed by Quenouille in mid 1950's.
- In fact, the jackknife predates the bootstrap.
- The jackknife (with $m = n - 1$) is less computer-intensive than the bootstrap.
- *Jackknife* describes a swiss penknife, easy to carry around. By analogy, Tukey (1958) coined the term in statistics as a general approach for testing hypotheses and calculating confidence intervals.

Jackknife samples

Definition

The Jackknife samples are computed by leaving out one observation x_i from $\mathbf{x} = (x_1, x_2, \cdots, x_n)$ at a time:

$$\mathbf{x}_{(i)} = (x_1, x_2, \cdots, x_{i-1}, x_{i+1}, \cdots, x_n)$$

- The dimension of the jackknife sample $\mathbf{x}_{(i)}$ is $m = n - 1$
- n different Jackknife samples : $\{\mathbf{x}_{(i)}\}_{i=1\cdots n}$.
- No sampling method needed to compute the n jackknife samples.

Available BOOTSTRAP MATLAB TOOLBOX, by Abdelhak M. Zoubir and D. Robert Iskander, http://www.csp.curtin.edu.au/downloads/bootstrap_toolbox.html

Jackknife replications

Definition

The ith jackknife replication $\hat{\theta}_{(i)}$ of the statistic $\hat{\theta} = s(\mathbf{x})$ is:

$$\hat{\theta}_{(i)} = s(\mathbf{x}_{(i)}), \quad \forall i = 1, \cdots, n$$

Jackknife replication of the mean

$$\begin{aligned} s(\mathbf{x}_{(i)}) &= \tfrac{1}{n-1} \textstyle\sum_{j \neq i} x_j \\ &= \tfrac{(n\overline{x} - x_i)}{n-1} \\ &= \overline{x}_{(i)} \end{aligned}$$

Jackknife estimation of the standard error

1. Compute the n jackknife subsamples $\mathbf{x}_{(1)}, \cdots, \mathbf{x}_{(n)}$ from $\mathbf{x}$.
2. Evaluate the n jackknife replications $\hat{\theta}_{(i)} = s(\mathbf{x}_{(i)})$.
3. The jackknife estimate of the standard error is defined by:

$$\hat{\mathrm{se}}_{jack} = \left[\frac{n-1}{n} \sum_{i=1}^{n} (\hat{\theta}_{(\cdot)} - \hat{\theta}_{(i)})^2 \right]^{1/2}$$

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

Jackknife estimation of the standard error of the mean

For $\hat{\theta} = \overline{x}$, it is easy to show that:

$$\begin{cases} \overline{x}_{(i)} = \frac{n\overline{x} - x_i}{n-1} \\ \overline{x}(\cdot) = \frac{1}{n}\sum_{i=1}^{n} \overline{x}_{(i)} = \overline{x} \end{cases}$$

Therefore:

$$\begin{aligned} \widehat{\mathrm{se}}_{jack} &= \left\{ \sum_{i=1}^{n} \frac{(x_i - \overline{x})^2}{(n-1)n} \right\}^{1/2} \\ &= \frac{\overline{\sigma}}{\sqrt{n}} \end{aligned}$$

where $\overline{\sigma}$ is the unbiased variance.

Jackknife estimation of the standard error

- The factor $\frac{n-1}{n}$ is much larger than $\frac{1}{B-1}$ used in bootstrap.
- Intuitively this inflation factor is needed because jackknife deviation $(\hat{\theta}_{(i)} - \hat{\theta}_{(\cdot)})^2$ tend to be smaller than the bootstrap $(\hat{\theta}^*(b) - \hat{\theta}^*(\cdot))^2$ (the jackknife sample is more similar to the original data $\mathbf{x}$ than the bootstrap).
- In fact, the factor $\frac{n-1}{n}$ is derived by considering the special case $\hat{\theta} = \overline{x}$ (somewhat arbitrary convention).

Comparison of Jackknife and Bootstrap on an example

Example A: $\hat{\theta} = \overline{x}$

$f(x) = 0.2\ \mathcal{N}(\mu{=}1, \sigma{=}2) + 0.8\ \mathcal{N}(\mu{=}6, \sigma{=}1) \rightsquigarrow \mathbf{x} = (x_1, \cdots, x_{100})$.

- Bootstrap standard error and bias w.r.t. the number B of bootstrap samples:

B	10	20	50	100	500	1000	10000
$\widehat{\mathrm{se}}_B$	0.1386	0.2188	0.2245	0.2142	0.2248	0.2212	0.2187
$\widehat{\mathrm{Bias}}_B$	0.0617	-0.0419	0.0274	-0.0087	-0.0025	0.0064	0.0025

- Jackknife: $\widehat{\mathrm{se}}_{jack} = 0.2207$ and $\widehat{\mathrm{Bias}}_{jack} = 0$
- Using textbook formulas: $\mathrm{se}_{\hat{f}} = \frac{\hat{\sigma}}{\sqrt{n}} = 0.2196$ ($\frac{\overline{\sigma}}{\sqrt{n}} = 0.2207$).

Jackknife estimation of the bias

1. Compute the n jackknife subsamples $\mathbf{x}_{(1)}, \cdots, \mathbf{x}_{(n)}$ from $\mathbf{x}$.
2. Evaluate the n jackknife replications $\hat{\theta}_{(i)} = s(\mathbf{x}_{(i)})$.
3. The jackknife estimation of the bias is defined as:

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

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

Jackknife estimation of the bias

- Note the inflation factor $(n-1)$ (compared to the bootstrap bias estimate).
- $\hat{\theta} = \overline{x}$ is unbiased so the correspondence is done considering the plug-in estimate of the variance $\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n}$.
- The jackknife estimate of the bias for the plug-in estimate of the variance is then:

$$\widehat{\mathrm{Bias}}_{jack} = \frac{-\overline{\sigma}^2}{n}$$

Histogram of the replications

Example A

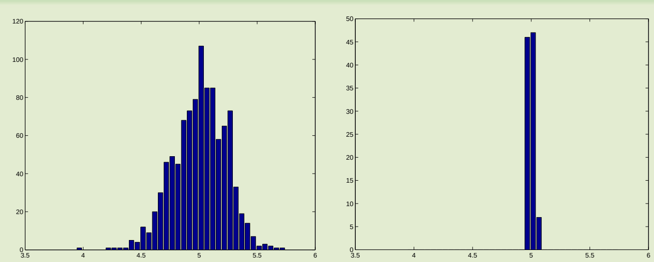
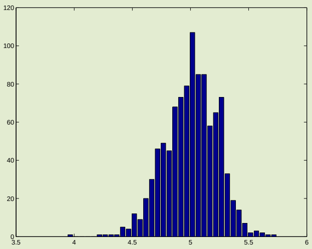


Figure: Histograms of the bootstrap replications $\{\hat{\theta}^*(b)\}_{b\in\{1,\cdots,B=1000\}}$ (left), and the jackknife replications $\{\hat{\theta}(_{}i)\}_{i\in\{1,\cdots,n=100\}}$ (right).

Histogram of the replications

Example A



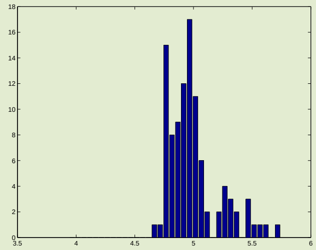


Figure: Histograms of the bootstrap replications $\{\hat{\theta}^*(b)\}_{b\in\{1,\cdots,B=1000\}}$ (left), and the inflated jackknife replications $\{\sqrt{n-1}(\hat{\theta}_{(i)} - \hat{\theta}_{(\cdot)}) + \hat{\theta}_{(\cdot)}\}_{i\in\{1,\cdots,n=100\}}$ (right).

Relationship between jackknife and bootstrap

- When n is small, it is easier (faster) to compute the n jackknife replications.
- However the jackknife uses less information (less samples) than the bootstrap.
- In fact, the jackknife is an approximation to the bootstrap!

Relationship between jackknife and bootstrap

- Considering a linear statistic :

$$\begin{aligned}\hat{\theta} &= s(\mathbf{x}) = \mu + \tfrac{1}{n}\textstyle\sum_{i=1}^n \alpha(x_i)\\ &= \mu + \tfrac{1}{n}\textstyle\sum_{i=1}^n \alpha_i\end{aligned}$$

Mean $\hat{\theta} = \overline{x}$

The mean is linear $\mu = 0$ and $\alpha(x_i) = \alpha_i = x_i, \quad \forall i \in \{1, \cdot, n\}$.

- There is no loss of information in using the jackknife to compute the standard error (compared to the bootstrap) for a linear statistic. Indeed the knowledge of the n jackknife replications $\{\hat{\theta}_{(i)}\}$, gives the value of $\hat{\theta}$ for any bootstrap data set.
- For non-linear statistics, the jackknife makes a linear approximation to the bootstrap for the standard error.

Relationship between jackknife and bootstrap

- Considering a quadratic statistic

$$\hat{\theta} = s(\mathbf{x}) = \mu + \tfrac{1}{n}\textstyle\sum_{i=1}^n \alpha(x_i) + \tfrac{1}{n^2}\beta(x_i, x_j)$$

Variance $\hat{\theta} = \hat{\sigma}^2$

$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (x_i - \overline{x})^2$ is a quadratic statistic.

- Again the knowledge of the n jackknife replications $\{s(\hat{\theta}_{(i)})\}$, gives the value of $\hat{\theta}$ for any bootstrap data set. The jackknife and bootstrap estimates of the bias agree for quadratic statistics.

Relationship between jackknife and bootstrap

The Law school example: $\hat{\theta} = \widehat{\mathrm{corr}}(\mathbf{x}, \mathbf{y})$.

The correlation is a non linear statistic.

- From B=3200 bootstrap replications, $\widehat{\mathrm{se}}_{B=3200} = 0.132$.
- From $n = 15$ jackknife replications, $\widehat{\mathrm{se}}_{jack} = 0.1425$.
- Textbook formula: $\mathrm{se}_{\hat{f}} = (1 - \widehat{\mathrm{corr}}^2)/\sqrt{n-3} = 0.1147$

Failure of the jackknife

The jackknife can fail if the estimate $\hat{\theta}$ is not smooth (i.e. a small change in the data can cause a large change in the statistic). A simple non-smooth statistic is the median.

On the mouse data

Compute the jackknife replications of the median $\mathbf{x}_{Cont} = (10, 27, 31, 40, 46, 50, 52, 104, 146)$ (Control group data).

- You should find 48,48,48,48,45,43,43,43,43 [a].
- Three different values appears as a consequence of a lack of smoothness of the median[b].

[a] The median of an even number of data points is the average of the middle 2 values.
[b] the median is not a differentiable function of x.

Delete-d Jackknife samples

Definition

The delete-d Jackknife subsamples are computed by leaving out d observations from $\mathbf{x}$ at a time.

- The dimension of the subsample is $n - d$.
- The number of possible subsamples now rises $\binom{n}{d} = \frac{n!}{d!(n-d)!}$.
- Choice: $\sqrt{n} < d < n - 1$

Delete-d jackknife

1. Compute all $\binom{n}{d}$ d-jackknife subsamples $\mathbf{x}_{(1)}, \cdots, \mathbf{x}_{(n)}$ from $\mathbf{x}$.
2. Evaluate the jackknife replications $\hat{\theta}_{(i)} = s(\mathbf{x}_{(i)})$.
3. Estimation of the standard error (when $n = r \cdot d$):

$$\widehat{\text{se}}_{d-jack} = \left\{ \frac{r}{\binom{n}{d}} \sum_i (\hat{\theta}_{(i)} - \hat{\theta}(\cdot))^2 \right\}^{1/2}$$

where $\hat{\theta}(\cdot) = \frac{\sum_i \hat{\theta}_{(i)}}{\binom{n}{d}}$.

Concluding remarks

- The inconsistency of the jackknife subsamples with non-smooth statistics can be fixed using delete-d jackknife subsamples.
- The subsamples (jackknife or delete-d jackknife) are actually samples (of smaller size) from the true distribution f whereas resamples (bootstrap) are samples from $\hat{f}$.

Summary

- Bias and standard error estimates have been introduced using jackknife replications.
- The Jackknife standard error estimate is a linear approximation of the bootstrap standard error.
- The Jackknife bias estimate is a quadratic approximation of the bootstrap bias.
- Using smaller subsamples (delete-d jackknife) can improve for non-smooth statistics such as the median.