Lectures 17: Jackknife sampling
error propagation for nonlinear functions of fit parameters

with material from https://www.scss.tcd.ie/Rozenn.Dahyot/

Jackknife sampling

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 sampling

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 sampling

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 sampling

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

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

Jackknife sampling

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:

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

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

Jackknife sampling

Jackknife estimation of the standard error of the mean

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

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

Therefore:

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

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

Jackknife sampling

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 **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).

Jackknife sampling

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 sampling

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 sampling

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:

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

Jackknife sampling

Histogram of the replications

Example A

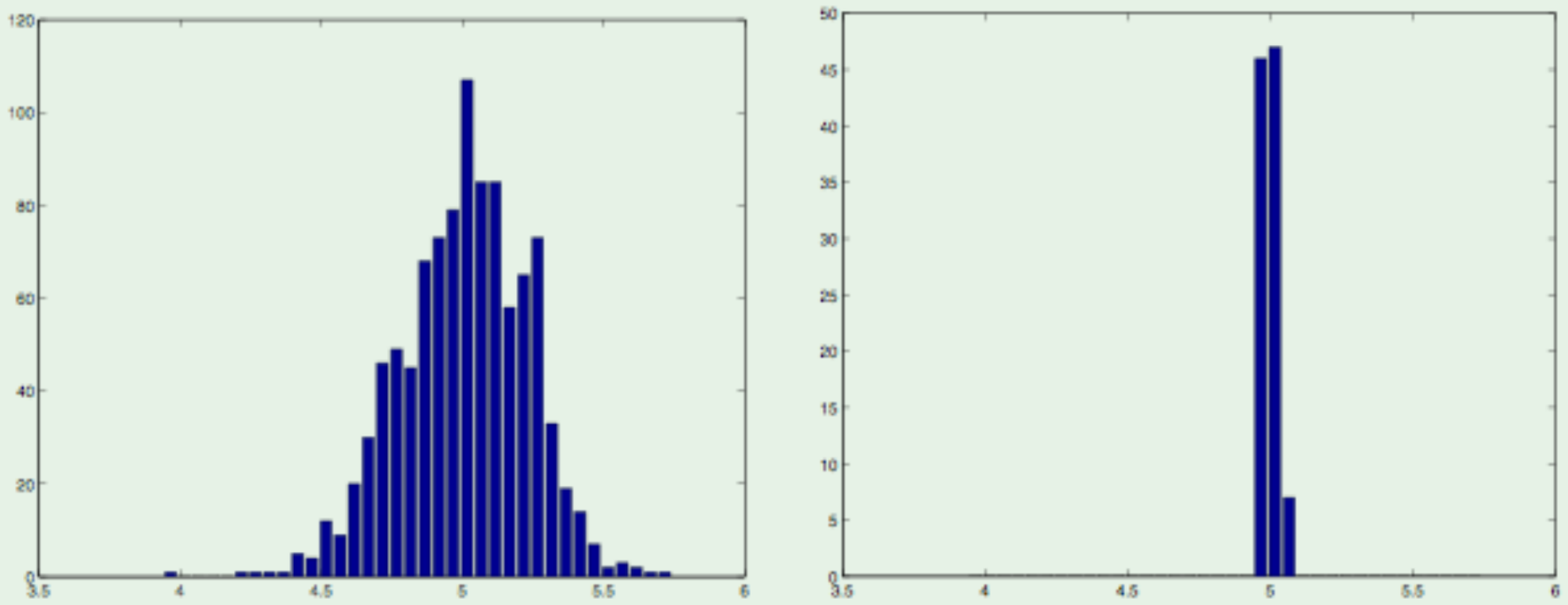

Figure: Histograms of the bootstrap replications $\{\hat{\theta}^*(b)\}_{b\in\{1,\cdots,B=1000\}}$ (left), and the jackknife replications $\{\hat{\theta}_{(}i)\}_{i\in\{1,\cdots,n=100\}}$ (right).

Jackknife sampling

Histogram of the replications

Example A

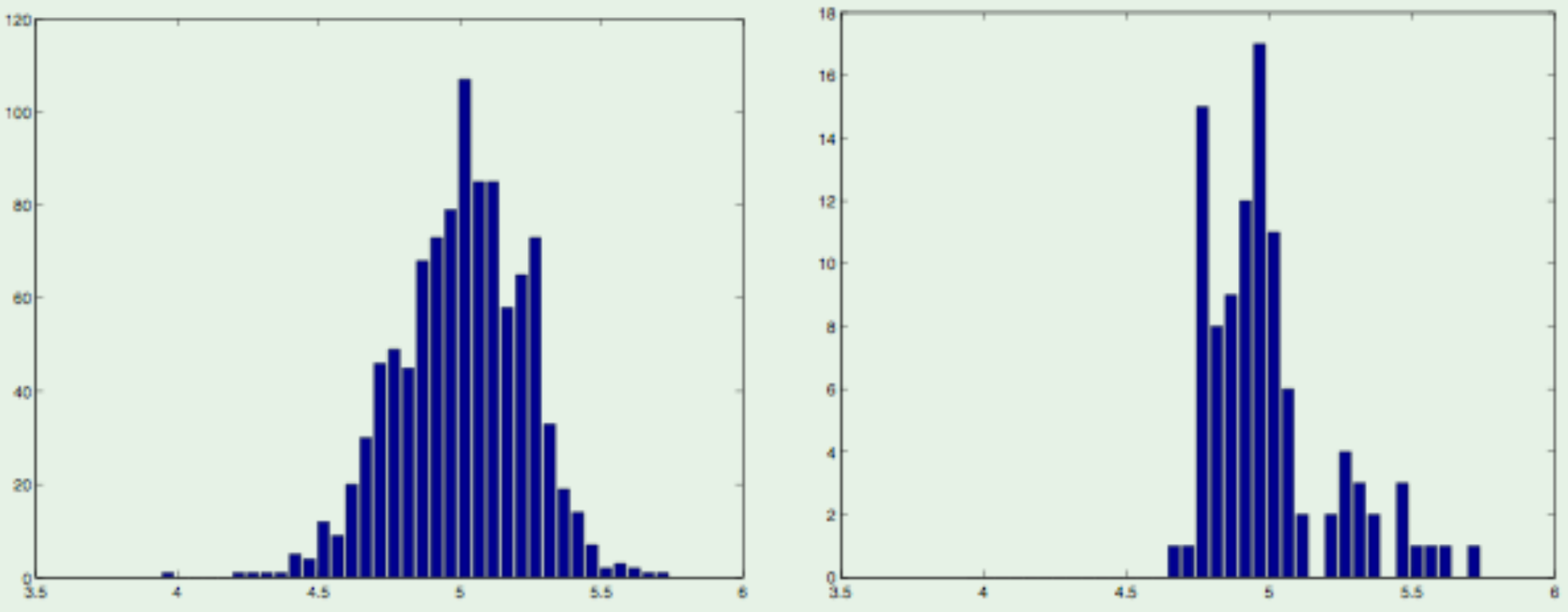

Figure: Histograms of the bootstrap replications $\{\hat{\theta}^*(b)\}_{b\in\{1,\cdots,B=1000\}}$ (left), and the inflated jackknife replications $\{\sqrt{n-1}(\hat{\theta}_{(i)} - \hat{\theta}_{(\cdot)}) + \hat{\theta}_{(\cdot)}\}_{i\in\{1,\cdots,n=100\}}$ (right).

Jackknife sampling

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!

Jackknife sampling

Relationship between jackknife and bootstrap

- Considering a linear statistic :

$$\hat{\theta} = s(\mathbf{x}) = \mu + \frac{1}{n}\sum_{i=1}^{n} \alpha(x_i)$$

$$= \mu + \frac{1}{n}\sum_{i=1}^{n} \alpha_i$$

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

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.

Jackknife sampling

Relationship between jackknife and bootstrap

- Considering a quadratic statistic

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

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

$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ 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.

Jackknife sampling

Relationship between jackknife and bootstrap

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

The correlation is a non linear statistic.

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

Jackknife sampling

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: $\sqrt{n} < d < n - 1$

Jackknife sampling

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 $n = r \cdot d$):

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

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

Jackknife sampling

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.

Jackknife sampling Matlab code

```
%% Jackknife Resampling
%

% Copyright 2015 The MathWorks, Inc.


%%
% Similar to the bootstrap is the jackknife, which uses resampling to
% estimate the bias of a sample statistic. Sometimes it is also used to
% estimate standard error of the sample statistic. The jackknife is
% implemented by the Statistics and Machine Learning Toolbox(TM) function
% |jackknife|.

%%
% The jackknife resamples systematically, rather than at random as the
% bootstrap does. For a sample with |n| points, the jackknife computes
% sample statistics on |n| separate samples of size |n|-1. Each sample is
% the original data with a single observation omitted.
```

```
%%
% In the bootstrap example, you measured the uncertainty in estimating the
% correlation coefficient. You can use the jackknife to estimate the bias,
% which is the tendency of the sample correlation to over-estimate or
% under-estimate the true, unknown correlation. First compute the sample
% correlation on the data.
load lawdata
rhohat = corr(lsat,gpa)

%%
% Next compute the correlations for jackknife samples, and compute their
% mean.
rng default;  % For reproducibility
jackrho = jackknife(@corr,lsat,gpa);
meanrho = mean(jackrho)

%%
% Now compute an estimate of the bias.
n = length(lsat);
biasrho = (n-1) * (meanrho-rhohat)

%%
% The sample correlation probably underestimates the true correlation by
% about this amount.
```