Lectures 13: Jackknife II.

(Jackknife/Bootstrap loose ends)

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

Jackknife sampling from Lecture 13

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 from Lecture 13

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

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

Jackknife sampling from Lecture 13

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 from Lecture 13

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).

Jackknife sampling from Lecture 13

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:

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

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

bootstrap review and bias   from Lecture 12

The Bootstrap algorithm for Estimating standard errors

1. Select B independent bootstrap samples $\mathbf{x}^{*(1)}, \mathbf{x}^{*(2)}, \cdots, \mathbf{x}^{*(B)}$ drawn from $\mathbf{x}$
2. Evaluate the bootstrap replications:

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

3. Estimate the standard error $\mathrm{se}_F(\hat{\theta})$ by the standard deviation of the B replications:

$$\hat{\mathrm{se}}_B = \left[\frac{\sum_{b=1}^{B} [\hat{\theta}^*(b) - \hat{\theta}^*(\cdot)]^2}{B-1} \right]^{\frac{1}{2}}$$

where $\hat{\theta}^*(\cdot) = \frac{\sum_{b=1}^{B} \hat{\theta}^*(b)}{B}$

Bootstrap estimate of bias

1. B independent bootstrap samples $\mathbf{x}^{*(1)}, \mathbf{x}^{*(2)}, \cdots, \mathbf{x}^{*(B)}$ drawn from $\mathbf{x}$
2. Evaluate the bootstrap replications:

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

3. Approximate the bootstrap expectation :

$$\hat{\theta}^*(\cdot) = \frac{1}{B}\sum_{b=1}^{B} \hat{\theta}^*(b) = \frac{1}{B}\sum_{b=1}^{B} s(\mathbf{x}^{*(b)})$$

4. the bootstrap estimate of bias based on B replications is:

$$\widehat{\mathrm{Bias}}_B = \hat{\theta}^*(\cdot) - \hat{\theta}$$

Bootstrap estimate of the standard Error

Example A

From the distribution F: $F(x) = 0.2\,\mathcal{N}(\mu{=}1,\sigma{=}2) + 0.8\,\mathcal{N}(\mu{=}6,\sigma{=}1)$. We draw the sample $\mathbf{x} = (x_1, \cdots, x_{100})$:

$$\mathbf{x} = \left\{\begin{array}{ccccc}
7.0411 & 4.8397 & 5.3156 & 6.7719 & 7.0616 \\
5.2546 & 7.3937 & 4.3376 & 4.4010 & 5.1724 \\
7.4199 & 5.3677 & 6.7028 & 6.2003 & 7.5707 \\
4.1230 & 3.8914 & 5.2323 & 5.5942 & 7.1479 \\
3.6790 & 0.3509 & 1.4197 & 1.7585 & 2.4476 \\
-3.8635 & 2.5731 & -0.7367 & 0.5627 & 1.6379 \\
-0.1864 & 2.7004 & 2.1487 & 2.3513 & 1.4833 \\
-1.0138 & 4.9794 & 0.1518 & 2.8683 & 1.6269 \\
6.9523 & 5.3073 & 4.7191 & 5.4374 & 4.6108 \\
6.5975 & 6.3495 & 7.2762 & 5.9453 & 4.6993 \\
6.1559 & 5.8950 & 5.7591 & 5.2173 & 4.9980 \\
4.5010 & 4.7860 & 5.4382 & 4.8893 & 7.2940 \\
5.5741 & 5.5139 & 5.8869 & 7.2756 & 5.8449 \\
6.6439 & 4.5224 & 5.5028 & 4.5672 & 5.8718 \\
6.0919 & 7.1912 & 6.4181 & 7.2248 & 8.4153 \\
7.3199 & 5.1305 & 6.8719 & 5.2686 & 5.8055 \\
5.3602 & 6.4120 & 6.0721 & 5.2740 & 7.2329 \\
7.0912 & 7.0766 & 5.9750 & 6.6091 & 7.2135 \\
4.9585 & 5.9042 & 5.9273 & 6.5762 & 5.3702 \\
4.7654 & 6.4668 & 6.1983 & 4.3450 & 5.3261
\end{array}\right\}$$

We have $\mu_F = 5$ and $\overline{x} = 4.9970$.

bootstrap review and bias from Lecture 12

Bootstrap estimate of the standard Error

Example A

1. $B = 1000$ bootstrap samples $\{\mathbf{x}^{*(b)}\}$
2. $B = 1000$ replications $\{\overline{x}^*(b)\}$
3. Bootstrap estimate of the standard error:

$$\widehat{\text{se}}_{B=1000} = \left[\frac{\sum_{b=1}^{1000}[\overline{x}^*(b) - \overline{x}^*(\cdot)]^2}{1000 - 1}\right]^{\frac{1}{2}} = 0.2212$$

where $\overline{x}^*(\cdot) = 5.0007$. This is to compare with $\hat{se}(\overline{x}) = \frac{\hat{\sigma}}{\sqrt{n}} = 0.22$.

bootstrap review and bias from Lecture 12

Distribution of $\hat{\theta}$

When enough bootstrap resamples have been generated, not only the standard error but any aspect of the distribution of the estimator $\hat{\theta} = t(\hat{F})$ could be estimated. One can draw a histogram of the distribution of $\hat{\theta}$ by using the observed $\hat{\theta}^*(b)$, $b = 1, \cdots, B$.

Example A

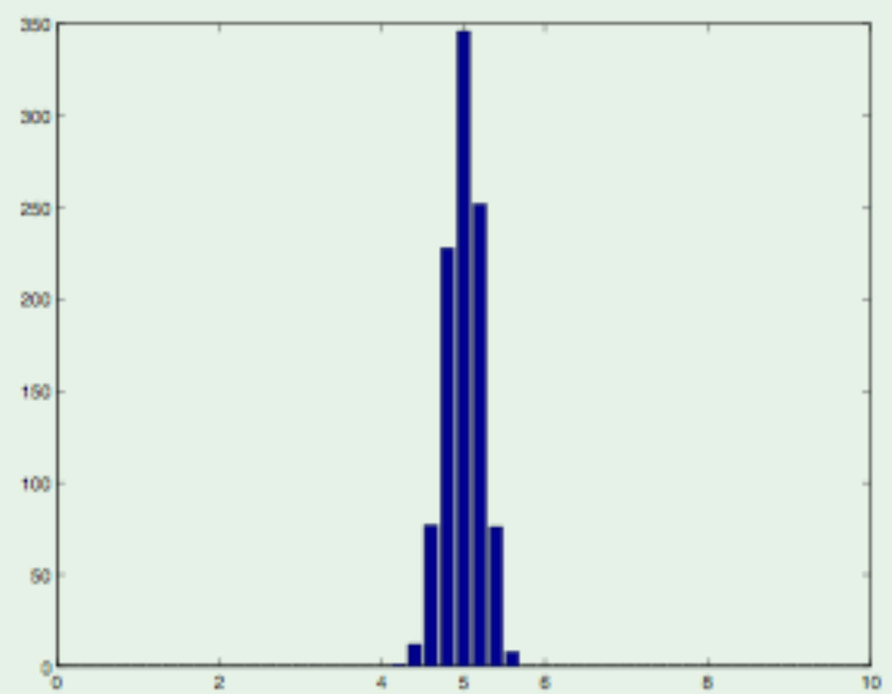

Figure: Histogram of the replications $\{\bar{x}^*(b)\}_{b=1\cdots B}$.

bootstrap review and bias   from Lecture 12

Bootstrap estimate of the standard error

Definition

The ideal bootstrap estimate $\text{se}_{\hat{F}}(\theta^*)$ is defined as:

$$\lim_{B \to \infty} \hat{\text{se}}_B = \text{se}_{\hat{F}}(\theta^*)$$

$\text{se}_{\hat{F}}(\theta^*)$ is called a non-parametric bootstrap estimate of the standard error.

bootstrap review and bias   from Lecture 12

Bootstrap estimate of the standard Error

How many B in practice ?

you may want to limit the computation time. In practice, you get a good estimation of the standard error for B in between 50 and 200.

Example A

B	10	20	50	100	500	1000	10000
$\widehat{\text{se}}_B$	0.1386	0.2188	0.2245	0.2142	0.2248	0.2212	0.2187

Table: Bootstrap standard error w.r.t. the number B of bootstrap samples.

bootstrap review and bias from Lecture 12

Bootstrap estimate of bias

Definition

The bootstrap estimate of bias is defined to be the estimate:

$$\text{Bias}_{\hat{F}}(\hat{\theta}) = \mathbb{E}_{\hat{F}}[s(\mathbf{x}^*)] - t(\hat{F})$$

$$= \theta^*(\cdot) - \hat{\theta}$$

Example A

B	10	20	50	100	500	1000	10000
$\mathbb{E}_{\hat{F}}(\bar{x}^*)$	5.0587	4.9551	5.0244	4.9883	4.9945	5.0035	4.9996
$\widehat{\text{Bias}}$	0.0617	-0.0419	0.0274	-0.0087	-0.0025	0.0064	0.0025

Table: $\widehat{\text{Bias}}$ of $\bar{x}^*$ ($\bar{x} = 4.997$ and $\mu_F = 5$).

Comparison of Jackknife and Bootstrap on an example

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

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

- 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: $\mathrm{se}_{\hat{F}} = \frac{\hat{\sigma}}{\sqrt{n}} = 0.2196$ ($\frac{\bar{\sigma}}{\sqrt{n}} = 0.2207$).

Jackknife sampling from Lecture 13

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:

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

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

Jackknife sampling from Lecture 13

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:

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

Jackknife sampling from Lecture 13

Histogram of the replications

Example A

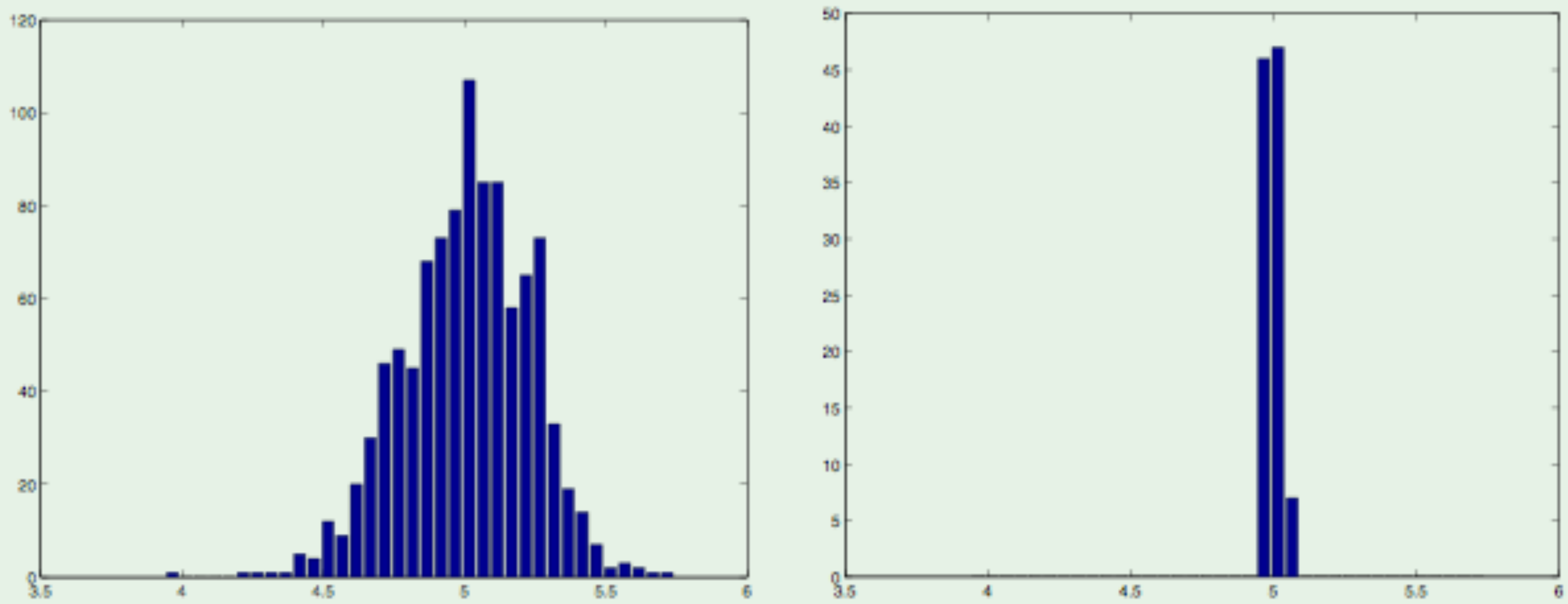

Figure: Histograms of the bootstrap replications $\{\hat{\theta}^*(b)\}_{b\in\{1,\cdots,B=1000\}}$ (left), and the jackknife replications $\{\hat{\theta}_{(}i)\}_{i\in\{1,\cdots,n=100\}}$ (right).

Jackknife sampling from Lecture 13

Histogram of the replications

Example A

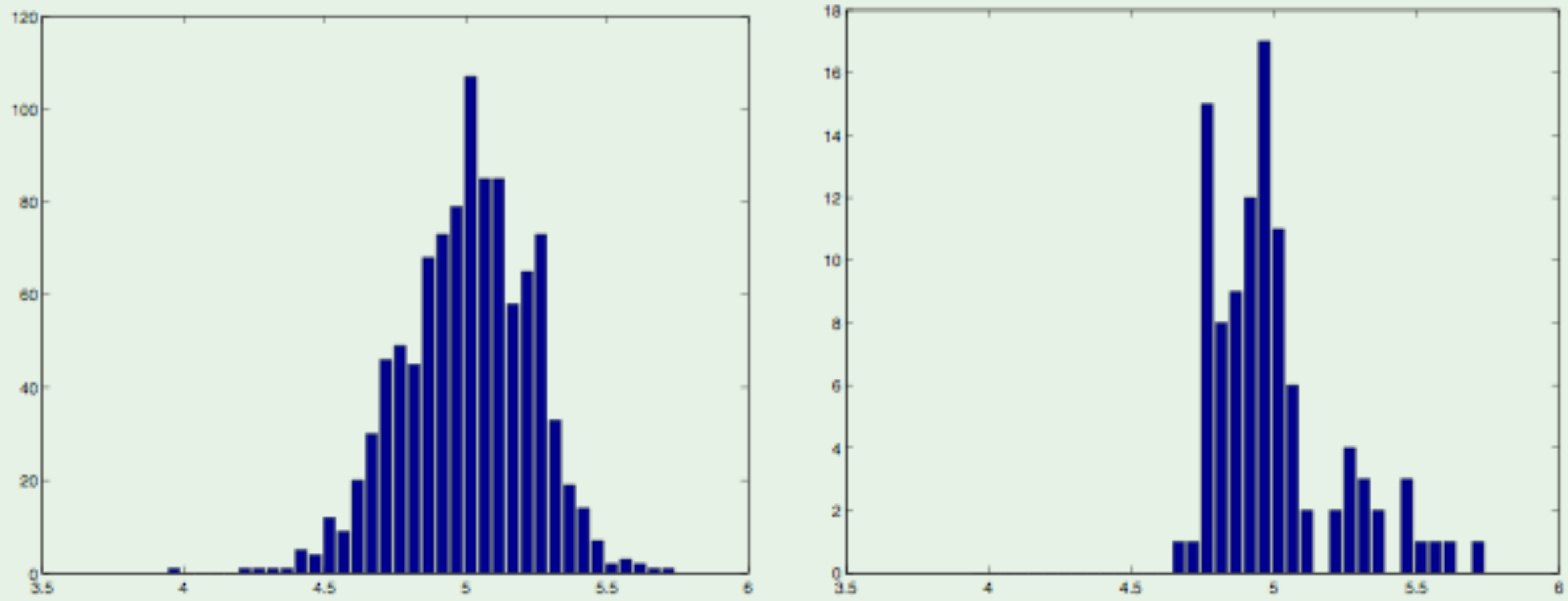

Figure: Histograms of the bootstrap replications $\{\hat{\theta}^*(b)\}_{b\in\{1,\cdots,B=1000\}}$ (left), and the inflated jackknife replications $\{\sqrt{n-1}(\hat{\theta}_{(i)} - \hat{\theta}_{(\cdot)}) + \hat{\theta}_{(\cdot)}\}_{i\in\{1,\cdots,n=100\}}$ (right).

Jackknife sampling

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 $n = r \cdot d$):

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

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

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 :

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

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

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

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

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

$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$ 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: $\hat{\theta} = \widehat{\text{corr}}(\mathbf{x}, \mathbf{y})$.

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

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.
```