Lectures 16-17: Jackknife I.-II.

compared with bootstrap

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

The Empirical density function

Statistical inference concerns learning from experience: we observe a random sample $\mathbf{x} = (x_1, x_2, \cdots, x_n)$ and wish to infer properties of the complete population $\mathcal{X}$ that yielded the sample. A complete knowledge is obtained from the population density function $F(.)$ from which $\mathbf{x}$ has been generated $F \rightsquigarrow \mathbf{x} = (x_1, x_2, \cdots, x_n)$

Definition

The empirical density function $\hat{F}(.)$ is defined as:

$$\hat{F}(x) = \frac{1}{n} \sum_{i=1}^{n} \delta(x - x_i)$$

where $\delta(\cdot)$ is the Dirac delta function. So the probability of $x = x_j$ is :

$$\int \hat{F}(x_j)\, dx = \int \frac{1}{n} \sum_{i=1}^{n} \delta(x_j - x_i)\, dx = \begin{cases} \frac{1}{n}, & x_j \in \{x_1, \cdots, x_n\} \\ 0, & \text{otherwise} \end{cases}$$

statistics reviewed

Parameters

Definition

A parameter, θ, is a function of the probability density function (p.d.f.) F, e.g.:

$$\theta = t(F)$$

if θ is the mean

$$\theta = \mathbb{E}_F(x) = \int_{-\infty}^{+\infty} x \; F(x)dx = \mu_F$$

if θ is the variance

$$\theta = \mathbb{E}_F[(x - \mu_F)^2] = \int_{-\infty}^{+\infty} (x - \mu_F)^2 \; F(x)dx = \sigma_F^2$$

statistics reviewed

Statistics or estimates

Definition

A statistic (also called estimates, estimators) $\hat{\theta}$ is a function of $\hat{F}$ or the sample $\mathbf{x}$, e.g.:

$$\hat{\theta} = t(\hat{F})$$

or also written $\hat{\theta} = s(\mathbf{x})$.

if $\hat{\theta}$ is the mean:

$$\begin{aligned} \hat{\theta} &= t(\hat{F}) &= \int_{-\infty}^{+\infty} x \, \hat{F}(x) dx \\ & &= \int_{-\infty}^{+\infty} x \, \frac{1}{n} \sum_{i=1}^{n} \delta(x - x_i) \, dx \\ & &= \frac{1}{n} \sum_{i=1}^{n} x_i \\ & &= s(\mathbf{x}) = \overline{x} \end{aligned}$$

statistics reviewed

Statistics or estimates

if $\hat{\theta}$ is the variance:

$$\begin{aligned}\hat{\theta} &= \int_{-\infty}^{+\infty} (x - \overline{x})^2 \, \hat{F}(x) dx \\ &= \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2 \\ &= \hat{\sigma}^2\end{aligned}$$

The Plug-in principle

Definition

The Plug-in estimate of a parameter $\theta = t(F)$ is defined to be:

$$\hat{\theta} = t(\hat{F}).$$

The function $\theta = t(F)$ of the probability density function F is estimated by the same function $t(.)$ of the empirical density $\hat{F}$.

- $\bar{x}$ is the plug-in estimate of μ_F.
- $\hat{\sigma}$ is the plug-in estimate of σ_F

Computing the mean knowing F

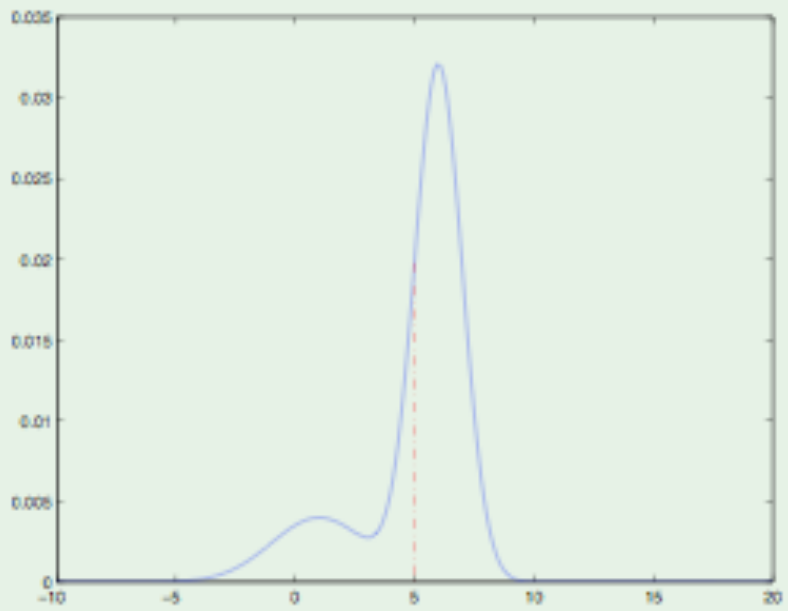
Example A

Lets assume we know the p.d.f. F:

$$F(x) = 0.2\ \mathcal{N}(\mu{=}1,\sigma{=}2) + 0.8\ \mathcal{N}(\mu{=}6,\sigma{=}1)$$

Then the mean is computed:

$$\begin{aligned}\mu_F = \mathbb{E}_F(x) &= \int_{-\infty}^{+\infty} x\ F(x)\ dx \\ &= 0.2 \cdot 1 + 0.8 \cdot 6 \\ &= 5\end{aligned}$$



statistics reviewed

Estimating the mean knowing the observations $\mathbf{x}$

Example A

Observations $\mathbf{x} = (x_1, \cdots, x_{100})$:

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

From the samples, the mean can be computed:

$$\bar{x} = \frac{\sum_{i=1}^{100} x_i}{100} = 4.9970$$

statistics reviewed

Accuracy of estimates $\hat{\theta}$

We can compute an estimate $\hat{\theta}$ of a parameter θ from an observation sample $\mathbf{x} = (x_1, x_2, \cdots, x_n)$. But how accurate is $\hat{\theta}$ compared to the real value θ ?

Our attention is focused on questions concerning the probability distribution of $\hat{\theta}$. For instance we would like to know about:

- its standard error
- its confidence interval
- its bias
- etc.

Standard error of $\hat{\theta}$

Definition

The standard error is the standard deviation of a statistic $\hat{\theta}$. As such, it measures the precision of an estimate of the statistic of a population distribution.

$$se(\hat{\theta}) = \sqrt{var_F[\hat{\theta}]}$$

Standard error of $\overline{x}$

We have:

$$\mathbb{E}_F\left[(\overline{x} - \mu_F)^2\right] = \frac{\sum_{i=1}^{n} \mathbb{E}_F\left[(x_i - \mu_F)^2\right]}{n^2} = \frac{\sigma_F^2}{n}$$

Then

$$se_F(\overline{x}) = [\mathrm{var}_F(\overline{x})]^{1/2} = \frac{\sigma_F}{\sqrt{n}}$$

Plug in estimate of the standard error

Suppose now that F is unknown and that only the random sample $\mathbf{x} = (x_1, \cdots, x_n)$ is known. As μ_F and σ_F are unknown, we can use the previous formula to compute a plug-in estimate of the standard error.

Definition

The estimated standard error of the estimator $\hat{\theta}$ is defined as:

$$\widehat{\mathrm{se}}(\hat{\theta}) = \mathrm{se}_{\hat{F}}(\hat{\theta}) = [\mathrm{var}_{\hat{F}}(\hat{\theta})]^{1/2}$$

Estimated standard error of $\overline{x}$

$$\widehat{\mathrm{se}}(\overline{x}) = \frac{\hat{\sigma}}{\sqrt{n}}$$

Bias of $\hat{\theta}$

Definition

The Bias is the difference between the expectation of an estimator $\hat{\theta}$ and the quantity θ being estimated:

$$\mathrm{Bias}_F(\hat{\theta}, \theta) = \mathbb{E}_F(\hat{\theta}) - \theta$$

Bias of the mean $\overline{x}$

we have:

$$\mathbb{E}_F(\overline{x}) = \mathbb{E}_F\left(\frac{\sum_{i=1}^n x_i}{n}\right) = \frac{\sum_{i=1}^n \mathbb{E}_F(x_i)}{n} = \mu_F$$

then:

$$\mathrm{Bias}_F(\overline{x}, \mu_F) = \mathbb{E}_F(\overline{x}) - \mu_F = 0$$

statistics reviewed

Bias of $\hat{\theta}$

Bias of $\hat{\sigma}^2$

$$\begin{aligned}\hat{\sigma}^2 &= \tfrac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \tfrac{1}{n}\sum_{i=1}^{n}((x_i - \mu_F) + (\mu_F - \bar{x}))^2 \\ &= \left(\tfrac{1}{n}\sum_{i=1}^{n}(x_i - \mu_F)^2\right) - (\bar{x} - \mu_F)^2\end{aligned}$$

The first term has an expected value of σ_F^2 and the second term has expected value σ_F^2/n. So the bias of $\hat{\sigma}^2$ is:

$$\mathrm{Bias}_F(\hat{\sigma}^2, \sigma_F^2) = \sigma_F^2 - \frac{\sigma_F^2}{n} - \sigma_F^2 = -\frac{\sigma_F^2}{n}$$

Bias of $\hat{\theta}$

Instead of using $\hat{\sigma}^2$ as an estimate of the variance, you should try to choose an unbiased estimate.

Bias of $\overline{\sigma}^2$

Let define:

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

then by computing its bias:

$$\begin{aligned} \mathrm{Bias}_F(\overline{\sigma}^2, \sigma_F^2) &= \mathbb{E}_F(\overline{\sigma}^2) - \sigma_F^2 \\ &= 0 \end{aligned}$$

$\overline{\sigma}$ is an unbiased estimator of the standard deviation.

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

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

Jackknife sampling

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:

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

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 $\hat{\theta} = \overline{x}$, 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.

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 $\mathbf{x}$ than the bootstrap).
- In fact, the factor $\frac{n-1}{n}$ is derived by considering the special case $\hat{\theta} = \bar{x}$ (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:

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

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

Jackknife sampling

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

Histogram of the replications

Example A

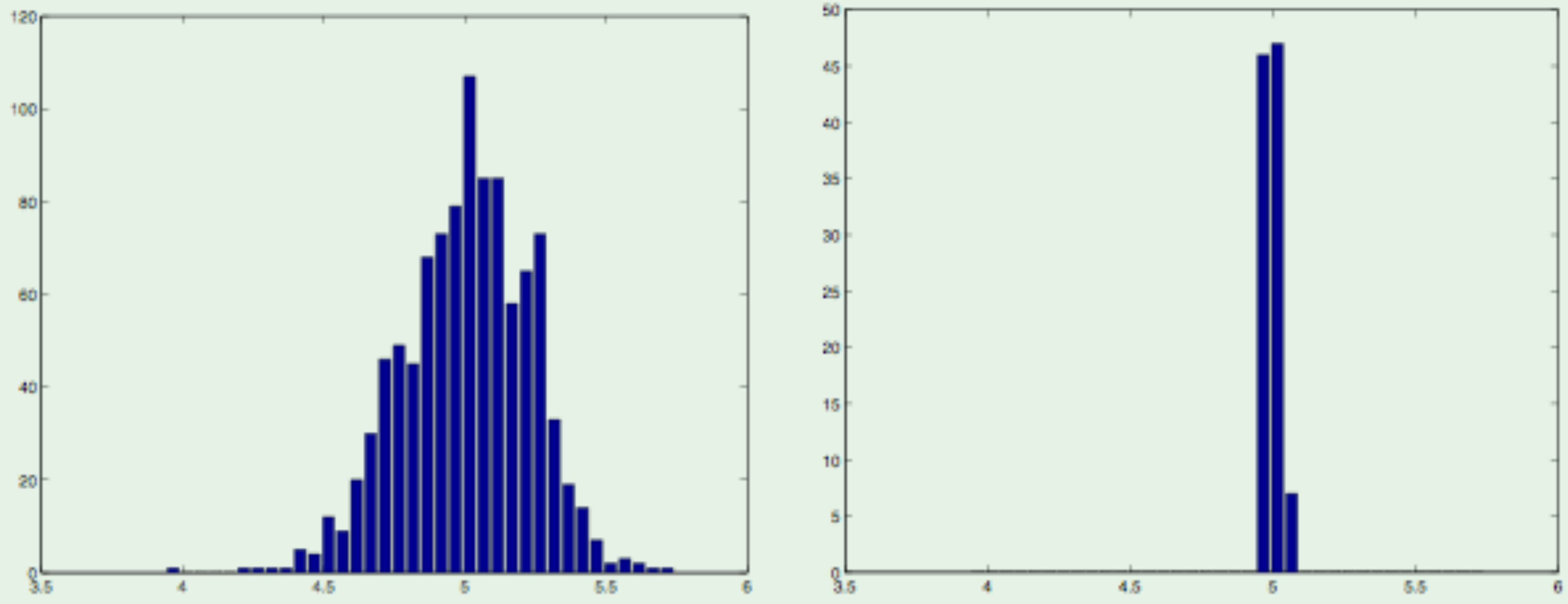


Figure: Histograms of the bootstrap replications $\{\hat{\theta}^*(b)\}_{b\in\{1,\cdots,B=1000\}}$ (left), and the jackknife replications $\{\hat{\theta}_(i)\}_{i\in\{1,\cdots,n=100\}}$ (right).

Histogram of the replications

Example A

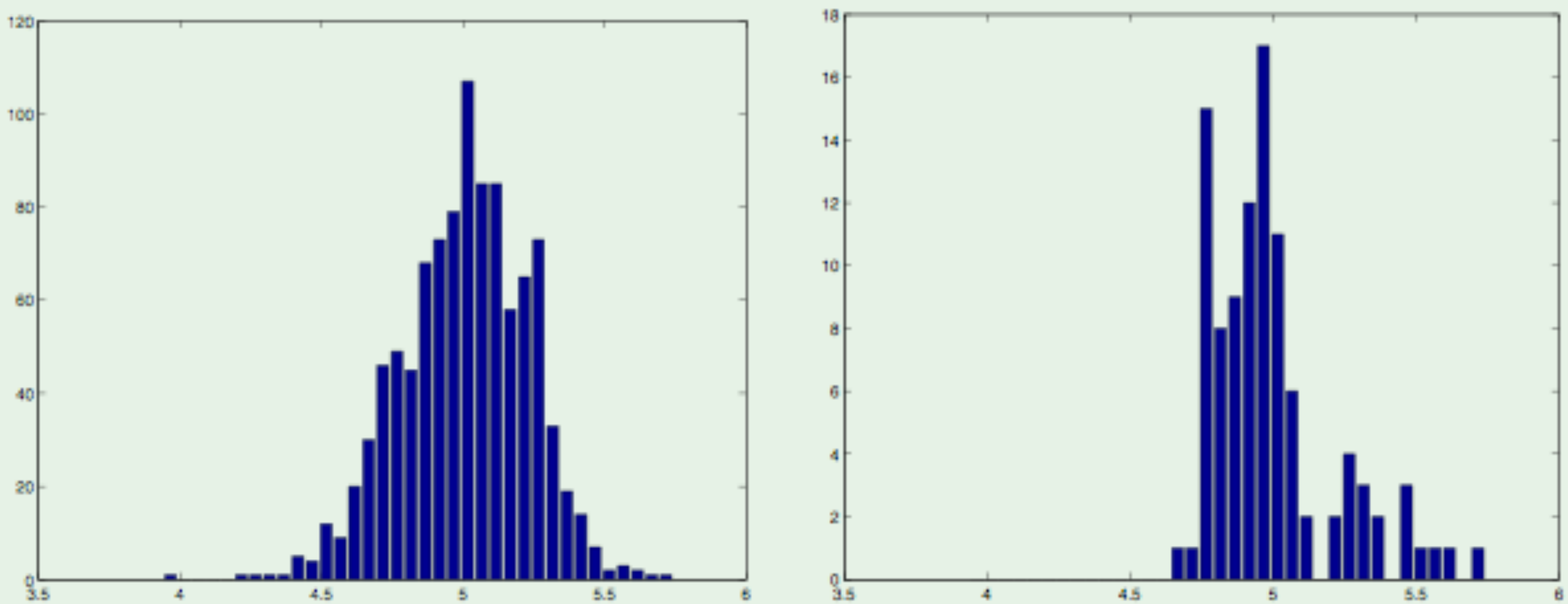


Figure: Histograms of the bootstrap replications $\{\hat{\theta}^*(b)\}_{b\in\{1,\cdots,B=1000\}}$ (left), and the inflated jackknife replications $\{\sqrt{n-1}(\hat{\theta}_{(i)} - \hat{\theta}_{(\cdot)}) + \hat{\theta}_{(\cdot)}\}_{i\in\{1,\cdots,n=100\}}$ (right).

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

bootstrap review and bias

bootstrap review and bias

Bootstrap samples and replications

Definition

A bootstrap sample $\mathbf{x}^* = (x_1^*, x_2^*, \cdots, x_n^*)$ is obtained by randomly sampling n times, with replacement, from the original data points $\mathbf{x} = (x_1, x_2, \cdots, x_n)$.

Considering a sample $\mathbf{x} = (x_1, x_2, x_3, x_4, x_5)$, some bootstrap samples can be:

$$\mathbf{x}^{*(1)} = (x_2, x_3, x_5, x_4, x_5)$$
$$\mathbf{x}^{*(2)} = (x_1, x_3, x_1, x_4, x_5)$$

etc.

Definition

With each bootstrap sample $\mathbf{x}^{*(1)}$ to $\mathbf{x}^{*(B)}$, we can compute a bootstrap replication $\hat{\theta}^*(b) = s(\mathbf{x}^{*(b)})$ using the plug-in principle.

bootstrap review and bias

How to compute Bootstrap samples

Repeat B times:

1. A random number device selects integers $i_1, \cdots, i_n$ each of which equals any value between 1 and n with probability $\frac{1}{n}$.
2. Then compute $\mathbf{x}^* = (x_{i_1}, \cdots, x_{i_n})$.

Some matlab code available on the web

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

bootstrap review and bias

How many values are left out of a bootstrap resample ?

Given a sample $\mathbf{x} = (x_1, x_2, \cdots, x_n)$ and assuming that all x_i are different, the probability that a particular value x_i is left out of a resample $\mathbf{x}^* = (x_1^*, x_2^*, \cdots, x_n^*)$ is:

$$\mathcal{P}(x_j^* \neq x_i, 1 \leq j \leq n) = \left(1 - \frac{1}{n}\right)^n$$

since $\mathcal{P}(x_j^* = x_i) = \frac{1}{n}$. When n is large, the probability $\left(1 - \frac{1}{n}\right)^n$ converges to $e^{-1} \approx 0.37$.

bootstrap review and bias

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 review and bias

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

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

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

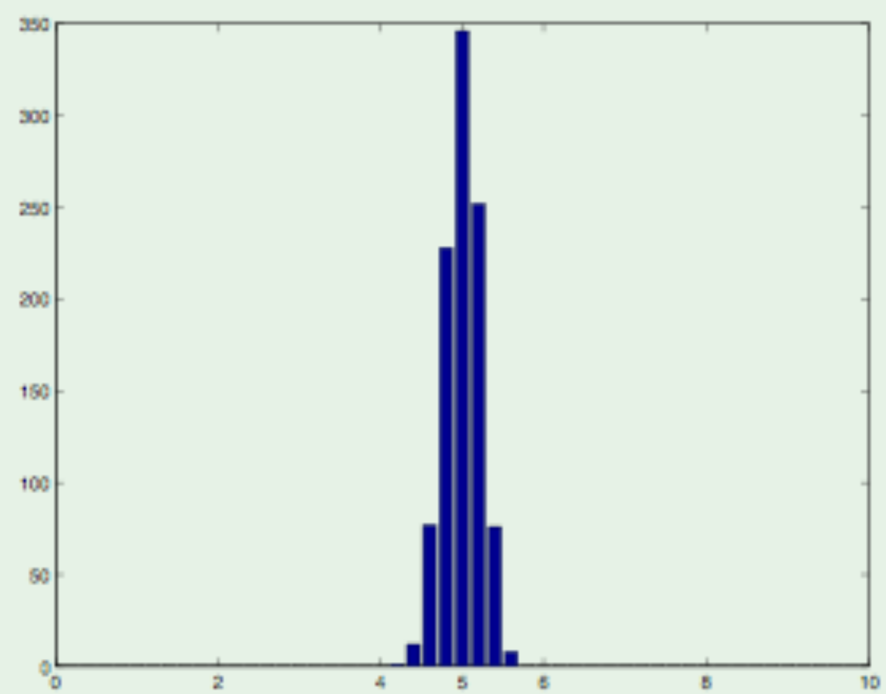


Figure: Histogram of the replications $\{\overline{x}^*(b)\}_{b=1\cdots B}$.

Bootstrap estimate of the standard error

Definition

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

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

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

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
$\widehat{se}_B$	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

Bootstrap estimate of bias

Definition

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

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

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

Example A

B	10	20	50	100	500	1000	10000
$\mathbb{E}_{\hat{F}}(\overline{x}^*)$	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 $\overline{x}^*$ ($\overline{x} = 4.997$ and $\mu_F = 5$).

bootstrap review and bias

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:

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

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 :

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

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.

Jackknife sampling

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.

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.
```