Chapter 4

Resampling Methods

4.1 Non-parametric computational estimation

Let $x_1, \ldots, x_n$ be a realization of the i.i.d. r.vs $X_1, \ldots, X_n$ with a c.d.f. $F$.

We are interested in the precision of estimation of a population parameter $\theta_F$. One possibility is to estimate $\theta_F$ by $\hat{\theta}_F$, where $\hat{F}$ is the empirical distribution function. We will denote an estimator of a parameter $\theta$ by $\hat{\theta}$.

Examples

1. 

$$\theta_F = E(X) = \int_{-\infty}^{\infty} x f(x) dx,$$

where $f(x) = \frac{dF(x)}{dx}$. Then

$$\hat{\theta}_F = \frac{1}{n} \sum_{i=1}^{n} x_i,$$

which is the sample mean. Here we assign equal probability, $\frac{1}{n}$, to each realization of $X$.

2. 

$$\theta_F = var(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f(x) dx.$$

Then

$$\hat{\theta}_F = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2,$$

which is the variance of the sample.
3.  

\[ \theta_F = F(c) = P(X \leq c) \]

Then

\[ \hat{\theta} = \frac{1}{n} \# \{ i : x_i \leq c \} \]

One might ask: How good is \( \hat{\theta} = \theta_F \) as an estimator of \( \theta_F \)?

Three common measures of goodness are:

\[ Bias_\theta(\hat{\theta}) = E_F(\hat{\theta}) - \theta \quad (4.1) \]

\[ se_\theta(\hat{\theta}) = \sqrt{\text{var}(\hat{\theta})} \quad (4.2) \]

\[ MSE_\theta(\hat{\theta}) = E_F[(\hat{\theta} - \theta)^2] \quad (4.3) \]

It is easy to see that

\[ MSE_\theta(\hat{\theta}) = \text{var}(\hat{\theta}) + (Bias_\theta(\hat{\theta}))^2 \quad (4.4) \]

Namely:

\[ E[(\hat{\theta} - \theta)^2] = E(\hat{\theta}^2 - 2\hat{\theta}\theta + \theta^2) = \]

\[ E[\hat{\theta}^2 - 2\hat{\theta}E(\hat{\theta}) - Bias_\theta(\hat{\theta}) + (E(\hat{\theta}) - Bias_\theta(\hat{\theta}))^2] = \]

\[ E[\hat{\theta}^2 - 2\hat{\theta}E(\hat{\theta}) - 2\hat{\theta}Bias_\theta(\hat{\theta}) + (E(\hat{\theta})^2 - 2E(\hat{\theta})Bias_\theta(\hat{\theta}) + (Bias_\theta(\hat{\theta}))^2] = \]

\[ E[\hat{\theta}^2 - 2\hat{\theta}E(\hat{\theta}) + (E(\hat{\theta})^2) + (Bias_\theta(\hat{\theta}))^2 = \]

\[ \text{var}(\hat{\theta}) + (Bias_\theta(\hat{\theta}))^2 \]

Also note that

\[ \sqrt{MSE_\theta(\hat{\theta})} = \sqrt{\text{var}(\hat{\theta}) + (Bias_\theta(\hat{\theta}))^2} = se_\theta(\hat{\theta}) \times \sqrt{1 + (Bias_\theta(\hat{\theta}))^2} \]

\[ se_\theta(\hat{\theta}) \times \left[ 1 + \frac{1}{2} \left( \frac{Bias_\theta(\hat{\theta})}{se_\theta(\hat{\theta})} \right)^2 \right] \]
4.2. Bootstrap Estimates of Bias, Standard Error and MSE

Problem
How to calculate $Bias_\theta(\hat{\theta})$, $se_\theta(\hat{\theta})$ and $MSE_\theta(\hat{\theta})$?

If we knew the distribution $F$ then we could calculate expected value and variance of the estimator $\hat{\theta}$ directly from definitions. It may be difficult if $f(x) = \frac{dF}{dx}$ is complicated. Then a practical alternative is simulation:

- Generate a large number of random samples from the population with the c.d.f. $F$ and calculate a value of $\hat{\theta}$ for each sample.
- The mean and variance of the set of generated values of $\hat{\theta}$ will give a good approximation to $E_F(\hat{\theta})$ and $var_F(\hat{\theta})$.

What if $F$ is unknown? Then the simulation from $F$ is impossible. In such situations a further approximation is to replace $F$ by $\hat{F}$. Let $\theta^*$ be the value of $\theta$ calculated at the random sample from $\hat{F}$. The idea is that

$$Bias_\theta(\hat{\theta}) \approx Bias_\hat{\theta}(\theta^*)$$

and

$$var_F(\hat{\theta}) \approx var_{\hat{F}}(\theta^*).$$

The heuristic reasoning is that $\hat{F}$ is close to $F$ and so the relationship of $\theta_F$ to $\theta_F^*$ should be close to the relationship of $\theta_F^*$ to $\theta_F$, as shown in the diagram below.

<table>
<thead>
<tr>
<th>True unknown</th>
<th>Empirical</th>
<th>Resampled</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F$</td>
<td>$\hat{F}$</td>
<td>$F^*$</td>
</tr>
<tr>
<td>$\downarrow$ data</td>
<td>$\downarrow$ resampling</td>
<td>$\downarrow$</td>
</tr>
<tr>
<td>$\theta_F$</td>
<td>$\theta_{\hat{F}}$</td>
<td>$\theta_{F^*}$</td>
</tr>
</tbody>
</table>

4.2 Bootstrap estimates of bias, standard error and MSE

Assume we do not know $F$. The bootstrap estimates of $Bias_\theta(\hat{\theta})$, $se_\theta(\hat{\theta})$ and $MSE_\theta(\hat{\theta})$ are obtained by substituting $\hat{F}$ for $F$ in (4.1), (4.2) and (4.3), respectively. $\hat{F}$ is the distribution which assigns probability $\frac{1}{n}$ to each observation $x_i$. So, a random sample from $\hat{F}$ is just a random sample from the set $\{x_1, \ldots, x_n\}$ with replacement. The procedure to calculate the estimates is the following:
CHAPTER 4. RESAMPLING METHODS

- construct \( N \) samples of size \( n \) from \( \{x_1, \ldots, x_n\} \) with replacement;
- denote the bootstrap samples by \( \{x_1, \ldots, x_n\}_i, i = 1, \ldots, N \);
- denote by \( \theta_i^* \) the value of the estimator calculated for the \( i \)-th bootstrap sample;
- calculate sample mean and variance of bootstrap estimates \( \theta_i^* \), \( i = 1, \ldots, n \), that is \( \bar{\theta}^* = \frac{1}{N} \sum_{i=1}^{N} \theta_i^* \), \( \frac{1}{N-1} \sum_{i=1}^{N} (\theta_i^* - \bar{\theta}^*)^2 \);

\( \text{Bias}_\theta(\hat{\theta}), \var_F(\hat{\theta}), \text{se}_F(\hat{\theta}) \) and \( \text{MSE}_\theta(\hat{\theta}) \) are approximated, respectively, by \( \hat{\text{Bias}}_\theta(\hat{\theta}), \hat{\var}_F(\hat{\theta}), \hat{\text{se}}_F(\hat{\theta}) \) and \( \hat{\text{MSE}}_\theta(\hat{\theta}) \) which are further approximated by

\[
\hat{\text{Bias}}_\theta(\hat{\theta}) = \bar{\theta}^* - \hat{\theta},
\]
\[
\hat{\var}_F(\theta^*) = \frac{1}{N-1} \sum_{i=1}^{N} (\theta_i^* - \bar{\theta}^*)^2,
\]
\[
\hat{\text{se}}_F(\theta^*) = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (\theta_i^* - \bar{\theta}^*)^2},
\]
\[
\hat{\text{MSE}}_\theta(\theta^*) = \frac{1}{N-1} \sum_{i=1}^{N} (\theta_i^* - \bar{\theta}^*)^2 + (\bar{\theta}^* - \hat{\theta})^2.
\]

The following diagram represents the bootstrap resampling method:

<table>
<thead>
<tr>
<th>Empirical distribution</th>
<th>Bootstrap samples of size ( n )</th>
<th>Bootstrap replications of ( \hat{\theta} )</th>
<th>Bootstrap estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{F} )</td>
<td>( {x_1, \ldots, x_n}_1 )</td>
<td>( \theta_1^* )</td>
<td>( \theta_1^* )</td>
</tr>
<tr>
<td></td>
<td>( {x_1, \ldots, x_n}_2 )</td>
<td>( \theta_2^* )</td>
<td>( \theta_2^* )</td>
</tr>
<tr>
<td></td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
</tr>
<tr>
<td>{x_1, \ldots, x_n}</td>
<td>( {x_1, \ldots, x_n}_N )</td>
<td>( \theta_N^* )</td>
<td>( \theta_N^* )</td>
</tr>
</tbody>
</table>

bias: \( \hat{\theta}^* - \hat{\theta} \)
variance: \( \frac{1}{N-1} \sum_{i=1}^{N} (\theta_i^* - \bar{\theta}^*)^2 \)

Example: sample mean and sample median
Let \( x_1, \ldots, x_n \) be a realization of the i.i.d. r.v.s \( X_1, \ldots, X_n \) with a c.d.f. \( F \). Consider \( \theta_F = E_F(X_i) \) and let \( \hat{\theta}_F = \bar{X} \). We know that the sample mean \( \bar{X} \) is an unbiased estimator of \( E(X_i) \). What is the bootstrap bias of the mean?
4.2. **BOOTSTRAP ESTIMATES OF BIAS, STANDARD ERROR AND MSE**

\[
\text{Bias}_θ(θ^*) = E_θ(θ^*) - \hat{θ} = E_θ(\bar{X}) - \bar{X} = 0
\]

Is the estimate of the bootstrap bias of the sample mean, \(\hat{\text{Bias}}_θ(\bar{X})\), equal to zero as well?

Let \(\{2.3\ 3.4\ 2.5\ 3.2\ 2.7\ 2.6\ 3.1\ 3.5\ 2.9\ 2.5\}\) be a sample from a population with a c.d.f. \(F\).

Here the sample mean is 2.87 and the sample median is 2.80.

<table>
<thead>
<tr>
<th>15 bootstrap samples</th>
<th>replicates of sample mean median</th>
</tr>
</thead>
<tbody>
<tr>
<td>{3.2 2.5 3.2 3.2 3.4 3.2 2.5 2.7 2.5 2.5}</td>
<td>2.89 2.95</td>
</tr>
<tr>
<td>{2.3 3.4 3.5 2.9 2.6 3.5 2.5 2.9 2.9 3.1}</td>
<td>2.96 2.90</td>
</tr>
<tr>
<td>{2.5 2.3 3.1 2.5 3.4 3.1 2.3 3.1 3.5 3.5}</td>
<td>2.93 3.10</td>
</tr>
<tr>
<td>{2.3 2.3 2.6 3.4 2.5 2.6 2.3 3.1 2.6 2.5}</td>
<td>2.62 2.55</td>
</tr>
<tr>
<td>{2.5 3.5 2.9 3.4 2.5 3.4 2.6 2.3 2.3 3.2}</td>
<td>2.86 2.75</td>
</tr>
<tr>
<td>{2.5 2.5 2.5 3.4 2.9 3.5 3.5 2.7 3.5 3.2}</td>
<td>3.02 3.05</td>
</tr>
<tr>
<td>{3.5 3.4 2.6 2.5 2.9 3.4 3.2 3.2 3.1 3.1}</td>
<td>2.98 3.00</td>
</tr>
<tr>
<td>{3.1 2.5 3.1 2.3 2.6 3.2 3.4 3.4 3.4 2.5}</td>
<td>2.88 2.90</td>
</tr>
<tr>
<td>{3.1 3.4 3.1 3.1 3.1 3.1 3.2 3.2 3.5 3.2}</td>
<td>2.95 3.10</td>
</tr>
<tr>
<td>{2.6 2.5 2.3 2.5 2.3 2.6 3.2 3.2 3.2 3.2}</td>
<td>2.57 2.50</td>
</tr>
<tr>
<td>{2.9 3.5 2.9 2.3 2.7 2.7 2.6 2.5 2.9 3.1}</td>
<td>2.81 2.80</td>
</tr>
<tr>
<td>{2.9 3.1 3.5 2.3 2.7 2.3 3.5 3.2 2.5 2.9}</td>
<td>2.89 2.90</td>
</tr>
<tr>
<td>{3.4 2.9 3.2 3.2 3.2 3.1 2.9 3.4 2.7 3.5 3.2}</td>
<td>3.15 3.20</td>
</tr>
<tr>
<td>{2.7 3.1 3.4 3.2 3.5 2.5 3.2 2.9 2.5 3.4}</td>
<td>3.04 3.15</td>
</tr>
<tr>
<td>{2.5 3.2 3.5 2.5 2.7 3.1 3.2 3.2 2.9 2.5 3.4}</td>
<td>2.95 3.00</td>
</tr>
</tbody>
</table>

The bootstrap estimate of the sample mean is the average of the replicates \(θ^*_i\), that is

\[
\bar{θ} = \frac{1}{15} (2.89 + 2.96 + 2.93 + \ldots + 2.95) = 2.898
\]

Hence, the estimate of bootstrap bias of the sample mean is

\[
\hat{\text{Bias}}_θ(\bar{X}) = \bar{θ} - \hat{θ} = 2.898 - 2.87 = 0.098 \neq 0.
\]

So, the answer to the question if the estimate of the bootstrap bias of the sample mean, \(\hat{\text{Bias}}_θ(\bar{X})\), is equal to zero, is GENERALLY NOT.
Similar calculations for the above data and the bootstrap samples show that the estimate of the bootstrap bias for the sample median is

\[
\hat{\text{Bias}}_{\theta}(\theta^*) = \bar{\theta}^* - \hat{\theta} = 2.916 - 2.80 = 0.116
\]

However, we do not know what is the true bias of the sample median, so we do not know how good this estimate is.

The following question arises: How big should the bootstrap sample be to get a high probability that the estimate of the bootstrap bias of an estimator is close to the true value of the bias (known or unknown)?

The Central Limit Theorem says that the distribution of an average is approximately normal if the sample size is large and the variance is finite. Applying it to the bootstrap replicates \( \theta^*_i \) we get

\[
P \left( \left| \bar{\theta}^* - E_{\hat{\theta}}(\theta^*) \right| < 2 \frac{\hat{\text{se}}_\theta(\theta^*)}{\sqrt{N}} \right) \approx 0.95. \tag{4.5}
\]

This means

\[
P \left( \left| \hat{\text{Bias}}_{\theta}(\theta^*) - \text{Bias}_{\theta}(\theta^*) \right| < 2 \frac{\hat{\text{se}}_\theta(\theta^*)}{\sqrt{N}} \right) \approx 0.95.
\]

Hence, we need \( N \) such that \( 2 \frac{\hat{\text{se}}_\theta(\theta^*)}{\sqrt{N}} \) is very small, for example 0.001.

**Example: The patch data** (Efron, Tibshirani, 1993, page 127)

Eight subjects wore medical patches designed to increase the blood levels of a certain natural hormone. Each subject had his blood levels of the hormone measured after wearing three different patches: a placebo patch, which had no medicine in it, an old patch which was from a lot manufactured at an old plant, and a new patch, which was from a lot manufactured at a newly opened plant. The purpose of the experiment was to show that the new plant was producing patches equivalent to those from the old plant. The observations are in the table below.
4.2. Bootstrap Estimates of Bias, Standard Error and MSE

<table>
<thead>
<tr>
<th>subject</th>
<th>placebo ( p_i )</th>
<th>old patch ( old_i )</th>
<th>new patch ( new_i )</th>
<th>( z_i = old_i - p_i )</th>
<th>( y_i = new_i - old_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9243</td>
<td>17649</td>
<td>16499</td>
<td>8406</td>
<td>-1200</td>
</tr>
<tr>
<td>2</td>
<td>9671</td>
<td>12013</td>
<td>14614</td>
<td>2342</td>
<td>2601</td>
</tr>
<tr>
<td>3</td>
<td>11792</td>
<td>19979</td>
<td>17274</td>
<td>8187</td>
<td>-2705</td>
</tr>
<tr>
<td>4</td>
<td>13357</td>
<td>21816</td>
<td>23798</td>
<td>8459</td>
<td>1982</td>
</tr>
<tr>
<td>5</td>
<td>9055</td>
<td>13850</td>
<td>12560</td>
<td>4795</td>
<td>-1290</td>
</tr>
<tr>
<td>6</td>
<td>6290</td>
<td>9806</td>
<td>10157</td>
<td>3516</td>
<td>351</td>
</tr>
<tr>
<td>7</td>
<td>12412</td>
<td>17208</td>
<td>16570</td>
<td>4796</td>
<td>-638</td>
</tr>
<tr>
<td>8</td>
<td>18806</td>
<td>29044</td>
<td>26325</td>
<td>10238</td>
<td>-2719</td>
</tr>
<tr>
<td>mean:</td>
<td></td>
<td></td>
<td></td>
<td>6342</td>
<td>-452.3</td>
</tr>
</tbody>
</table>

The Food and Drug Administration (FDA) criterion for the bioequivalence is that the expected value of the new patches match that of the old patches in the sense that

\[
\frac{|E(new) - E(old)|}{E(old) - E(placebo)} \leq 0.2
\]

This means that the new patch should match the old one within 20% of the amount of hormone the old drug adds to placebo blood levels. Denote the parameter of interest by \( \theta \), i.e.,

\[
\theta = \frac{E(new) - E(old)}{E(old) - E(placebo)}
\]

and let us assume that the pairs \( x_i = (z_i, y_i) \) are realization of i.i.d. bivariate r.vs \( X_i = (Z_i, Y_i) \) with unknown c.d.f. \( F \). Then the parameter \( \theta \) is

\[
\theta = \frac{E_F(Y)}{E_F(Z)}.
\]

The natural estimator of \( \theta \) is

\[
\hat{\theta} = \frac{Y}{Z}
\]

and its value for the given observations is

\[
\hat{\theta} = \frac{-452.3}{6342} = -0.0713
\]

whose absolute value is less then 0.2.

GenStat program doing the bootstrap calculations will be shown during lectures.
40  

CHAPTER 4. RESAMPLING METHODS

4.3 The Jackknife

The Jackknife is one of the oldest resampling methods. Here we get replications of an estimator \( \hat{\theta} \) by constructing new samples simply omitting one observation at a time. So, we get \( n \) samples of size \( n - 1 \). Here is the procedure:

| Empirical distribution | Jackknife samples of size \( n - 1 \) | Jackknife replications of \( \hat{\theta} \) | Jackknife estimates of
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{F} )</td>
<td>( {x_2, x_3, \ldots, x_n}^* )</td>
<td>( \theta^*_1 )</td>
<td></td>
</tr>
<tr>
<td>( {x_1, \ldots, x_n} )</td>
<td>( {x_1, x_3, \ldots, x_n}^* )</td>
<td>( \theta^*_2 )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \vdots )</td>
<td>( \vdots )</td>
<td>( \vdots )</td>
</tr>
<tr>
<td>{x_1, \ldots, x_{n-1}}^*</td>
<td>( {x_1, \ldots, x_{n-1}}^* )</td>
<td>( \theta^*_n )</td>
<td>( \theta^*_J )</td>
</tr>
</tbody>
</table>

Here, \( \theta^*_i \) is calculated in the same way as \( \hat{\theta} \) except that the \( i \)-th observation is omitted. The Jackknife estimator of bias and variance of \( \hat{\theta} \) are defined to be:

- Bias:
  \[
  \text{Bias}_{\hat{\theta}}(\theta^*_\text{Jack}) = (n - 1)(\bar{\theta}^* - \hat{\theta})
  \]

- Variance:
  \[
  \text{var}_{\hat{\theta}}(\theta^*_\text{Jack}) = \frac{n - 1}{n} \sum_{i=1}^{n} (\theta^*_i - \bar{\theta}^*)^2
  \]

For simple types of \( \theta \) the Jackknife estimator can be calculated explicitly.

**Example: Jackknife estimator of the mean and of its variance**

**Mean**

Let \( \theta \) be the expected value of a r.v. \( X \) with a c.d.f. \( F \) and let the estimator of \( \theta \) be the average of a random sample of size \( n \), i.e., \( \hat{\theta} = \bar{X} \). The Jackknife replications of \( \hat{\theta} \) are calculated as:

\[
\theta^*_i = \frac{1}{n - 1} \sum_{j=1, j \neq i}^{n} X_j.
\]

Is the Jackknife estimate of bias of the mean equal to zero?

\[
\bar{\theta}^* = \frac{1}{n} \sum_{i=1}^{n} \theta^*_i = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{n - 1} \sum_{j=1, j \neq i}^{n} X_j = \frac{1}{n} \frac{1}{n - 1} \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} X_j =
\]
4.3. THE JACKKNIFE

\[
\frac{1}{n} \frac{1}{n-1} (n-1) \sum_{i=1}^{n} X_i = \bar{X} = \hat{\theta}.
\]

The Jackknife estimate of bias is

\[
\text{Bias}_{\hat{\theta}}(\theta^{\ast}_{\text{Jack}}) = (n-1)(\bar{\theta}^* - \hat{\theta}) = (n-1)(\bar{X} - \bar{X}) = 0.
\]

**Variance of the mean**

Here we calculate the Jackknife variance of the mean.

\[
\text{var}_{\hat{\theta}}(\theta^{\ast}_{\text{Jack}}) = \frac{n-1}{n} \sum_{i=1}^{n} (\theta^*_{(i)} - \bar{\theta}^*)^2 = \frac{n-1}{n} \sum_{i=1}^{n} \left( \frac{1}{n-1} \sum_{j=1,j\neq i}^{n} X_j - \bar{X} \right)^2 = \frac{n-1}{n} \sum_{i=1}^{n} \left( \frac{1}{n} - \frac{1}{n-1} (n\bar{X} - X_i) - \bar{X} \right)^2 = \frac{n-1}{n} \sum_{i=1}^{n} (\bar{X} - X_i)^2 = \frac{1}{n(n-1)} \sum_{i=1}^{n} (X_i - \bar{X})^2 = \frac{1}{n} \sigma^2.
\]

So, the Jackknife estimator of the variance of the mean is the familiar one.

**Example: Opinion survey**

An opinion survey asked a random sample of 200 people a yes/no question of whom 75 answered yes. The estimate of the population proportion \( p \) of those who would answer yes is estimated as \( \hat{p} = \frac{75}{200} = \frac{3}{8} \). A social science researcher is interested in a parameter

\[
\theta = p(1-p) = pq.
\]

A natural estimate of the parameter is

\[
\hat{\theta} = \hat{p}\hat{q} = \frac{3}{8} \frac{5}{8} \approx 0.2344.
\]

Is the estimator \( \hat{\theta} = \hat{p}\hat{q} \) biased?

The calculations will be shown during lectures.