Introduction to Bootstrap techniques with R

UGARTE, M.D.(*)

(*)DEPARTAMENTO DE ESTADÍSTICA E I. O., UNIVERSIDAD PÚBLICA DE NAVARRA, PAMPLONA, SPAIN

E-MAIL: LOLA@UNAVARRA.ES

Material from the book *Probability and Statistics with R*
by Ugarte, Militino, and Arnholt. Chapman and Hall/CRC, 2008

Outline

- Introduction
- **R** Tools
- Bootstrap Paradigm
- Bootstrap Estimate of Standard Error
- Bootstrap Estimate of Bias
- Bootstrap Confidence Intervals

Introduction

The term bootstrapping is due to Efron (1979), and is an allusion to a German legend about a Baron Münchhausen, who was able to lift himself out of a swamp by pulling himself up by his own hair.

In later versions he was using his own boot straps to pull himself out of the sea which gave rise to the term **bootstrapping**.

As improbable as it may seem, taking samples from the original data and using these **resamples** to calculate statistics can actually give more accurate answers than using the single original sample to calculate an estimate of a parameter.

Introduction -Cont.

In fact, **resampling methods require fewer assumptions than traditional parametric methods and generally give more accurate answers.**

The price to pay is that Bootstrap methods are computationally intensive techniques. However, today's computers are many times faster than those of a generation ago.

The fundamental concept in bootstrapping is the building of a sampling distribution for a particular statistic by resampling from the data that is at hand. In this sense, bootstrap methods are both parametric and nonparametric; however, attention now is focused exclusively on the nonparametric bootstrap.

Introduction -Cont.

Bootstrap methods offer the practitioner valuable tools for dealing with complex problems.

Even though resampling procedures rely on the new power of the computer to perform simulations, they are based on the old statistical principles such as populations, parameters, samples, sampling variation, pivotal quantities, and confidence intervals.

For most students, the idea of a sampling distribution for a particular statistic is completely abstract; however, once work begins with **the bootstrap distribution, the bootstrap analog to the sampling distribution, the concreteness of the bootstrap distribution promotes a conceptual understanding of the more abstract sampling distribution**.

R Tools

R is a free statistical software with many possibilities allowing the practitioner to use bootstrap techniques very easily

There exist two important R packages:

- **bootstrap** by Efron and Tibshirani (1993) (ported to **R** from **S-PLUS**® by Friedrich Leisch).
- **boot** by Angelo Canty (ported to **R** from **S-PLUS**® by B. D. Ripley)

The **boot** library provides functions and data sets from the book *Bootstrap Methods and Their Applications* by Davison and Hinkley (1997).

The Bootstrap Paradigm

Suppose a random sample $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ is taken from an unknown probability distribution, F, and the values $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ are observed.

Using $\mathbf{x}$, the parameter $\theta = t(F)$ is to be estimated.

The traditional approach of estimating θ is to make some assumptions about the population structure and to derive the sampling distribution of $\hat{\theta}$ based on these assumptions.

This, of course, assumes the derivation of the sampling distribution of the statistic of interest has either been done or that the individual who needs to do the deriving has the mathematical acumen to do so. Often, the use of the bootstrap will be preferable to extensive mathematical calculations.

Bootstrap Paradigm -Cont

In the bootstrap paradigm, the original sample, $\mathbf{x}$, takes the place the population holds in the traditional approach. Subsequently, a random sample of size n is drawn from $\mathbf{x}$ with replacement.

The resampled values are called a bootstrap sample and are denoted $\mathbf{x}^*$.

Sampling with replacement means that after we randomly draw an observation from the original sample we put it back before drawing the next observation. Think of drawing a number from a hat, then putting it back before drawing it again.

That is, given $\mathbf{x} = \{4, 5, 6, 2, 8, 12\}$, one possible bootstrap sample $\mathbf{x}^*$ might be $\mathbf{x}^* = \{6, 6, 5, 12, 2, 8\}$

Some values from the original sample $\mathbf{x}$ may appear once, more than once, or not at all in the bootstrap sample $\mathbf{x}^*$.

Bootstrap Paradigm -Cont

Remember that the star notation indicates that $\mathbf{x}^*$ is not the original data set $\mathbf{x}$, but rather, it is a random sample of size n drawn with replacement from $\mathbf{x}$.

The idea of calculating the sampling distribution of a statistic in the classical approach is to collect the values of the statistic from many samples. The bootstrap distribution of a statistic collects its values from many resamples.

These values are used to calculate an estimate of the statistic of interest $s(\mathbf{x}) = \hat{\theta}$.

Bootstrap Paradigm -Cont

The fundamental bootstrap assumption is that the sampling distribution of the statistic under the unknown probability distribution F may be approximated by the sampling distribution of $\hat{\theta}^*$ under the empirical probability distribution $\hat{F}$.

Remember that the empirical probability distribution puts probability $1/n$ for each value x_i.

Bootstrap Paradigm -Cont

The process of creating a bootstrap sample $\mathbf{x}^*$ and a bootstrap estimate $\hat{\theta}^*$ of the parameter of interest is repeated B times.

The B bootstrap estimates of θ, the $\hat{\theta}^*$s, are subsequently used to estimate specific properties of the bootstrap sampling distribution of $\hat{\theta}^*$.

There are a total of $\binom{2n-1}{n}$ distinct bootstrap samples. Yet, a reasonable estimate of the standard error of $\hat{\theta}^*$, $\hat{\sigma}_{\hat{\theta}^*} \equiv \widehat{\mathrm{SE}}_B$, can be achieved with only $B = 200$ bootstrap replications in most problems.

For confidence intervals and quantile estimation, B generally should be at least 999.

Bootstrap Estimate of Standard Error

Under the fundamental bootstrap assumption, we may write

$$\text{se}_F(\hat{\theta}) = \sqrt{\text{var}_F(\hat{\theta})} \doteq \text{se}_{\hat{F}}\hat{\theta}^*$$

The algorithm that we will describe soon will allow us to calculate a good numerical approximation of $\text{se}_{\hat{F}}\hat{\theta}^*$

Bootstrap Estimate of Standard Error

The drawing of bootstrap samples can be easily done in a computer. We just need a selection procedure of integer random numbers among 1 and n with probability $1/n$: $i_1, \ldots, i_n$.

The bootstrap sample corresponding to a single drawing is

$$x_1^* = x_{i_1}, x_2^* = x_{i_2}, \ldots, x_n^* = x_{i_n}$$

In **R** the function **sample** does this task:

```
x<-c(10, 15, 25, 37, 48, 23, 44, 19, 32, 20)
set.seed(30)  #to reproduce the same result
indices<-sample(1:10, replace=T)
indices
[1]  1  5  4  5  4  2  9  3 10  2
x.asterisco<-x[indices]
x.asterisco
[1] 10 48 37 48 37 15 32 25 20 15
```

Also (and easier)

```
set.seed(30)
sample(c(10, 15, 25, 37, 48, 23, 44, 19, 32, 20), replace=T)
 [1] 10 48 37 48 37 15 32 25 20 15
```

Bootstrap Estimate of Standard Error

- The bootstrap algorithm works drawing independent bootstrap samples, and calculating the corresponding statistic using these samples. **The bootstrap standard error of a statistic is the standard deviation of the bootstrap distribution of that statistic.**

- The result is called bootstrap standard error and it is denoted by $\hat{\text{se}}_B$, where B is the number of replications.

- To apply the bootstrap idea we must start with a statistic that estimates the parameter we are interested in. We usually come up with a suitable statistic by appealing to another principle that we often apply without thinking: **the plug-in principle**

THE PLUG-IN PRINCIPLE

To estimate a parameter, a quantity that describes the population, use a statistic that is the corresponding quantity for the sample.

- For example, to estimate μ we use $\bar{x}$ or to estimate the population standard deviation σ, we use the sample standard deviation s.

- **The bootstrap idea itself is a form of the plug-in principle:** substitute the sample data for the population, then draw samples (resamples) to mimic the process of building a sampling distribution.

Bootstrap Estimate of Standard Error

The general procedure for estimating the standard error of $\hat{\theta}^*$ is:

1. Generate B independent bootstrap samples $\mathbf{x}^{*1}, \mathbf{x}^{*2}, \ldots \mathbf{x}^{*B}$, each consisting of n values drawn with replacement from the original sample.

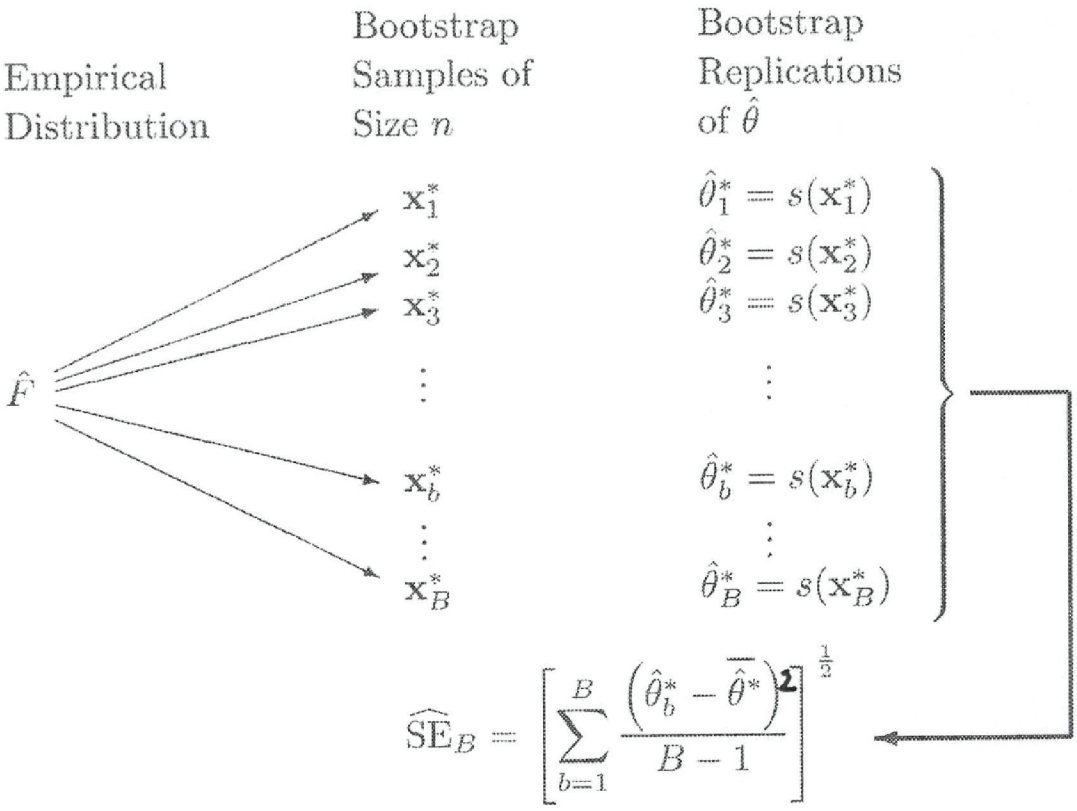
2. Compute the statistic of interest for each bootstrap sample b.

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

3. Estimate the standard error of $\hat{\theta}$ by computing the sample standard deviation of the bootstrap replications of $\hat{\theta}^*_b, b = 1, 2, \ldots, B$.

$$\hat{\text{se}}_B = \left\{ \sum_{i=1}^{B} \frac{(\hat{\theta}^*_b - \bar{\hat{\theta}}^*)^2}{(B-1)} \right\}^{1/2}, \quad \text{where } \bar{\hat{\theta}}^* = \sum_{b=1}^{B} \frac{\hat{\theta}^*_b}{B}$$

FIGURE 10.16: Graphical representation of the bootstrap



Bootstrap Estimate of Standard Error

- The number of replications needed to calculate the bootstrap standard error is rarely superior to 200 (Efron and Tibshirani, 1993)
- The limit of $\hat{\text{se}}_B$ when B goes to infinity is the ideal bootstrap estimate of $\text{se}_F(\hat{\theta})$.
- The fact that $\hat{\text{se}}_B$ is approximately equal to $\text{se}_{\hat{F}}(\hat{\theta}^*)$ when B goes to infinity is similar to saying that the empirical standard deviation is approximately equal to the population standard deviation when the number of replications increases.
- The ideal bootstrap estimate $\text{se}_{\hat{F}}(\hat{\theta}^*)$ and its numerical approximation $\hat{\text{se}}_B$ are called non-parametric bootstrap estimates because they are based on $\hat{F}$, a non-parametric estimator of F.

Bootstrap Standard Error- Example

We generate 10 values from a $N(0,1)$. Think in the sample mean $\bar{X}$ as an estimator of μ (in this case the true value of μ is known and equal to 0). We know theoretically that the standard error of the sample mean estimator is $\text{ee}(\bar{X}) = \sqrt{\sigma^2/n}$ and the corresponding estimator of this standard error is $\hat{\text{ee}}(\bar{X}) = \sqrt{S^2/n}$. Then, the true standard error of $\bar{X}$ is $\sqrt{1/10} = 0{,}3162$.
Let us calculate a bootstrap numerical approximation using B=200 replicates.

The student should repeat this little experiment generating a sample of size 100. Fix the seed in 10 using the command **set.seed(10)** to be able to reproduce results.

SOLUTION

```
set.seed(10)
n<-10
muestra.original<-rnorm(n)
muestra.original
> muestra.original
 [1]  0.01874617 -0.18425254 -1.37133055 -0.59916772  0.29454513  0.38979430
 [7] -1.20807618 -0.36367602 -1.62667268 -0.25647839
sqrt(var(muestra.original)/n) # With theoretical results
[1] 0.2213258
B<-200
muestras.bootstrap<-matrix(0,B,n)
estadistico.boot<-array(0,B)
i<-1
while (i < (B+1)){
muestras.bootstrap[i,]<-sample(muestra.original,replace=T)
estadistico.boot[i]<-mean(muestras.bootstrap[i,])
i<-i+1}
error.estandar<-sd(estadistico.boot)
> error.estandar
[1] 0.2059542
```

Bootstrap Estimate of Bias

A statistic is biased as an estimate of a population parameter if its sampling distribution is not centered at the true value of the parameter.

We may check bias by seeing whether the bootstrap distribution of the statistic is centered at the value of the statistic for the original sample.

More precisely, the bias of $\hat{\theta} = s(\mathbf{X})$ is the difference between the expected value of $\hat{\theta}$ and the true parameter value $\theta = t(F)$.

$$\text{Bias}(s(\mathbf{X})|F) = E_F[s(\mathbf{X})] - t(F)$$

We may use bootstrap to estimate the bias of any estimator $\hat{\theta}$ by writing in the previous expression $\hat{F}$ instead of F

$$\text{Bias}(s(\mathbf{X})|\hat{F}) = E_{\hat{F}}[s(\mathbf{X}^*)] - t(\hat{F})$$

In words, the bootstrap estimate of bias is the difference between the mean of the bootstrap distribution and the value of the statistic in the original sample

We will calculate the bootstrap bias of $s(\mathbf{X})$ using B resamples of the original sample

$$\widehat{\text{Bias}}_B[s(\mathbf{X})] = \bar{\hat{\theta}}^* - \hat{\theta} \quad \text{donde} \quad \bar{\hat{\theta}}^* = \sum_{i=1}^{B} \frac{\hat{\theta}^*}{B}$$

Bootstrap Confidence Intervals

- With estimates of the standard error (standard deviation) and bias of some statistic of interest, various types of confidence intervals for the parameter θ can be constructed.
- Although exact confidence intervals for specific problems can be computed, most confidence intervals are approximate.
- The most common confidence interval for a parameter θ when $\hat{\theta}$ follows either a normal or approximately normal distribution is

$$C.I._{1-\alpha}(\theta) = [\hat{\theta} - z_{1-\alpha/2}\hat{\sigma}_{\hat{\theta}}, \hat{\theta} + z_{1-\alpha/2}\hat{\sigma}_{\hat{\theta}}]$$

The student may remember that this interval is easily obtained from the distribution of $\hat{\theta} - \theta$ using the pivotal quantity

$$\frac{\hat{\theta} - \theta}{\sqrt{\hat{\text{var}}(\hat{\theta})}} \approx N(0, 1)$$

- The confidence interval described above works well when the distribution of $\hat{\theta} - \theta$ is normal, at least approximately, but this is not always the case. In some cases we could know the approximate normality but we may have difficulties deriving the variance of the estimator.

Bootstrap Confidence Intervals

- **The confidence interval that we will call normal** is a slight modification to the traditional CI that incorporates both a bootstrap adjustment for bias and a bootstrap estimate of the standard error. The normal confidence interval is calculated as

$$C.I._{1-\alpha}(\theta) = [\hat{\theta} - \widehat{\text{Bias}}_B(\hat{\theta}) - z_{1-\alpha/2}\widehat{\text{SE}}_B(\hat{\theta}), \hat{\theta} - \widehat{\text{Bias}}_B(\hat{\theta}) + z_{1-\alpha/2}\widehat{\text{SE}}_B(\hat{\theta})]$$

- To use this confidence interval it is convenient to check normality, using for instance a qq-norm of $\hat{\theta}^*_1 = s(\mathbf{x}^*_1), \ldots, \hat{\theta}^*_B = s(\mathbf{x}^*_B)$.

Bootstrap Confidence Intervals

- **The basic bootstrap confidence interval** is based on the idea that the quantity $\hat{\theta}^* - \hat{\theta}$ has roughly the same distribution as $\hat{\theta} - \theta$, and then it is possible to approximate the percentiles of $\hat{\theta} - \theta$ by the percentiles of $\hat{\theta}^* - \hat{\theta}$

$$P\left[\hat{\theta}^*_{((B+1)\alpha/2)} - \hat{\theta} \leq \hat{\theta}^* - \hat{\theta} \leq \hat{\theta}^*_{((B+1)(1-\alpha/2))} - \hat{\theta}\right] \doteq 1 - \alpha$$

$$P\left[\hat{\theta}^*_{((B+1)\alpha/2)} - \hat{\theta} \leq \hat{\theta} - \theta \leq \hat{\theta}^*_{((B+1)(1-\alpha/2))} - \hat{\theta}\right] \doteq 1 - \alpha$$

And then,

$$P\left[2\hat{\theta} - \hat{\theta}^*_{((B+1)(1-\alpha/2))} \leq \theta \leq 2\hat{\theta} - \hat{\theta}^*_{((B+1)\alpha/2)}\right] \doteq 1 - \alpha$$

$$I.C._{1-\alpha}(\theta) = [2\hat{\theta} - \hat{\theta}^*_{((B+1)(1-\alpha/2))}, 2\hat{\theta} - \hat{\theta}^*_{((B+1)\alpha/2)}]$$

Bootstrap Confidence Intervals

- **The bootstrap-t interval or Studentized interval** is based on the idea of replacing the approximation of $Z = \frac{\hat{\theta}-\theta}{\hat{\sigma}_{\hat{\theta}}}$ to the standard normal distribution $N(0,1)$ by a bootstrap approximation $Z^* = (\hat{\theta}^* - \hat{\theta})/SE_B$.
- We generate B bootstrap samples and compute $Z^*(b)$. The p percentile of the Z distribution is approximated by the $(B+1)p$ percentile of $Z^*(b)$.
- The interval takes the form

$$C.I._{1-\alpha}(\theta) = [\hat{\theta} + z^*_{((B+1)(\alpha/2))}\hat{\sigma}_{\hat{\theta}}, \hat{\theta} + z^*_{((B+1)(1-\alpha/2))}\hat{\sigma}_{\hat{\theta}}]$$

- The notation $z^*_{(\text{Integer})}$ is used to denote the $(\text{Integer})^{th}$ z^* of the B sorted Z^* values. The values of B and α are generally chosen so that $(B+1)\cdot\alpha/2$ is an integer. In cases where $(B+1)\cdot\alpha/2$ is not an integer, interpolation can be used. (Note that different programs use different interpolation techniques.)

Bootstrap Confidence Intervals

- The **percentile** confidence interval is based on the quantiles of the B bootstrap replications of $s(\mathbf{X})$.

  Specifically, the $(1-\alpha)$ percentile confidence interval of θ uses the $\alpha/2$ and the $1-\alpha/2$ quantiles of the $\hat{\theta}^*$ values to create a $(1-\alpha)\cdot 100\,\%$ confidence interval for θ.

$$C.I._{1-\alpha} = [\hat{\theta}^*_{((B+1)\alpha/2)}, \hat{\theta}^*_{((B+1)(1-\alpha/2))}]$$

Bootstrap Confidence Intervals

- At this point, a reasonable question might be **which confidence interval is recommended for general usage** since the normal confidence interval is based on large sample properties, the t-bootstrap confidence interval is not recommended if the bootstrap distribution is not normal or shows substantial bias, and the percentile and basic bootstrap confidence interval formulas give different answers when the distribution of $\hat{\theta}^*$ is skewed?
- In fact, the answer is to use *none* of the confidence intervals discussed thus far. The bootstrap confidence interval procedure recommended for general usage is the BC_a method, which stands for bias-corrected and accelerated.
- The bottom line is that there are theoretical reasons to prefer the BC_a confidence interval over the normal, percentile, and basic bootstrap confidence intervals.

- It is possible that both the percentile and basic methods provide confidence intervals not centered around the true value of the parameter. Only if the bootstrap distribution is symmetric around $\hat{\theta}$, all of the methods provide the same results. Otherwise, the BC_a CI corrects for bias and skewness.

- The underlying idea of the BC_a CI is to assume that there exist a transformation of $\hat{\theta}$ whose distribution is normal and its mean and standard error depend on θ. Then, one derives an interval of the transformed parameter and then back-transformed the confidence limits to obtain an interval for θ.

- **The most interesting thing is that it is possible to calculate the interval without knowing the explicit form of the transformation by using bootstrap.**

Bootstrap Confidence Intervals

To compute a BC_a interval for θ, $C.I._{1-\alpha}(\theta) = \left[\hat{\theta}^*_{\text{lower}}, \hat{\theta}^*_{\text{upper}}\right]$, first compute the bias factor, z, where

$$z = \mathbf{\Phi}^{-1}\left[\frac{\sum_{i=1}^{B}\mathbf{I}\{\hat{\theta}^*_b < \hat{\theta}\}}{B}\right]$$

$\mathbf{\Phi}^{-1}$ is the inverse of the cumulative distribution function of the standard normal distribution and $\mathbf{I}$ is the indicator function.

- Provided the estimated bootstrap distribution of $\hat{\theta}^*$ is symmetric with respect to $\hat{\theta}$, and if $\hat{\theta}$ is unbiased, then $\frac{\sum_{b=1}^{B}\mathbf{I}\left\{\hat{\theta}^*_b < \hat{\theta}\right\}}{B}$ will be close to $0{,}5$, and the biased correction factor z will be close to zero since $\mathbf{\Phi}^{-1}(0{,}5) = 0$.

- Using $\mathbf{R}$, type qnorm(.5)=0.

Bootstrap Confidence Intervals

Next, we compute the skewness correction factor

$$a = \frac{\sum_{i=1}^{n} \left(\bar{\hat{\theta}}_{(-i)} - \hat{\theta}_{(-i)} \right)^3}{6 \left[\sum_{i=1}^{n} (\bar{\hat{\theta}}_{(-i)} - \hat{\theta}_{(-i)})^2 \right]^{3/2}},$$

where $\hat{\theta}_{(-i)}$ is the value of $\hat{\theta} = s(\mathbf{X})$ when the i-th value is deleted from the sample of n values and $\bar{\hat{\theta}}_{(-i)} = \sum_{i=1}^{n} \frac{\hat{\theta}_{(-i)}}{n}$.

Using z and a, we compute

$$a_1 = \mathbf{\Phi} \left[z + \frac{z + z_{\alpha/2}}{1 - a(z + z_{\alpha/2})} \right] \quad \text{and} \quad a_2 = \mathbf{\Phi} \left[z + \frac{z + z_{1-\alpha/2}}{1 - a(z + z_{1-\alpha/2})} \right]$$

The BC_a confidence interval for θ is

$$C.I._{1-\alpha}(\theta) = [\hat{\theta}^*_{((B+1)a_1)}, \hat{\theta}^*_{((B+1)a_2)}]$$

When either lower= $((B+1)a_1$ or upper=$((B+1)a_2$ is not an integer, interpolation can be used to obtain the lower and upper endpoints of the BC_a confidence interval.

The function boot.ci of the package boot computes all of the confidence intervals just shown.

Example

The times recorded are those for 41 successive vehicles travelling northwards along the M1 motorway in England when passing a fixed point near Junction 13 in Bedfordshire on Saturday, March 23, 1985. After subtracting the times, the following 40 interarrival times were reported to the nearest second.

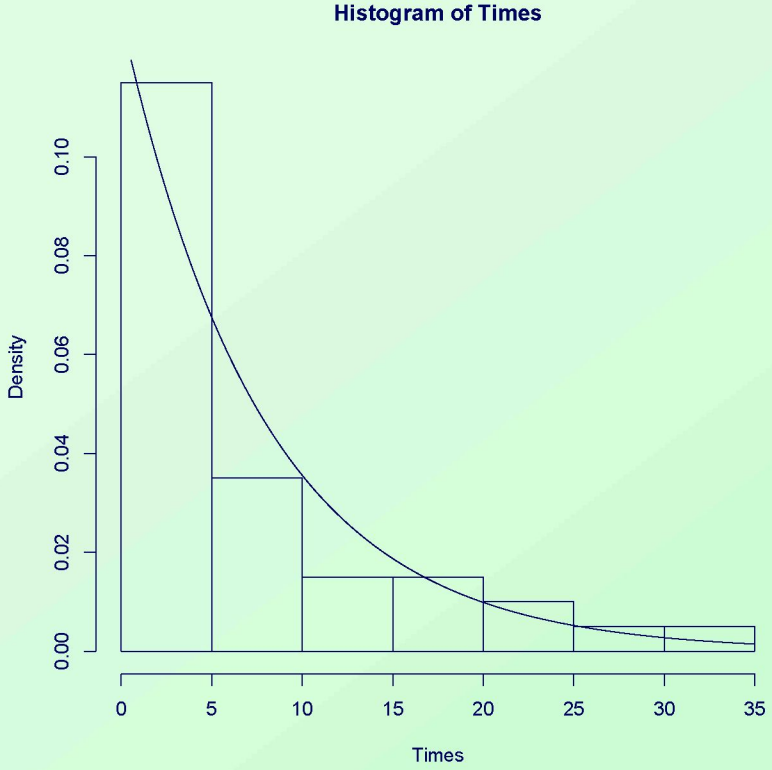
```
Times<-c(12,2,6,2,19,5,34,4,1,4,8,7,1,21,6,11,8,28,6,
    4,5,1,18,9,5,1,21,1,1,5,3,14,5,3,4,5,1,3,16,2)
```

Determine the distribution of the interarrival times and calculate bootstrap confidence intervals for the mean of those times using the function **boot.ci** from the **boot** library. In addition, compute an exact confidence interval for the mean knowing that $2n\bar{X}/\theta \sim \chi^2_{2n}$ where θ is the mean of the exponential distribution. What type of confidence interval is closer to the exact interval?

Solution

To determine the distribution of interarrival times, a histogram is created and the mean and standard deviation are calculated.

```
> hist(Times,prob=T)
> mean(Times)
[1] 7.8
> sd(Times)
[1] 7.871402
> lamb<-1/mean(Times)
> x<-seq(0,35,length=800)
> f<-lamb*exp(-lamb*x)
> lines(x,f,lwd=1)
```


Histogram of Times
Density
0.00
0.02
0.04
0.06
0.08
0.10
0
5
10
15
20
25
30
35
Times

Example

- It seems an approximate Poisson process for the number of cars passing Junction 13 Saturday 23 March 1985 with $\bar{x} = 1/\hat{\lambda} = 7{,}8$ is present.
- Recall that the waiting time between outcomes in a Poisson process has an exponential distribution. Note that the interarrival times seems to be fit well with an exponential density with $\hat{\lambda} = 1/7{,}8$.
- Recall that mean and standard deviation of the exponential are $1/\lambda$.

Example

The bootstrap confidence intervals are now constructed for the mean using the function **boot**.

To use the package **boot** we need first to build the following function

```
library(boot)
times.fun <- function(data, i)
{    media <- mean(data[i]) # compute the mean of each bootstrap sample
     n <- length(i)
     v <- (n-1)*var(data[i])/n^2 # compute the variance of the sample mean
                                 it is needed only for the t-bootstrap CI
     c(media, v)
}
```

Example -Cont.

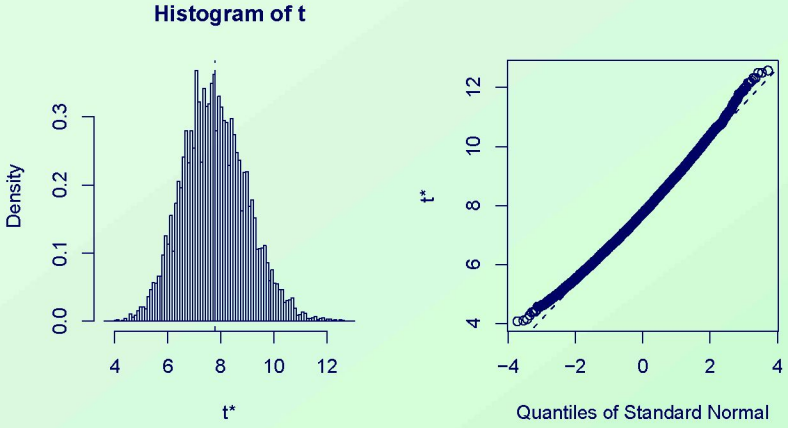
Set the number of bootstrap replications B to 9999 and generate the bootstrap distribution of $\bar{X}$ denoted by t^* when using the **boot** package. Note that R in **boot** is the number of bootstrap replications which we denoted B before, so R is set equal to B. A random seed value of 10 is used so the reader can reproduce the results.

```
B<-9999
set.seed(10)
b.obj<-boot(Times, times.fun, R=B)
> b.obj
ORDINARY NONPARAMETRIC BOOTSTRAP
Call:
boot(data = Times, statistic = times.fun, R = B)
Bootstrap Statistics :
    original      bias    std. error
t1*  7.80000 -0.01249375   1.2149888 #mean
t2*  1.51025 -0.04141159   0.4712177 #variance of the mean
```

Example -Cont.

- Let us represent now the bootstrap distribution of the mean (t_1^*) (we will observe from the histogram and qq-norm that the distribution is slightly skew to the right)
- Hence, there will be small differences between the alternative confidence intervals
- Type in **R**:

```
plot(b.obj)
```


Histogram of t
Density
0.0
0.1
0.2
0.3
4
6
8
10
12
t*
t*
4
6
8
10
12
−4
−2
0
2
4
Quantiles of Standard Normal

Example -Cont.

Next, use the function **boot.ci** on the object **b.obj** to create the five types of bootstrapped confidence intervals.

```
 > boot.ci(b.obj)
BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
Based on 9999 bootstrap replicates

CALL :
boot.ci(boot.out = b.obj)

Intervals :
Level      Normal              Basic             Studentized
95%   ( 5.431, 10.194 )   ( 5.275, 10.050 )   ( 5.681, 11.070 )

Level     Percentile            BCa
95%   ( 5.550, 10.325 )   ( 5.800, 10.700 )
Calculations and Intervals on Original Scale
```

Exact Confidence Interval

From $\frac{2n\bar{X}}{\theta} \sim \chi^2_{2n}$, we may write

$$P\left(\chi^2_{2n;\alpha/2} \leq \frac{2n\bar{X}}{\theta} \leq \chi^2_{2n;1-\alpha/2}\right) = 1-\alpha$$

From the above expression, it follows

$$P\left(\chi^2_{2n;\alpha/2} \geq \frac{\theta}{2n\bar{X}} \geq \chi^2_{2n;1-\alpha/2}\right) = 1-\alpha$$

Then a $1-\alpha$ confidence interval for θ is

$$C.I._{1-\alpha}(\theta) = \left[\frac{2n\bar{X}}{\chi^2_{2n;1-\alpha/2}}, \frac{2n\bar{X}}{\chi^2_{2n;\alpha/2}}\right]$$

Exact Confidence Interval

In **R**:

```
> lower<-2*length(Times)*mean(Times)/qchisq( 0.975, 2*length(Times))
> upper<-2*length(Times)*mean(Times)/qchisq( 0.025, 2*length(Times))
> intervalo<-round(c(lower,upper),3)
> intervalo
[1]  5.852 10.918
```

Comparison of Results

Method	Lower Limit	Upper Limit
Normal	5.431	10.194
Basic	5.275	10.050
t or Studentized	5.681	11.070
Percentile	5.550	10.325
BC_a	5.800	10.700
Exact	5.852	10.918

More Bibliography

Books:

An Introduction to Bootstrap, B. Efron and R. J. Tibshirani, Chapman and Hall, 1998.

Bootstrap Methods and Their Application, A. Davidson and D. Hinkley, Cambridge University Press, 1997.

Randomization, Bootstrapping, and Monte Carlo Methods in Biology, B. Manly, Chapman and Hall, 1997.