ETH
Swiss Federal Institute of Technology Zurich

Seminar for Statistics

Department of Mathematics

Master Thesis Autumn 2019

**Eufemiano Fuentes Pérez**

# Review of bootstrap principles and coverage analysis of bootstrap confidence intervals for common estimators

Submission Date: 17 December 2019

Adviser: Dr. Markus Kalisch

To my classmates, from whom I have learnt so much and with whom I have gone through one of the most beautiful stages of my life, and especially to Christopher Salahub and Lorenz Herger, who were always there to give me a hand when I needed it.

To Markus Kalisch, coordinator of the MSc Statistics and supervisor of this master thesis, who has always been friendly, approachable and a great guidance during my thesis.

To all the people that helped me adapt to Zürich during these past two years, especially my flatmates.

To my family, for their love and unwavering support from the distance during all my years as a student.

And especially to my grandmother Clorinda, who recently passed away, and to my grandfather Sixto, for being both always an infinite source of love and for being the most hard-working, humble people of the utmost integrity I have known. The best role models to look up to.

To all of you, thank you.

# Abstract

The bootstrap is a statistical technique that has been around for 40 years, since it was introduced by Efron (1979). Its use in practice is widespread, but many practitioners do not fully understand its limits, and under which circumstances it works or does not work. This thesis tries to address that, first by diving into the theoretical underpinnings of the bootstrap and then by analysing its performance in some scenarios in practice. Chapter 1 starts with an introduction to the bootstrap, its fundamental principles and how it works. Chapter 2 gets into when the bootstrap works (consistency) and when it does not, and if it works at which rate it does so (accuracy). Chapter 3 explores a number of first- and second-order accurate bootstrap confidence intervals, introduces a general technique to improve bootstrap intervals called *double bootstrap* and presents the two most well-known R packages to implement the bootstrap. Chapter 4 illustrates the results from a coverage analysis of bootstrap intervals in different scenarios for the sample mean, sample median and sample (Pearson) correlation coefficient, computed via simulations. Chapter 5 makes a summary of the thesis and presents a list of conclusions. Finally, Chapter 6 outlines interesting points and topics we wanted to explore further but had no time for them.

# Contents

# List of Figures

# List of Tables

# Notation

| | |
|---|---|
| $X^{**}$ | A second-layer bootstrap replicate of original sample X |
| $X^{*b}$ | $b$-th bootstrap replicate of original sample X |
| $X^{*}$ | A bootstrap replicate of original sample X |
| $\mathbb{1}\{A\}$ | Indicator function for event A |
| $\mathbb{E}(X)$ | Expected value of random variable X |
| $\mathcal{O}_{\mathbb{P}}$ | Big O in probability notation |
| $\mathcal{O}$ | Big O notation |
| $\mathbb{P}_{F}$ | Probability measure under distribution $F$ |
| $\Phi$ | Standard normal CDF |
| $\mathrm{Var}(X)$ | Variance of random variable X |
| $\xrightarrow{a.s.}$ | Almost sure convergence |
| $\xrightarrow{D}$ | Convergence in distribution |
| $\xrightarrow{\mathbb{P}}$ | Convergence in probability |
| $\phi$ | Standard normal PDF |
| $\sigma_{F}^{2}$ | Variance of population distribution $F$ |
| $\mathrm{Skew}(X)$ | Skewness of random variable X |
| $\hat{\theta}^{**}$ | A second-layer bootstrap replicate of original sample X |
| $\hat{\theta}^{*b}$ | $b$-th bootstrap replicate of original sample X |
| $\hat{\theta}^{*}$ | A bootstrap replicate of original sample X |
| $o_{\mathbb{P}}$ | Small O in probability notation |
| $q_F(\alpha)$ | $\alpha$-quantile of distribution F |

| | |
|---|---|
| $s^2$ | Unbiased sample variance of original sample X |
| CDF | Cumulative Distribution Function |
| CLT | Central Limit Theorem |
| ML | Maximum Likelihood |
| PDF | Probability Distribution Function |
| iid or i.i.d. | Indepentent and Identically Distributed |

# Chapter 1

# Introduction to the Bootstrap

The bootstrap was first introduced by Efron (1979) in his paper "Bootstrap Methods: Another Look at the Jackknife". He got inspired by earlier work on the jackknife, as he was trying to find a better method for estimating the variance of a statistic of interest. It is one of the most widely used statistical methods, and its main use is to estimate properties of an estimator $\widehat{\theta}$ of $\theta$ by sampling from an approximating distribution for the population distribution.

The following paragraph by J. Geyer (2017) gives a good explanation on why the term Bootstrap is so controversial —for some people it suggests that it "creates something out of nothing"— and what the real purpose of the name is.

> The term "bootstrap" recalls the English idiom "pull oneself up by one's bootstrap". The literal meaning of "bootstrap" in non-technical language is: leather loops at the top of boots used to pull them on. So the literal meaning of "pull oneself up by one's bootstraps" is to reach down, grab your shoes, and lift yourself off the ground — a physical impossibility. But, idiomatically, it doesn't mean do the physically impossible; it means something like "succeed by one's own efforts", especially when this is difficult. The technical meaning in statistics plays off this idiom. It means to get a good approximation to the sampling distribution of an estimator without using any theory (at least not using any theory in the computation).

Historically, when one wanted to construct a confidence interval for a parameter $\theta$ of a population, one had to rely on classical frequentist statistical theory. A sample of independent and identically-distributed (iid) observations $X_1, X_2, ..., X_n$ is taken from the population $F$, and then one tries to do inference based on the sample to get an estimate for the parameter $\theta$. In frequentist statistics, one constructs an estimator $\widehat{\theta}$ for $\theta$, which is itself a random variable with its own distribution. The goal is then to use the distribution of $\widehat{\theta}$ to construct a confidence interval for $\theta$ with a desired confidence level.

The distribution of the estimator $\widehat{\theta}$ might be known theoretically, making the compu-

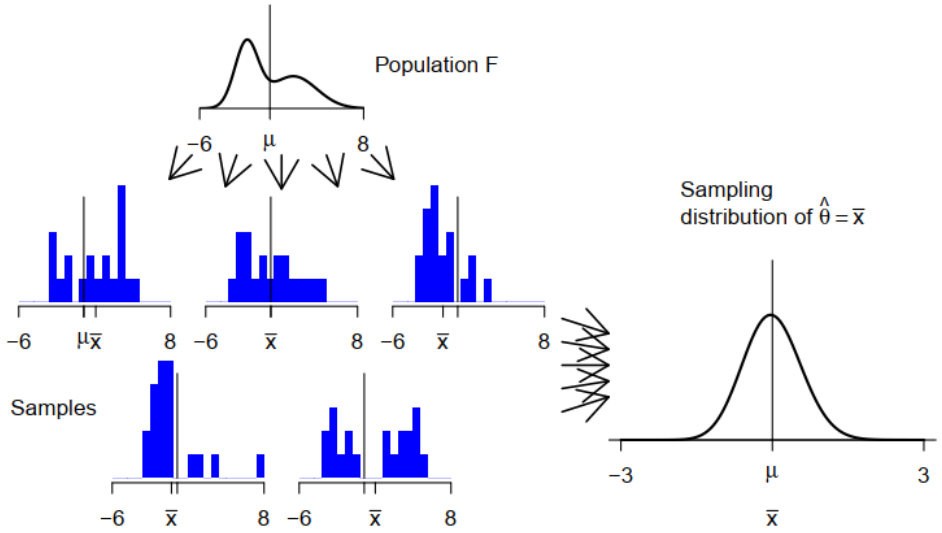


Figure 1.1: Ideal world (Hesterberg, 2015). Sampling distributions are obtained by drawing repeated samples from the population, computing the statistic of interest for each, and collecting (an infinite number of) those statistics to form the sampling distribution.

tation of confidence intervals straightforward. If it is unknown, many times we can approximate the distribution of $\widehat{\theta}$ with an asymptotic distribution. We then hope for the error between the actual unknown distribution and the approximate known distribution to be small for the data available.

Ideally, one would get an infinite number of samples from the population $F$, compute the value of $\widehat{\theta}$ for each sample and use all those computed values to estimate the distribution of $\widehat{\theta}$. This type of computation is usually known as Monte Carlo simulation. Figure 1.1 shows the process for the sample average $\bar{X}$ as an estimator for the true mean $\mu$ of the population $F$. The problem is that it is almost never possible to draw an infinite or very large number of samples, either because the population distribution is not known or because it is too expensive.

## 1.1 Bootstrap Principles

The main idea behind the bootstrap is, in fact, very simple. First, it computes an estimate $\widehat{F}$ of the original population $F$ given the sample data $X_1, X_2, ..., X_n$. Second, it draws samples $X^*$ from $\widehat{F}$ as if it was $F$, known as bootstrap samples. Third, it evaluates the value of the estimator $\widehat{\theta}$ at each bootstrap sample, yielding bootstrap estimates $\widehat{\theta}^*$. These bootstrap estimates $\widehat{\theta}^*$ of $\widehat{\theta}$ make up the bootstrap distribution of $\widehat{F}$, which is used to compute bootstrap confidence intervals for the parameter $\theta$. This process is

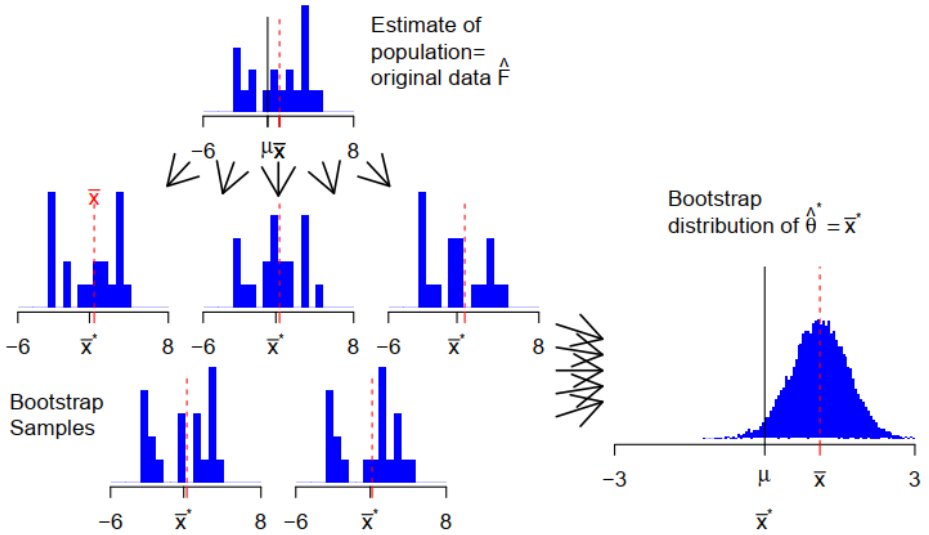


Figure 1.2: Bootstrap world (Hesterberg, 2015). The bootstrap distribution is obtained by drawing repeated samples from an estimate $\widehat{F}$ of the population, computing the statistic of interest for each, and collecting those statistics. The bootstrap distribution is centred at the observed statistic ($\bar{X}$), not the parameter ($\mu$).

illustrated in Figure 1.2.

Figure 1.3 shows what Efron calls the "Real World" and the "Bootstrap World". The first describes the estimation process of the distribution of $\widehat{\theta}$ given access to the original population's probability model $P$. The second describes the estimation process, but using the estimated probability model $\widehat{P}$ given the data. As pointed out by Efron, the only inference step is the "travel" from the "Real World" to the "Bootstrap World" illustrated by the double arrow. All other arrows on the right are exact analogs of those on the left. This "travel" is done via the so-called plug-in principle.

**Definition 1.1.** (Plug-In Principle) If we want to estimate a functional $J(\cdot)$ (usually a parameter) of an unknown distribution $F$, the *plug-in principle* states that a reasonable estimator $\hat{J}$ evaluates the functional $J(\cdot)$ using the empirical distribution $F_n$ of the sample data. That is, estimate

$$J = J(F) \tag{1.1}$$

by using

$$\hat{J} = J(F_n) \tag{1.2}$$

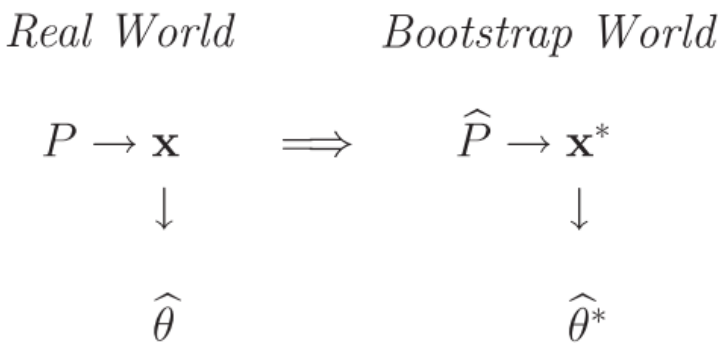


Figure 1.3: Typical bootstrap diagram (Efron, 2003). Unknown probability model $P$ gives observed data $\boldsymbol{x}$ and we wish to know the accuracy of the estimator $\widehat{\theta} = \theta(\boldsymbol{x})$ for estimating the parameter of interest $\theta = \theta(P)$. Point estimate $\widehat{P}$ for $P$ yields bootstrap data sets $\boldsymbol{x}^*$. Accuracy is inferred from observed variability of bootstrap replications $\widehat{\theta}^* = \theta(\boldsymbol{x}^*)$.

In short, *if a functional (or parameter) is unknown, then substitute an estimate for it based on the data.*

In the bootstrap literature it is common to see the use of the Greek symbol $\theta$ in place of $J$ to represent a functional, even though $\theta$ usually refers to the parameters of a distribution. So, sometimes one has the functional $\theta(\cdot)$ that gives rise to the estimator $\widehat{\theta}$ for the parameter $\theta$. To avoid that ambiguity, some prefer to use the capital letter $T$ when solely studying or referring to the functional.

Usually in the literature $\widehat{\theta}$ is used to represent the value of the estimator $\widehat{\theta}$ evaluated at the sample data, i.e. $\widehat{\theta}(x_1, \ldots, x_n)$, as well as the estimator $\widehat{\theta}$ as a random variable, i.e. $\widehat{\theta}(X_1, \ldots, X_n)$, where $x_1, \ldots, x_n$ are the specific values drawn from $F$ and $X_1, \ldots, X_n$ are the respective random variables. In this text we use $\hat{\theta}$ to refer to the former and $\widehat{\theta}$ to refer to the latter, trying to make a distinction by using different sizes of the "hat" symbol. If not specified, it should be clear from the context which one is being referred to.

As an example of the plug-in principle, one can assume sample data $X_1, \ldots, X_n$ has been iid sampled from $X$ which follows a normal distribution with mean $\mu$ and variance $\sigma^2$. Since $\mu$ and $\sigma^2$ are unknown, we can use the functionals $\bar{X} = n^{-1} \sum X_i$, $s^2 = (n-1)^{-1} \sum (X_i - \bar{X})^2$ as estimators for $\mu$ and $\sigma^2$ respectively, and then use the plug-in principle with the collected data to get an estimate of $\mu$ and $\sigma^2$. Furthermore, one can then use those estimates to estimate the original distribution $\mathcal{N}(\mu, \sigma^2)$ by using $\mathcal{N}(\bar{X}, s^2)$.

The first step of the bootstrap is based on the plug-in principle, as it tries to find an

estimate $\widehat{F}$ for the original population $F$. By the plug-in principle, we can "plug-in" our sample and get an estimate $\widehat{F}$ of the whole distribution instead of an estimate for a single parameter or functional (exemplified in the last sentence of the previous paragraph). Then we use this estimate $\widehat{F}$ as a proxy for $F$, and we sample from $\widehat{F}$ as if it was the original population $F$.

What are possible estimates $\widehat{F}$ of $F$? There are various ways in which this estimation can be done. The following two are the most popular ways to do so.

1) *Nonparametric bootstrap.* This is the most common type and the first version of the bootstrap, introduced by Efron (1979). It uses the empirical distribution function $\hat{F}_n$ directly as an estimate for $F$. Then it randomly samples from $\hat{F}_n$, which is equivalent to resampling with replacement from the original data.

2) *Parametric bootstrap.* Here it is assumed that the population distribution follows a parametric distribution $P \in \mathcal{P}$, where $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$. The data is used to estimate $\theta$ with some estimator $\widehat{\theta}$, and then $\widehat{F}(\cdot) = \widehat{P}_\theta(X \leq \cdot)$ is used as an estimate for $F$. We write $P_{\hat{\theta}} = \widehat{P}_\theta$.

There are other bootstrap variations such as the smoothed bootstrap (smoothed $\widehat{F}$ as estimate for $F$) or the block bootstrap (for correlated data, works with "blocks" of data) that we will not get into. This thesis is only concerned about the nonparametric bootstrap procedure.

Ultimately, the nonparametric bootstrap can be summarized in two basic principles stated by Hesterberg (2015):

**Fundamental Bootstrap Principle** Under certain regularity conditions, *the plug-in principle works* for the bootstrap resampling scheme.

Most of the time, the bootstrap distribution tells us something useful about the sampling distribution of an estimator $\widehat{\theta} = \theta(\widehat{F})$ for the parameter $\theta = \theta(F)$. However, there are cases in which the regularity conditions are not met and this principle does not hold. These cases are studied in the next chapter.

**Second Bootstrap Principle** The bootstrap samples are obtained by *sampling with replacement* from the original data.

As pointed out by Hesterberg, this is not that much of a principle but rather an instruction on how to implement the nonparametric bootstrap.

The bootstrap samples should also be drawn the same way they were drawn in real life, so that the right arrows in 1.3 are analogs of those on the left. That is, if the original sample was drawn using stratified sampling, the bootstrap samples must be drawn the same way. The same holds for other types of sampling such as simple random sampling or finite-population sampling.

There is another important sampling detail worth bearing in mind. Because the bootstrap mimics the way the original data was sampled, that means we have

to *sample conditioned on the observed information.* For instance, if we compare two samples of size $n_1$ and $n_2$, those numbers must be fixed, even if the original sampling process could have given different counts.

## 1.2 How Many Bootstrap Samples Are There? The Exact Bootstrap Distribution

We already know that the bootstrap first computes an estimate $\widehat{F}$ of $F$ and then it draws so-called bootstrap samples $X^*$ from $\widehat{F}$, mimicking the original data collection procedure. But how many samples can be drawn?

Consider this example. Let the sample data be $\boldsymbol{x} = \{x_1, x_2\}$. What are the possible bootstrap samples we can draw? There are 4 possibilities: $\{x_1, x_1\}, \{x_1, x_2\}, \{x_2, x_1\}$ or $\{x_2, x_2\}$. However, you can see that samples $\{x_1, x_2\}$ and $\{x_2, x_1\}$ are equivalent, since the estimation of the empirical distribution function $\hat{F}_n$ does not take into account the order. So there are 4 total bootstrap samples from which 3 are unique. If now we let the sample data be $\boldsymbol{x} = \{x_1, x_2, x_3\}$, we would get 27 bootstrap samples from which 10 are unique.

The collection of all possible bootstrap samples that can be drawn makes up the *exact bootstrap distribution.* There are $n^n$ possible bootstrap samples that can be drawn, but many of them are equivalent. We say that two bootstrap samples are equivalent if their empirical distribution functions are the same.

To understand how many different (not equivalent) bootstrap samples exist, we are going to introduce some new notation. Consider the sample data $\boldsymbol{x} = \{x_1, x_2, x_3, x_4\}$ and the specific bootstrap sample $\boldsymbol{x}^* = \{x_3, x_4, x_1, x_3\}$. We can rewrite $\boldsymbol{x}^*$ in terms of the number of occurrences of each observation from the original sample in vector form, that would be $\boldsymbol{x}^* = (1, 0, 2, 1)$. Notice that equivalent bootstrap samples would be mapped to the same vector. Hence, the new notation "maps" all possible bootstrap samples to their unique representation. The set of all different/unique bootstrap samples can be defined as

$$C_n = \left\{(k_1, \ldots, k_n),\ k_i \in \mathbb{N} \cup \{0\},\ \sum k_i = n\right\} \tag{1.3}$$

The number of different bootstrap samples is the size of $C_n$. For a vector in $C_n$, consider each component to be a box. There are $n$ boxes (number of observations), and $n$ balls (sum of occurrences of each observation, in last example $n = 1 + 0 + 2 + 1$). We want to count the number of ways in which the $n$ balls can be placed into the $n$ boxes. If we use $n - 1$ separators to make boxes, and have $n$ balls, there will be $n - 1 + n = 2n - 1$ positions from which to choose $n - 1$ separators' positions. In our previous example,

$\boldsymbol{x}^* = (1, 0, 2, 1)$ corresponds to o||oo|o, where o represents a ball and | a separator. Thus, we have that

$$|C_n| = \binom{2n-1}{n-1} \tag{1.4}$$

There is another nice interpretation followed by the representation of bootstrap samples in the vector form. We can think of bootstrap samples as vectors $\mathbf{k} = (k_1, \ldots, k_n)$ being drawn from a multinomial distribution with each of the $n$ categories being equaly likely, with $p_i = 1/n$. Then we can compute the probability of drawing a specific vector as

$$\mathbb{P}\Big(\mathbf{k} = (k_1, \ldots, k_n)\Big) = \frac{n!}{k_1! \cdots k_n!}\Big(\frac{1}{n}\Big)^{k_1+\ldots+k_n} = \binom{n}{k_1 \cdots k_n} n^{-n} \tag{1.5}$$

The largest probability corresponds to the vector of all 1's, which is the original sample. Hence, the original sample itself is the most likely to be sampled as a bootstrap sample.

Even for fairly small $n$, say $n = 10$, the total number of bootstrap samples is too large to compute ($10^{10}$). What is done in practice is to sample with replacement from the original sample a large number $B$ of bootstrap samples, and use those bootstrap samples as a representative set from all possible $n^n$ bootstrap samples. In the next section we investigate a way to define $B$ given a desired accuracy.

## 1.3 How Many Bootstrap Samples to Draw? Monte Carlo Accuracy

So far, we know that in the bootstrap setting we sample many times, say $B$, from the estimated distribution $\widehat{F}$ of the population $F$ to get bootstrap samples $X^{*(1)}, \ldots, X^{*(B)}$. Then we evaluate the estimator $\widehat{\theta}$ at every bootstrap sample $X^{*b}$, yielding bootstrap estimates $\hat{\theta}^{*(1)}, \ldots, \hat{\theta}^{*(B)}$ which we use to get an estimate of the sampling distribution of $\widehat{\theta}$.

Let

$$G(t) = \mathbb{P}(\widehat{\theta} \leq t) \tag{1.6}$$

be the CDF of the estimator $\widehat{\theta}$ for a sample of size $n$, and let

$$\widehat{G}(t) = \mathbb{P}(\widehat{\theta}^* \leq t \mid X_1, \ldots, X_n) \tag{1.7}$$

be the exact bootstrap distribution approximating $G$. The notation $\mathbb{P}(\,\cdot \mid X_1, \ldots, X_n)$ emphasizes that the bootstrap distribution is conditioned on the sample data. We usually approximate the exact bootstrap distribution $\widehat{G}$ with its Monte Carlo version

$$\widehat{G}^*(t) = \mathbb{P}^*\left(\widehat{\theta}^* \leq t \mid X_1, \ldots, X_n\right) = \frac{1}{B}\sum_{j=1}^{B} \mathbb{1}\{\hat{\theta}^{*j} \leq t\} \tag{1.8}$$

where $\hat{\theta}^{*j}$ is the evaluation of the estimator $\widehat{\theta}$ at the $j$-th bootstrap sample $X^{*(j)}$.

There is still one important question we have not addressed yet. How many times should we resample to get an accurate estimate $\widehat{G}^*$ of the sampling distribution $\widehat{G}$?

By the Strong Law of Large Numbers, $\widehat{G}^* \xrightarrow{a.s.} \widehat{G}$. Moreover, by the Glivenko–Cantelli theorem we know that this convergence happens uniformly over $t$. That is,

$$\left\|\widehat{G}^* - \widehat{G}\right\|_\infty = \sup_{t \in \mathcal{T}} \left|\widehat{G}^*(t) - \widehat{G}(t)\right| \xrightarrow{a.s.} 0 \tag{1.9}$$

Equation (1.9) guarantees that, by taking a large number of resamples $B$, $\widehat{G}^*$ will be a good approximation of $\widehat{G}$. It is of interest to know which value of $B$ we should pick to get a good enough approximation of $\widehat{G}$.

Let's define what value of $B$ gives a "good enough" approximation. Let $\widehat{G}^{-1}(p)$ be the $p$-quantile of the exact bootstrap distribution $\widehat{G}$. Then it holds that

$$p = \mathbb{P}\left(\widehat{\theta}^* \leq \widehat{G}^{-1}(p)\right) = \mathbb{E}\left[\mathbb{1}\{\widehat{\theta}^* \leq \widehat{G}^{-1}(p)\}\right] \tag{1.10}$$

Thus, one could build an estimate of $p$ by computing

$$\hat{p} = \frac{1}{B}\sum_{j=1}^{B} \mathbb{1}\{\hat{\theta}^{*(j)} \leq \widehat{G}^{-1}(p)\} \tag{1.11}$$

Note that $p$ and $\hat{p}$ are just $\widehat{G}$ and $\widehat{G}^*$ evaluated at $t = \widehat{G}^{-1}(p)$. We consider that $B$ leads to a "good enough" approximation if there is a 95% chance that the Monte Carlo estimate $\hat{p}$ of $p$ is within 10% of the value of $p$, assuming $p = 0.025$. That is, $B$ such that $\hat{p} \in [0.0225, 0.0275]$ with 95% probability. To compute it, we need to know the distribution of $\hat{p}$.

We can see each $\hat{\theta}^{*(j)}$ in Equation (1.11) as one random realization of $\widehat{\theta}^* \sim \widehat{G}$. Then we have that $\mathbb{1}\{\hat{\theta}^{*(j)} \leq \widehat{G}^{-1}(p)\} \sim \text{Ber}(p)$. Hence

$$\hat{p} \sim \frac{1}{B}\,\text{Bin}(B, p) \approx \mathcal{N}\left(p, \frac{p(1-p)}{B}\right) \tag{1.12}$$

We can now compute $B$ by solving from a 95%-confidence $z$ interval's half width

$$z_{0.975}\sqrt{\frac{p(1-p)}{B}} \leq 0.1p \tag{1.13}$$

Plugging in the assumed value of $p = 0.025$ results in $B \geq 14982$.

On the other hand, if $B$ is computed for a $t$ interval with bootstrap standard error (see 3.1.1) with the same "good enough" accuracy, then $B \geq 4371$ (Hesterberg, 2015).

Summing up, and as a rule of thumb, for a $t$ interval with bootstrap standard error use $B \geq 5000$, and for a bootstrap interval that relies on Monte Carlo simulation use $B \geq 15000$. However, if one takes the commonly used $B = 10000$, then the probability that $\hat{p} \in [0.0225, 0.0275]$ drops to 89%.

There is another detail to take into account when choosing the number of bootstrap samples $B$. Some part of the literature uses numbers such as $B = 999$ or $B = 9999$, whereas some other part uses $B = 1000$ or $B = 10000$, without subtracting one. This choice of $B$ relates to the selected method to estimate the quantiles of the bootstrap distribution. The main goal of choosing $B$ one way or another is to make the estimated quantile coincide with a resample data point, avoiding an interpolation between two resample data points to get the estimated quantile. In the literature it is more common to see the "R-6" (type 6 in R) method of computing quantiles, needing $N-1$ data points to get a data point to be a $p$-quantile $q(p)$, where $p = h/N$ with $h \in \{1, 2, \ldots, N\}$. On the other hand, it is not uncommon to see some people using the "R-7" method of computing quantiles, in this case needing $N$ data points instead of $N-1$.

## 1.4 $z$ and $t$ interval for skewed populations

The reader might be wondering why this section is present in a bootstrap introduction. The main purpose is to highlight some of the shortcomings of classical confidence intervals, especially in the presence of skewness, and how bootstrap intervals can overcome them by not making assumptions about the original populations (except for $z$ and $t$ intervals with bootstrap standard error).

Although so far the estimator $\widehat{\theta}$ for $\theta$ has been treated as arbitrary, arguably the most widely used is the sample average $\bar{X}$ to estimate the true mean $\mu$ of a population. It is widely known that the sample average converges to a normal distribution by the CLT, whatever the distribution of the population $X$. The $t$ statistic is known to follow a Student $t$ distribution when $X$ follows a normal distribution. However, when $X$ is positively skewed, $\bar{X}$ and $s$ are positively correlated, the correlation does not get smaller with large $n$ and the $t$ statistic does not follow a $t$ distribution. This section highlights why relying on the CLT for the estimation of the mean, either with $\bar{X}$ or $t$, is sometimes a bad idea.

Let $X_1, \ldots, X_n$ be iid with $\mathbb{E}[X_i] = \mu$, $\text{Var}(X_i) = \sigma^2$ and $\text{Skew}(X_i) = \mathbb{E}[(\frac{X-\mu}{\sigma})^3] = \mu_3/\sigma^3 = \gamma$. Let $\widehat{\theta} = \bar{X} = n^{-1}\sum_{i=1}^n X_i$ be the sample average. Then it is easy to show that the skewness of $\bar{X}$ is $\text{Skew}(\bar{X}) = \gamma/\sqrt{n}$. Now let $t = (\bar{X} - \mu)/(s/\sqrt{n})$ be the $t$ statistic following a $t_{n-1}$ distribution. It can be shown (see Hesterberg (2015) for more details) that the skewness of $t$ is equal to

$$\text{Skew}(t) = \frac{-2\gamma}{\sqrt{n}} + \mathcal{O}_{\mathbb{P}}(n^{-3/2}) \tag{1.14}$$

This means that the skewness of $t$ is twice the skewness of $\bar{X}$, but in the *opposite direction*. So, if the original population is skewed, the $t$ statistic not only does not follow a $t$ distribution, but is also skewed in the wrong direction. Although asymptotically the standard t interval converges to the correct size and coverage, it does so very slowly, at a rate of $Ou(n^{-1/2})$ with a large constant. One might argue that, for large $n$, this error would diminish and the $t$ interval would be accurate enough. But how large should $n$ be for $t$ to be accurate enough? We can inspect that by using a first-order Edgeworth expansion to approximate the CDF of the statistic $t$

$$\mathbb{P}(t \leq x) = \Phi(x) + \kappa(2x^2 + 1)\phi(x) + \mathcal{O}_{\mathbb{P}}(n^{-1}) \tag{1.15}$$

where $\Phi$ and $\phi$ are the standard normal CDF and PDF respectively, and $\kappa = \gamma/(6\sqrt{n})$.

We can use equation (1.15) to estimate the one-sided non-coverage probabilities of a $t$ interval. Plugging in the $\alpha/2$-quantile of the $t$ distribution we get

$$\mathbb{P}(t \leq t_{\alpha/2,n-1}) = \Phi(t_{\alpha/2,n-1}) + \kappa(2t^2_{\alpha/2,n-1} + 1)\phi(t_{\alpha/2,n-1}) + \mathcal{O}_{\mathbb{P}}(n^{-1}) \tag{1.16}$$

For large $n$, we can approximate the $\alpha$-quantiles of the $t$ distribution with those of the standard normal distribution

$$\mathbb{P}(t \leq z_{\alpha/2}) = \alpha/2 + \kappa(2z^2_{\alpha/2} + 1)\phi(z_{\alpha/2}) + \mathcal{O}_{\mathbb{P}}(n^{-1}) \tag{1.17}$$

The error is the difference between the probability and $\alpha/2$. We define "accurate enough" as stated above as follows: the first-order error term is bounded to be 10% of the desired $\alpha/2$ value (for $\alpha = 0.05$ the actual probabilities would be between 0.0225 and 0.0275). We then compute

$$\frac{\gamma}{6\sqrt{n}}(2z^2_{\alpha/2} + 1)\phi(z_{\alpha/2}) \leq 0.1\frac{\alpha}{2} \tag{1.18}$$

leading to

$$n \geq \left(\frac{\gamma}{6}\frac{20}{\alpha}(2z_{\alpha/2}^2+1)\phi(z_{\alpha/2})\right)^2 \tag{1.19}$$

For exponential distributions ($\gamma = 2$), it gives $n \geq 4579$. Obviously the CLT eventually kicks in, but for a sometimes unrealistic large $n$.

The previous calculations show that the usual rule of thumb of "$n \geq 30$" for relying on the CLT is not even close to the necessary $n$ when the original population $X$ is skewed. In these cases, the CLT takes effect at a much slower rate, leading to large approximation errors if the previous rule of thumb is used.

### 1.4.1 Skewness-adjusted $t$ interval

There exist a skewness-adjusted version of the $t$ statistic introduced by Johnson (1978), which takes the form

$$t_1 = t + \kappa(2t^2+1) \tag{1.20}$$

for use in hypothesis tests, with rejection if $|t_1| \geq t_{\alpha/2,n-1}$. Kleijnen et al. (1986) obtains a confidence interval by solving $t_1$ for $\mu$, yielding a quadratic equation. Hesterberg proposes a simpler $(1-\alpha)$-confidence interval for $\mu$

$$I_{tSkew} = \bar{X} + \left(\kappa\ (2t_{\alpha/2}^2+1) \pm t_{\alpha/2}\right)\frac{s}{\sqrt{n}} \tag{1.21}$$

A simple estimate for $\gamma$ is $(1/n)\sum((X_i-\bar{X}))^3/s^3$, which is biased towards zero. This makes $I_{tSkew}$ a bit closer to a $t$ interval for small sample sizes.

For comparison, Hesterberg also shows that the endpoints of a bootstrap percentile interval (presented in the next chapter) are

$$\bar{X} + \left(\kappa\ (z_{\alpha/2}^2-1) \pm z_{\alpha/2}\right)\frac{s}{\sqrt{n}} \tag{1.22}$$

For large $n$, with $t_{\alpha/2,n-1} \approx z_{\alpha/2} \approx 2$, this has about a third of the asymmetry of the skewness-adjusted $t$ interval.

# Chapter 2

# Bootstrap Consistency, Accuracy and Failure Cases

One of the main goals of learning about the distribution of an estimator $\widehat{\theta}$ is to construct confidence intervals for the parameter $\theta$ that it is trying to estimate. For that to be possible, we need to assume that the bootstrap is consistent for the estimator $\widehat{\theta}$. Section 2.1 explains in detail what consistency in the bootstrap setting means and some cases for which bootstrap consistency holds. Section 2.2 introduces the notion of accuracy and how bootstrap accuracy relates to the CLT and Edgeworth expansions. Section 2.3 exposes a number of situations in which the bootstrap estimator fails to consistently approximate the distribution of $\widehat{\theta}$, illustrates some cases with examples and presents a technique to solve some of them. Section 2.4 collects some interesting remarks scattered around in the literature and with no clear categorisation.

Usually $\widehat{\theta}$ refers to an estimator of $\theta$. In this chapter, however, $T_n$ might refer directly to some estimator $\widehat{\theta}$ of $\theta$ or to some function of it, usually of the form $T_n = \sqrt{n}(\widehat{\theta} - \theta)$, so that $T_n$ can admit a CLT approximation.

## 2.1 Bootstrap Consistency

Let $X_1, \ldots, X_n \overset{iid}{\sim} F$, $X_i \in \mathcal{X}$, with $\mathcal{X}$ some sample space. Let $\mathcal{F}$ be the space of CDFs on $\mathcal{X}$ and defined as a convex class of distributions containing the population distribution $F$ and all degenerate distributions. Let $T(X_1, \ldots, X_n)$ be a given functional defined on $\mathcal{F}$. The CDF of $T$ is defined as

$$H_n(x) = \mathbb{P}_F(T(X_1, \ldots, X_n) \leq x) \tag{2.1}$$

and the (Monte Carlo) bootstrap distribution (introduced in Chapter 1) of $T$ is defined as

$$H_{Boot}(x) = \mathbb{P}^*(T(X_1^*, \ldots, X_n^*) \leq x) \tag{2.2}$$

where $(X_1^*, \ldots, X_n^*)$ is a bootstrap sample. $\mathbb{P}_F$ denotes probabilities under distribution $F$ and $\mathbb{P}^*$ denotes probabilities under the bootstrap distribution.

**Definition 2.1.** Let $F$ and $G$ be two CDFs on a sample space $\mathcal{X}$, and let $\rho(F, G)$ be a metric on $\mathcal{F}$. We say that the bootstrap is (*weakly*) *consistent* under $\rho$ for $T$ if

$$\rho(H_n, H_{Boot}) \xrightarrow{\mathbb{P}} 0 \tag{2.3}$$

Similarly, we say that the bootstrap is *strongly consistent* under $\rho$ for $T$ if

$$\rho(H_n, H_{Boot}) \xrightarrow{a.s.} 0 \tag{2.4}$$

Consistency is the concept we use when we say that the bootstrap "works". Note that its definition requires the use of a metric, which ultimately tries to measure the "distance" or difference between the distribution of $H_n$ and the distribution of $H_{Boot}$. There are different types of metrics, but the most commonly used is the Kolmogorov metric, the distance generated by the supremum norm:

$$K(F,\ G) = \|F - G\|_\infty = \sup_{-\infty < x < \infty} |F(x) - G(x)| \tag{2.5}$$

There is another popular metric called Mallows-Wasserstein metric, defined on $\mathcal{F}_r = \{G \in \mathcal{F} : \int \|x\|^r dG(x) < \infty\}$, which takes the form

$$\ell_r(F,\ G) = \inf_{\Gamma_{r,F,G}} (\mathbb{E}\,|Y - X|^r)^{\frac{1}{r}} \tag{2.6}$$

where $X \sim F$, $Y \sim G$, and $\Gamma_{r,F,G}$ is the class of all joint distributions of $(X, Y)$ with marginals $F$ and $G$, each with a finite $r$th moment. The special case with $r = 2$ is the most used in the bootstrap literature.

Although the Kolmogorov metric is the easiest to understand and seems the most natural, $\ell_2$ has a very interesting property that makes it very appealing (Shao and Tu, 1995):

$$\ell_2(F_n, F) \xrightarrow{\mathbb{P}} 0 \iff F_n \xrightarrow{D} F \text{ and } \mathbb{E}_{F_n}[X^i] \xrightarrow{\mathbb{P}} \mathbb{E}_F[X^i] \text{ for } i = 1, 2 \tag{2.7}$$

The same property holds for almost sure convergence. Normally one uses the bootstrap to estimate the CDF, mean and/or variance of an estimator $T_n$, and that is exactly why $\ell_2$ can be very useful. If we prove that $\ell_2(H_n, H_{Boot}) \xrightarrow{\mathbb{P}} 0$, then not only can we estimate (asymptotically) accurately the CDF of $H_n$ through $H_{Boot}$, but also its mean and variance, since the first two moments of $H_{Boot}$ converge to those of $H_n$.

On the other hand, if one uses Kolmogorov metric instead, one can only guarantee that if $K(H_n, H_{Boot}) \xrightarrow{\mathbb{P}} 0$, then $H_{Boot} \xrightarrow{D} H_n$. In this case, convergence in distribution of a random sequence does not imply convergence in moments, unless the random sequence is also uniformly integrable. For more details, see Shao and Tu (1995).

There are a number of techniques that can be used to prove bootstrap consistency for an estimator. We will not present any technique here, since we are more interested in results and theorems proven with these techniques. However, if the reader is interested, Shao and Tu mention direct application of Mallows-Wasserstein metric, Berry-Esséen's inequality, imitation (remarked as the most common) and linearisation as some techniques to prove bootstrap consistency.

### 2.1.1 Bootstrap consistency for the sample mean

The following theorem states that the bootstrap is always strongly consistent for the sample mean, as long as the variance of the population distribution is finite.

**Theorem 2.1.** *Let $X_1, \ldots, X_n \overset{iid}{\sim} F$, $X_i \in \mathbb{R}^p$, with $\mathbb{E}(X_1) = \mu$ and $\Sigma = \mathrm{Cov}_F(X_1)$ finite. Let $T(X_1, \ldots, X_n) = \sqrt{n}(\bar{X}_n - \mu)$. Then the bootstrap estimator $H_{Boot}$ is strongly consistent under the Kolmogorov metric and the $\ell_2$ metric, i.e., $K(H_n, H_{Boot}) \xrightarrow{a.s.} 0$ and $\ell_2(H_n, H_{Boot}) \xrightarrow{a.s.} 0$, as $n \to \infty$.*

We omit a proof for Theorem 2.1. Three different proofs (using direct application of Mallows-Wasserstein metric, Berry-Esséen's inequality and imitation) are presented in Shao and Tu (1995).

Notice that the fact that $\mathrm{Cov}_F(X_1)$ is finite guarantees that $T$ admits a CLT. In fact, DasGupta (2008) suggests that this is a very good rule of thumb: if a functional $T(X_1, \ldots, X_n)$ admits a CLT, then the bootstrap would be at least weakly consistent for $T$.

**Rule of Thumb.** Let $X_1, \ldots, X_n \overset{iid}{\sim} F$ and let $T(X_1, \ldots, X_n)$ be a given functional. If $T$ admits a CLT approximation, that is,

$$T(X_1, \ldots, X_n) \xrightarrow{D} \mathcal{N}(0, \tau^2)$$

where $\tau^2 < \infty$ is the asymptotic variance of $T$, then (at least) $K(H_n, H_{Boot}) \xrightarrow{\mathbb{P}} 0$.

### 2.1.2 Delta method for the bootstrap

Theorem 2.1 states that the bootstrap is, under very mild conditions, strongly consistent for the sample mean. We also know that $T_n = \sqrt{n}(\bar{X}_n - \mu)$ admits a CLT, that is, its distribution is asymptotically normal based on the CLT. We can apply the delta method

to $T$ using a suitable transformation $g(\cdot)$, which implies that **not only is the bootstrap consistent for the sample mean, but also for some functions of the sample mean**. Let us briefly remember what the delta method states.

**Theorem 2.2** (Delta method)**.** *Let $T_n$ be a sequence of random variables in $\mathbb{R}^p$ used to estimate $\theta$, such that $\sqrt{n}(T_n - \theta) \xrightarrow{D} \mathcal{N}_p(0, \Sigma(\theta))$, with covariance matrix $\Sigma(\theta) \neq 0$ finite. Let $g : \mathbb{R}^p \to \mathbb{R}^m$ be once differentiable at $\theta$, with gradient matrix $\nabla g(\theta)$ finite and $\nabla g(\theta) \neq 0$. Then*

$$\sqrt{n}(g(T_n) - g(\theta)) \xrightarrow{D} \mathcal{N}_p(0, \nabla g(\theta)^T \Sigma(\theta) \nabla g(\theta)) \tag{2.8}$$

*provided $\nabla g(\theta)^T \Sigma(\theta) \nabla g(\theta)$ is positive definite.*

Let us now "merge" theorems 2.1 and 2.2 to obtain the following theorem.

**Theorem 2.3** (Delta method for the bootstrap)**.** *Let $X_1, \ldots, X_n \overset{iid}{\sim} F$, $X_i \in \mathbb{R}^p$, and let $\mu = \mathbb{E}_F(X_1)$ and $\Sigma = \mathrm{Cov}_F(X_1)$ be finite. Let $T_n = \sqrt{n}(g(\bar{X}_n) - g(\mu))$ and, for some $m \geq 1$, let $g : \mathbb{R}^p \to \mathbb{R}^m$. Suppose $\nabla g(\cdot)$ exists in a neighbourhood of $\mu$, with $\nabla g(\mu) \neq 0$, and suppose $\nabla g(\cdot)$ is continuous at $\mu$. Then the bootstrap estimator $H_{Boot}$ is strongly consistent under the Kolmogorov metric for $H_n$, i.e., $K(H_n, H_{Boot}) \xrightarrow{a.s.} 0$.*

See Shao and Tu (1995) for a proof of Theorem 2.3.

One might wonder what would happen if $X_1$ does not have a finite second moment. Babu (1984), Athreya et al. (1987), and Knight (1989) gave some examples showing that the bootstrap estimator $H_{Boot}$ is inconsistent when $\mathbb{E}_F \|X_1\|^2 = \infty$. Giné and Zinn (1989) and Hall (1990) further proved that, for $p = 1$ and $g(x) = x$, $\mathbb{E}_F(X_1^2) < \infty$ is a sufficient and necessary condition for the strong consistency of $H_{Boot}$.

Example 2.1 shows that the bootstrap is consistent for the sample variance, and Example 2.2 shows that the bootstrap is consistent for the sample Pearson correlation coefficient.

**Example 2.1** (DasGupta (2008), Example 29.2)**.** Let $X_1, \ldots, X_n \overset{iid}{\sim} F$, $X_i \in \mathbb{R}$, with $\mu_X = \mathbb{E}_F(X_1)$, and suppose $\mathbb{E}_F(X_1^4) < \infty$. Let $Y_i = (X_i,\ X_i^2)$. Then, $Y_1, \ldots, Y_n$ are iid 2-dimensional vectors with $\mathrm{Cov}(Y_1)$ finite. Note that $\bar{Y}_n = (\bar{X}_n,\ n^{-1} \sum_{i=1}^n X_i^2)$. Consider the transformation $g : \mathbb{R}^2 \to \mathbb{R}^1$ defined as $g(u, v) = v - u^2$. Then

$$\widehat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - (\bar{X}_n)^2 = g(\bar{Y}_n)$$

If we let $\mu = \mathbb{E}_F(Y_1)$, then $g(\mu) = \sigma^2 = \mathrm{Var}_F(X_1)$. Let's check if $g(u, v)$ fulfills the

conditions of Theorem 2.3. First, we check if $\nabla g(\cdot)$ exists. The gradient of $g(u, v)$ is

$$\nabla g(u, v) = \begin{pmatrix} -2u \\ 1 \end{pmatrix} \tag{2.9}$$

so it does exist and it is linear. Since

$$\mu = \mathbb{E}_F(Y_1) = \begin{pmatrix} \mu_X \\ \mathbb{E}_F(X_1^2) \end{pmatrix} \tag{2.10}$$

we have that

$$\nabla g(\mu) = \begin{pmatrix} -2\mu_X \\ 1 \end{pmatrix} \tag{2.11}$$

so we have that $\nabla g(\cdot)$ exists in a neighbourhood of $\mu$. We additionally have that $\nabla g(\mu) \neq 0$ and that $\nabla g(\cdot)$ is continuous at $\mu$.

Since $g(\cdot)$ satisfies the conditions of Theorem 2.3, we conclude that the bootstrap is strongly consistent under the Kolmogorov metric for $\sqrt{n}(\widehat{\sigma}_n^2 - \sigma^2)$, i.e. the bootstrap estimator $H_{Boot}$ is strongly consistent for the sample variance.

**Example 2.2.** Let $(Y_i, Z_i) \overset{iid}{\sim} F_2$, $(Y_i, Z_i) \in \mathbb{R}^2$, $i = 1, \ldots, n$, with $F_2$ the CDF of a bivariate distribution with finite covariance matrix $\Sigma$. Let $\mu_Y = \mathbb{E}_F(Y_1)$ and $\mu_Z = \mathbb{E}_F(Z_1)$, both finite, and let $\sigma_Y^2 = \mathrm{Var}(Y_1)$, $\sigma_Z^2 = \mathrm{Var}(Z_1)$ and $\sigma_{YZ} = \mathrm{Cov}(Y_1, Z_1)$. Let $\bar{Y}_n = n^{-1}\sum_{i=1}^n Y_i$ and $\bar{Z}_n = n^{-1}\sum_{i=1}^n Z_i$. A common estimator for the correlation coefficient $\rho$ between $Y_1$ and $Z_1$ is the sample Pearson correlation coefficient, defined as

$$\hat{\rho}_n = \frac{\sum_{i=1}^n (Y_i - \bar{Y}_n)(Z_i - \bar{Z}_n)}{\sqrt{\sum_{i=1}^n (Y_i - \bar{Y}_n)^2 \sum_{i=1}^n (Z_i - \bar{Z}_n)^2}}$$

which can be written as $g(\bar{X}_n)$ with $X_i = (Y_i, Z_i, Y_i^2, Z_i^2, Y_i Z_i)$ and

$$g(a, b, c, d, e) = \frac{e - ab}{\sqrt{(c - a^2)(d - b^2)}}$$

If we let $\mu = \mathbb{E}_F(X_1)$, then $g(\mu) = \rho = \mathrm{Cor}(Y_1, Z_1)$. Like in the previous example, let's check if $g$ fulfills the conditions of Theorem 2.3. First, we check the existence of $\nabla g(\cdot)$. The gradient of $g(a, b, c, d, e)$ is

$$\nabla g(a,b,c,d,e) = \begin{pmatrix} \frac{-b(c-a^2)+(e-ab)}{(c-a^2)^{3/2}(d-b^2)^{1/2}} \\ \frac{-a(d-b^2)+(e-ab)}{(c-a^2)^{1/2}(d-b^2)^{3/2}} \\ \frac{-(1/2)(e-ab)}{(c-a^2)^{3/2}(d-b^2)^{1/2}} \\ \frac{-(1/2)(e-ab)}{(c-a^2)^{1/2}(d-b^2)^{3/2}} \\ \frac{1}{(c-a^2)^{1/2}(d-b^2)^{1/2}} \end{pmatrix} \tag{2.12}$$

so it exists and it is nonlinear. Since

$$\mu = \mathbb{E}_F(X_1) = \begin{pmatrix} \mu_Y \\ \mu_Z \\ \mathbb{E}_F(Y_1^2) \\ \mathbb{E}_F(Z_1^2) \\ \mathbb{E}_F(Y_1 Z_1) \end{pmatrix} \tag{2.13}$$

we have that

$$\nabla g(\mu) = \begin{pmatrix} \frac{-\mu_Z \sigma_Y^2 + \sigma_{YZ}}{\sigma_Y^3 \sigma_Z} \\ \frac{-\mu_Y \sigma_Z^2 + \sigma_{YZ}}{\sigma_Y \sigma_Z^3} \\ \frac{-(1/2)\sigma_{YZ}}{\sigma_Y^3 \sigma_Z} \\ \frac{-(1/2)\sigma_{YZ}}{\sigma_Y \sigma_Z^3} \\ \frac{1}{\sigma_Y \sigma_Z} \end{pmatrix} \tag{2.14}$$

There are discontinuities when $\sigma_{_Y} = 0$ or $\sigma_{_Z} = 0$, but both are strictly positive by definition, so $\nabla g(\cdot)$ is continuous at $\mu$ and it exists in a neighbourhood of $\mu$. Since the last element is always $> 0$, we additionally have that $\nabla g(\mu) \neq 0$. Thus, $g(\cdot)$ satisfies the conditions of Theorem 2.3, and we conclude that the bootstrap is strongly consistent under the Kolmogorov metric for $\sqrt{n}(\hat{\rho}_n^2 - \rho)$, i.e. the bootstrap estimator $H_{Boot}$ is strongly consistent for the sample Pearson correlation coefficient.

Inference about $\rho$ is often based on the transformed statistic $\hat{\phi}_n = \frac{1}{2}\log[(1+\hat{\rho}_n)/(1-\hat{\rho}_n)]$ resulting from applying Fisher's $z$-transformation to $\hat{\rho}_n$ (when $F_2$ is believed to be a bivariate normal distribution), which is also a function of $\bar{X}_n$.

### 2.1.3 Bootstrap consistency for sample quantiles

It turns out that not only is the bootstrap consistent for the mean and some functions of the mean, but it is also consistent for the sample quantiles, as the following theorem states.

**Theorem 2.4.** *Let $X_1, \ldots, X_n \overset{iid}{\sim} F$, $X_i \in \mathbb{R}$. Let $q$ be a fixed constant, $0 < q < 1$, and let $\theta = F^{-1}(q)$ be the $q$-quantile of $F$ and $\widehat{\theta} = F_n^{-1}(q)$ be the sample $q$-quantile, with $F_n$ the empirical distribution function of $X_1, \ldots, X_n$. Let $T_n = \sqrt{n}(\widehat{\theta} - \theta)$. Suppose that $F$ is continuous at $\theta$ and that $f(\theta-)$ (left-hand derivative of $F$ at $\theta$) and $f(\theta-)$ (left-hand derivative of $F$ at $\theta$) exist, are positive and $f(\theta-) = f(\theta+)$. Then the bootstrap estimator $H_{Boot}$ is strongly consistent under the Kolmogorov metric for $H_n$, i.e., $K(H_n, H_{Boot}) \xrightarrow{a.s.} 0$.*

A proof for Theorem 2.4 can be found in Example 3.2 of Shao and Tu (1995). Furthermore, Example 3.2(continued) of Shao and Tu (1995) argues that the condition $f(\theta-) = f(\theta+)$ is essential for the consistency of $H_{Boot}$ in Theorem 2.4, and provides a reasoning showing that $H_{Boot}$ is inconsistent if $f(\theta-) \neq f(\theta+)$. After that, a method to "cure" the inconsistency of $H_{Boot}$ is used, known as "m/n bootstrap", that we present in Section 2.3.1.

### 2.1.4 Hadamard-differentiable functionals and bootstrap consistency

Hadamard differentiability is a type of directional derivative for maps between Banach spaces. It is a weaker notion than usual differentiability used in different applications in topology. Theorem 2.5 makes an interesting statement: $\rho$-Hadamard differentiability of a functional $T(F_n)$ is a sufficient condition for the bootstrap consistency under $\rho$ for $T$. That is, if we can prove that the functional $T$ is Hadamard-differentiable w.r.t. some metric $\rho$, then we are guaranteed that the bootstrap is going to work for the estimator $T_n$ based on the functional $T$.

Let $F$ be a CDF on a sample space $\mathcal{X}$. Let $\mathcal{F}$ be a convex class of distributions containing $F$ and all degenerate distributions. Let $\mathcal{D}$ be the linear space generated by members of $\mathcal{F}$. We now present the definition of Hadamard differentiability as introduced in Shao and Tu (1995).

**Definition 2.2** (Hadamard differentiability)**.** A functional $T$ defined on $\mathcal{F}$ is said to be $\rho$-Hadamard differentiable at $G \in \mathcal{F}$ if there is a linear functional $L_G$ on $\mathcal{D}$ such that for any sequence of numbers $t_k \to 0$ and $\{D, D_k, k = 1, 2, \ldots\} \subset \mathcal{D}$ satisfying $\rho(D_k, D) \to 0$ and $G + t_k D_k \in \mathcal{F}$,

$$\lim_{k \to \infty} \left[ \frac{T(G + t_k D_k) - T(G)}{t_k} - L_G(D_k) \right] = 0 \tag{2.15}$$

Similarly, a functional $T$ defined on $\mathcal{F}$ is said to be continuously $\rho$-Hadamard differentiable at $G \in \mathcal{F}$ if $T$ is $\rho$-Hadamard differentiable at $G$ and if for any $G_k \in \mathcal{F}$ satisfying $\rho(G_k, G) \to 0$,

$$\lim_{k\to\infty} \left[ \frac{T(G_k + t_k D_k) - T(G_k)}{t_k} - L_G(D_k) \right] = 0 \tag{2.16}$$

Hadamard differentiability is also referred to as compact differentiability. It is stronger than Gâteaux differentiability, but weaker than Fréchet differentiability.

Before introducing Theorem 2.5 we briefly remind the reader the definition of the influence function.

**Definition 2.3.** Let $X_1, \ldots, X_n \overset{iid}{\sim} F$, $X_i \in \mathcal{X} = \mathbb{R}^p$, and let $T_n = T(X_1, \ldots, X_n)$. Let $\delta_x$ denote the $p$-dimensional CDF degenerated at the point $x \in \mathcal{X}$, i.e. the "one-point probability distribution" putting probability 1 on $x$. The influence function $\phi_F$ of $T_n$ at $F$ is

$$\phi_F(x) = \lim_{\varepsilon\to 0} \frac{T\big((1-\varepsilon)F + \varepsilon\delta_x\big) - T(F)}{\varepsilon} \tag{2.17}$$

The function $\phi_F$ is known in operator theory as the Gâteaux derivative of the operator $T(F)$ at $F$ in the direction $\delta_x - F$. It measures the differential effect of modifying $F$ by putting additional probability on $x$.

Now that we have presented the concepts of Hadamard differentiability and influence function, we can introduce Theorem 2.5, which connects bootstrap consistency to $K$-Hadamard-differentiable functionals.

**Theorem 2.5.** *Let $T_n = T(F_n)$ where $F_n$ is the empirical distribution function and $T$ is a functional defined on $\mathcal{F}$. Suppose that $T$ is $K$-Hadamard differentiable at $F$ with the influence function satisfying $0 < \mathbb{E}[\phi_F(X_1)]^2 < \infty$. Then the bootstrap estimator $H_{Boot}$ is weakly consistent. If, in addition, $T$ is continuously $K$-Hadamard differentiable at $F$, then $H_{Boot}$ is strongly consistent.*

A proof for Theorem 2.5 can be found in Gill, Wellner, and Præstgaard (1989) and Liu, Singh, and Lo (1989). There is a similar theorem for Fréchet differentiable functionals, which we are not presenting here. For more details, see Shao and Tu (1995).

Statistics $T_n = T(F_n)$ with a $K$-Hadamard differentiable $T$ include M-estimators, trimmed smooth L-statistics, linear rank statistics and R-estimators (Shao and Tu, 1995).

## 2.2 Bootstrap Accuracy

The consistency of the bootstrap estimator $H_{Boot}$ as defined in Equation (2.2) implies that $H_{Boot}$ will converge to $H_n$ asymptotically. However, it does not state how fast the rate of convergence is. It is important to study the convergence rate of $K(H_n, H_{Boot})$ to 0 for different functionals $T$, so that we can compare the convergence rate of the bootstrap estimator to other types of estimators, such as CLT approximations or Edgeworth expansions.

First, we have to define what is known as $k$th-order accuracy.

**Definition 2.4.** An estimator $\hat{H}$ for the distribution $H$ is said to be $k$th order accurate ($k \geq 1$) if its convergence rate under the Kolmogorov metric is $\mathcal{O}_{\mathbb{P}}(n^{k/2})$. That is,

$$K(H, \hat{H}) = \mathcal{O}_{\mathbb{P}}(n^{-k/2}) \tag{2.18}$$

or equivalently,

$$n^{(k-1)/2} K(H, \hat{H}) \to 0 \tag{2.19}$$

There are many different techniques to study the convergence rate of an estimator. However, in the bootstrap literature, there are two that stand out of the rest: the Berry-Esséen inequalities and Edgeworth expansions. Although the convergence rate of bootstrap estimators has been studied by many researchers, because it is very difficult to establish the Berry-Esséen inequalities and the Edgeworth expansions, there are not many results available when the statistic $T_n$ is not a function of the sample mean (Shao and Tu, 1995). We now summarize some important results, namely for the sample mean, studentised statistics and sample quantiles.

### 2.2.1 Bootstrap accuracy for the sample mean

The following theorem gives the convergence rate of the bootstrap estimator for the sample mean and the standardised sample mean, when some conditions for the moments of the population distribution $F$ are imposed.

**Theorem 2.6.** *Let* $X_1, \ldots, X_n \overset{iid}{\sim} F$, $X_i \in \mathbb{R}$, *with* $\mathbb{E}(X_1) = \mu$ *and* $\mathrm{Var}(X_1) = \sigma^2 < \infty$. *Let* $H_{Boot}(x) = \mathbb{P}^*(\sqrt{n}(\bar{X}_n^* - \bar{X}_n) \leq x)$ *be the bootstrap estimator of* $H_n(x) = \mathbb{P}(\sqrt{n}(\bar{X}_n - \mu) \leq x)$, *and let* $\tilde{H}_{Boot}(x) = H_{Boot}(\hat{\sigma}x)$ *be the bootstrap estimator of* $\tilde{H}_n(x) = H_n(\sigma x)$, *the distribution of the standardised sample mean* $\sqrt{n}(\bar{X}_n - \mu)/\sigma$, *with* $\hat{\sigma}^2$ *the sample variance.*

*(i) If* $\mathbb{E}(X_1^4) < \infty$*, then*

$$\limsup_{n\to\infty} \frac{\sqrt{n}\, K(H_n, H_{Boot})}{\sqrt{\log\log n}} = \frac{\sqrt{\mathrm{Var}(X_1-\mu)^2}}{2\sigma^2\sqrt{\pi e}} \quad a.s. \tag{2.20}$$

*(ii) If* $\mathbb{E}\,|X_1|^3 < \infty$ *and* $F$ *is lattice in the sense that there are constants* $c$ *and* $h$ *such that* $\mathbb{P}(X_1 = c + kh, k = 0, 1, 2, \ldots) = 1$*, then*

$$\limsup_{n\to\infty} \sqrt{n}\, K(\tilde{H}_n, \tilde{H}_{Boot}) = \frac{h}{\sqrt{2\pi}\sigma} \quad a.s. \tag{2.21}$$

*(iii) If* $\mathbb{E}\,|X_1|^3 < \infty$ *and* $F$ *is nonlattice, then*

$$\sqrt{n}\, K(\tilde{H}_n, \tilde{H}_{Boot}) \xrightarrow{a.s.} 0 \tag{2.22}$$

The result in Theorem 2.6 can be extended to the multivariate case and to functions of the sample mean. See Shao and Tu (1995) for a proof of Theorem 2.6 and its extension.

Theorem 2.6 shows that if $\mathbb{E}(X_1^4) < \infty$, then $H_{Boot}$ has a convergence rate of $\mathcal{O}_{\mathbb{P}}(\sqrt{\log\log n/n})$, which is the same as the convergence rate of the normal approximation $\Phi_{\hat{\sigma}} = \Phi(x/\hat{\sigma})$; if $\mathbb{E}\,|X_1|^3 < \infty$, then the convergence rate of $\tilde{H}_{Boot}$ is $\mathcal{O}_{\mathbb{P}}(n^{-1/2})$ in general, but $\mathcal{O}_{\mathbb{P}}(n^{-1})$ when $F$ is nonlattice (Shao and Tu, 1995). Hall (1988) showed that finiteness of first three absolute moments is also a necessary condition for higher-order accuracy of the bootstrap in the standardised case.

Note that for the CLT approximation $\Phi$ to $\tilde{H}_n$, $K(\tilde{H}_n, \Phi) = \mathcal{O}_{\mathbb{P}}(n^{-1/2})$ by the Berry-Esséen theorem. This rate cannot be improved, whether $F$ is lattice or not. Therefore, (iii) of Theorem 2.6 implies that the bootstrap estimator is more accurate than the CLT approximation for a standardised variable, attaining second-order accuracy. This is one of the most attractive features of the bootstrap estimator, since it not only estimates the distribution of an estimator, but can also have a higher accuracy than the traditional CLT approximation. Because any normal distribution is symmetric, the CLT cannot capture information about the skewness in the finite sample distribution of $T$. So the bootstrap succeeds in correcting for skewness, just as an Edgeworth expansion would do.

### 2.2.2 Second-order accuracy of the bootstrap

Although second-order accuracy can be attained when $T_n = \sqrt{n}(\bar{X}_n - \mu)/\sigma$, for other functionals $T$ this is not automatic. It holds for certain types of $T$ but not for others. DasGupta (2008) suggests the following rule of thumb.

**Rule of Thumb.** Let $X_1, \ldots, X_n \overset{iid}{\sim} F$, and let $T(X_1, \ldots, X_n)$ be a functional. Let $T$ admit a CLT approximation, i.e. $T(X_1, \ldots, X_n) \xrightarrow{D} \mathcal{N}(0, \tau^2)$.

(i) If $\tau$ is independent of $F$, then second-order accuracy of the bootstrap is likely. Proving it will depend on the availability of an Edgeworth expansion for $T$.

(ii) If $\tau$ depends on $F$ ($\tau = \tau(F)$), then the bootstrap should be just first-order accurate.

### 2.2.3 Bootstrap accuracy for the $t$ statistic

Analogously to Section 2.2.1, the following theorem from DasGupta (2008) gives the convergence rate of the bootstrap estimator, but in this case for the $t$ statistic.

**Theorem 2.7.** *Let $X_1, \ldots, X_n \overset{iid}{\sim} F$. Suppose $F$ is nonlattice and that $\mathbb{E}_F(X_1^6) < \infty$. Let $T_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{s}$ and $T_n^* = \frac{\sqrt{n}(\bar{X}_n^* - \bar{X}_n)}{s^*}$, with $s$ the standard deviation of $X_1, \ldots, X_n$ and $s^*$ the standard deviation of $X_1^*, \ldots, X_n^*$. Then $\sqrt{n}(H_n, H_{Boot}) \xrightarrow{a.s.} 0$, i.e. $H_{Boot}$ is second-order accurate for $H_n$.*

A more general version for studentised linear combinations of sample means can be found in Theorem 3.12 of Shao and Tu (1995). It is not a surprising result taking into account the aforementioned rule of thumb, since we know that the $t$ statistic asymptotically follows a normal distribution whose variance is independent from $F$.

### 2.2.4 Bootstrap accuracy for sample quantiles

In this section we show a theorem from Shao and Tu (1995) that gives the convergence rate of the bootstrap estimator for sample quantiles. Note that its convergence rate is markedly slower than that for the mean.

**Theorem 2.8.** *Let $X_1, \ldots, X_n \overset{iid}{\sim} F$, $X_i \in \mathbb{R}$. Let $q$ be a fixed constant, $0 < q < 1$, and let $\theta = F^{-1}(q)$ be the $q$-quantile of $F$ and $\widehat{\theta} = F_n^{-1}(q)$ be the sample $q$-quantile, with $F_n$ the empirical distribution. Let $T_n = \sqrt{n}(\widehat{\theta} - \theta)$. Suppose that $F$ has a bounded second order derivative in a neighbourhood of $\theta$ and $f(\theta) \geq 0$, where $f(x) = dF/dx$. Then there is a constant $c_F$ such that*

$$\limsup_{n \to \infty} \frac{n^{1/4} K(H_n, H_{Boot})}{\sqrt{\log \log n}} = c_F \quad a.s. \tag{2.23}$$

By the Berry-Esséen inequality

$$K(H_n, \Phi_{\sigma_F}) = \mathcal{O}_{\mathbb{P}}(n^{-1/2})$$

where $\sigma_F = \sqrt{q(1-q)}/f(\theta)$ and $\Phi_{\sigma_F} = \Phi(x/\sigma_F)$. If we have an estimator $\widehat{\sigma}_F$ of $\sigma_F$ and we use $\Phi_{\widehat{\sigma}_F}$ to estimate $H_n$, then whether the bootstrap estimator $H_{Boot}$ is better than $\Phi_{\widehat{\sigma}_F}$ depends on the convergence rate of $\widehat{\sigma}_F$ to $\sigma_F$. It might also be of interest to study

the accuracy of the bootstrap estimator for the studentised statistic $\sqrt{n}(\widehat{\theta}-\theta)/\widehat{\sigma}_F$. But it involves Edgeworth expansions for sample quantiles and their bootstrap counterparts and, unfortunately, there are no results available (Shao and Tu, 1995).

Falk and Reiss (1989) studied the weak convergence of the process

$$\mathbb{P}^*(\sqrt{n}(F_n^{*-1}(q) - F_n^{-1}(q)) \leq x) - \mathbb{P}(\sqrt{n}(F_n^{-1}(q) - F^{-1}(q)) \leq x)$$

and showed that the bootstrap estimator $H_{Boot}$ has a convergence rate of $\mathcal{O}_{\mathbb{P}}(n^{-1/4})$.

## 2.3 Bootstrap Failure

So far we have only proven some cases in which the bootstrap works, i.e. in which the bootstrap is consistent for an estimator $T_n$. However, for other cases it is not guaranteed to work. DasGupta (2008) suggests that these are typically cases where the estimator $T_n$ fails to admit a CLT, and provides the following list of situations where the bootstrap fails to estimate the CDF of $T_n$ consistently:

1) $T_n = \sqrt{n}(\bar{X}_n - \mu)$ when $\mathrm{Var}(X_1) = \infty$.

2) $T_n = \sqrt{n}(g(\bar{X}) - g(\mu))$ and $g$ is not differentiable at $\mu$.

3) $T_n = \sqrt{n}(g(\bar{X}) - g(\mu))$ and $\nabla g(\mu) = 0$.

4) $T_n = \sqrt{n}(\widehat{\theta} - \theta)$ with $\theta = F^{-1}(q)$, $\widehat{\theta} = F_n^{-1}(q)$, and $f(\theta) = 0$ or $f(\theta+) \neq f(\theta-)$, where $0 < q < 1$ is a given constant, $F$ is the population CDF, $f$ is the population PDF and $F_n$ is the empirical distribution of $X_1, \ldots, X_n$.

5) The underlying population $F_\theta$ is indexed by a parameter $\theta$, and the support of $F_\theta$ depends on the value of $\theta$.

6) The underlying population $F_\theta$ is indexed by a parameter $\theta$, and the true value $\theta_0$ belongs to the boundary of the parameter space $\Theta$.

It is clear that failure case 1) is a consequence of $T_n$ not admitting a CLT due to infinite variance, which Theorem 2.1 states as a necessary condition for the consistency of the bootstrap. Failure cases 2) and 3) are a consequence of the limitations of the delta method, since the differentiability of $g$ and $\nabla g(\mu) \neq 0$ are necessary conditions, as stated in Theorem 2.3.

Failure case 4) happens when the conditions for the CLT approximation of $T_n$ (as defined in 4)) are not met. It is known that

$$\sqrt{n}({F_n}^{-1}(q) - F^{-1}(q)) \xrightarrow{D} \mathcal{N}\left(0, \frac{q(1-q)}{f^2(F^{-1}(q))}\right) \tag{2.24}$$

Thus, if $f(F^{-1}(q)) = 0$, the asymptotic variance blows up and there is no CLT approximation anymore. In the case with $f(F^{-1}(q)+) \neq f(F^{-1}(q)-)$ and a discontinuity at $f(F^{-1}(q))$, it is clear that it poses a problem when having to plug in a value for $f^2(F^{-1}(q))$ in Equation (2.24). However, in the case with $f(F^{-1}(q)+) \neq f(F^{-1}(q)-)$ but $f$ continuous at $F^{-1}(q)$, it is not clear to us if that would be a problem for the CLT approximation to work. It might have more to do with the bootstrap estimator itself.

The following examples give an idea of the type of cases in which the bootstrap estimator fails to be consistent.

**Example 2.3** (DasGupta (2008), Example 29.7)**.** This is a practical example of failure case 2). Let $X_1, \ldots, X_n \overset{iid}{\sim} F$, with $\mathbb{E}(X_1) = \mu$ and $\text{Var}(X_1) = \sigma^2$. Let $g(x) := |x|$ and $T_n := \sqrt{n}\Big(g(\bar{X}) - g(\mu)\Big)$. By the CLT, if the true value of $\mu$ is 0, then

$$\sqrt{n}(\bar{X}_n - \mu) = \sqrt{n}\bar{X}_n \xrightarrow{D} Z_\sigma$$

where $Z_\sigma \sim \mathcal{N}(0, \sigma^2)$. By the Continuous Mapping Theorem, applying $g(x)$ on both sides we have

$$g\Big(\sqrt{n}\bar{X}_n\Big) = \sqrt{n}|\bar{X}| =: T_n \xrightarrow{D} g\Big(Z_\sigma\Big) = |Z_\sigma|$$

We need to consider a couple of facts before showing that the bootstrap does not work in this case:

- For almost all sequences $X_1, \ldots, X_n$ the conditional distribution of $\sqrt{n}(\bar{X}_n^* - \bar{X}_n)$, given $\bar{X}_n$, converges in distribution to $\mathcal{N}(0, \sigma^2)$ by the triangular array CLT (Vaart, 1998).
- The joint asymptotic distribution of $\Big(\sqrt{n}(\bar{X}_n - \mu), \sqrt{n}(\bar{X}_n^* - \bar{X}_n)\Big) \xrightarrow{D} (Z_1, Z_2)$, where $Z_1, Z_2$ are iid $\mathcal{N}(0, \sigma^2)$.

Now, if the true $\mu$ is equal to 0:

$$\begin{aligned} T_n^* &:= \sqrt{n}(|\bar{X}_n^*| - |\bar{X}_n|) \\ &= \sqrt{n}|\bar{X}_n^* - \bar{X}_n + \bar{X}_n| - \sqrt{n}|\bar{X}_n| \\ &= |\sqrt{n}(\bar{X}_n^* - \bar{X}_n) + \sqrt{n}\bar{X}_n| - |\sqrt{n}\bar{X}_n| \\ &\xrightarrow{D} |Z_2 + Z_1| - |Z_1| \end{aligned}$$

where $Z_1, Z_2$ are iid $\mathcal{N}(0, \sigma^2)$. But this is not distributed as the absolute value of $\mathcal{N}(0, \sigma^2)$. Therefore, the bootstrap estimator $H_{Boot}$ is not consistent for $H_n$ when $\mu = 0$.

**Example 2.4** (Shao and Tu (1995), Example 3.6)**.** *Functions of the sample mean with null derivatives.* This is an example of failure case 3). Let $X_1, \ldots, X_n \overset{iid}{\sim} F$, $X_i \in \mathbb{R}^p$, with $\mu = \mathbb{E}(X_1)$ and $\Sigma = \mathrm{Cov}(X_1)$ finite. Let $g : \mathbb{R}^p \to \mathbb{R}$, $\theta = g(\mu)$ and $T_n = g(\bar{X}_n)$. Suppose that $g$ is continuously second-order differentiable in a neighbourhood of $\mu$, with $\nabla g(\mu) = 0$ and $\nabla^2 g(\mu) \neq 0$. Using Taylor's expansion and $\nabla g(\mu) = 0$, we obtain that

$$T_n - \theta = \frac{1}{2}(\bar{X}_n - \mu)^T \nabla^2 g(\mu)(\bar{X}_n - \mu) + o_{\mathbb{P}}(n^{-1})$$

which results in

$$n(T_n - \theta) = \frac{n}{2}(\bar{X}_n - \mu)^T \nabla^2 g(\mu)(\bar{X}_n - \mu) \xrightarrow{D} \frac{1}{2} Z_\Sigma^T \nabla^2 g(\mu) Z_\Sigma \tag{2.25}$$

where $Z_\Sigma$ is a $p$-dimensional normal distribution with mean 0 and covariance matrix $\Sigma$. Thus, we should study the bootstrap estimator of the distribution of $n(T_n - \theta)$, not $\sqrt{n}(T_n - \theta)$, in the case of $\nabla g(\mu) = 0$. Let $X_1^*, \ldots, X_n^* \overset{iid}{\sim} F_n$, with $F_n$ the empirical distribution of $X_1, \ldots, X_n$, and let $\bar{X}_n^*$ and $T_n^*$ be the bootstrap analogs of $\bar{X}_n$ and $T_n$, respectively. Then,

$$T_n^* - T_n = \nabla g(\bar{X}_n)^T(\bar{X}_n^* - \bar{X}_n) + \frac{1}{2}(\bar{X}_n^* - \bar{X}_n)^T \nabla^2 g(\bar{X}_n)(\bar{X}_n^* - \bar{X}_n) + o_{\mathbb{P}}(n^{-1}) \quad a.s.$$

and equivalently,

$$n(T_n^* - T_n) = n\nabla g(\bar{X}_n)^T(\bar{X}_n^* - \bar{X}_n) + \frac{n}{2}(\bar{X}_n^* - \bar{X}_n)^T \nabla^2 g(\bar{X}_n)(\bar{X}_n^* - \bar{X}_n) + o_{\mathbb{P}}(1) \quad a.s. \tag{2.26}$$

By Theorem 2.1, for almost all given sequences $X_1, X_2, \ldots$,

$$\frac{n}{2}(\bar{X}_n^* - \bar{X}_n)^T \nabla^2 g(\bar{X}_n)(\bar{X}_n^* - \bar{X}_n) \xrightarrow{D} \frac{1}{2} Z_\Sigma^T \nabla^2 g(\mu) Z_\Sigma$$

We would like $n\nabla g(\bar{X}_n)^T(\bar{X}_n^* - \bar{X}_n)$ to converge to zero, since then we would be able to know the distribution of $n(T_n^* - T_n)$. Using a first-order Taylor's expansion around $\bar{X}_n$ for the function $\nabla g(\cdot)$ we get

$$\nabla g(\bar{X}_n) = \nabla g(\mu) + \nabla^2 g(\mu)(\bar{X}_n - \mu) + o_{\mathbb{P}}(n^{-1/2})$$

Since $\nabla g(\mu) = 0$,

$$\sqrt{n}\nabla g(\bar{X}_n) = \sqrt{n}\nabla^2 g(\mu)(\bar{X}_n - \mu) + o_{\mathbb{P}}(1) \xrightarrow{D} \nabla^2 g(\mu) Z_\Sigma \tag{2.27}$$

Hence, for almost all given sequences $X_1, X_2, \ldots$, the conditional distribution of the term $\sqrt{n}\nabla g(\bar{X}_n)^T \sqrt{n}(\bar{X}_n^* - \bar{X}_n) = n\nabla g(\bar{X}_n)^T(\bar{X}_n^* - \bar{X}_n)$ does not have a limit. It follows

from Equation (2.26) that, for almost all given sequences $X_1, X_2, \ldots$, the conditional distribution of $n(T_n^* - T_n)$ does not have a limit. Therefore, the bootstrap estimator $H_{Boot}$ is inconsistent for the distribution of $\sqrt{n}(T_n - \theta)$.

Example 2.4 indicates that the bootstrap estimator $H_{Boot}$ may not have a limit while $H_n$ has a limit. This is caused by an inherent problem of the bootstrap: the bootstrap data are drawn from the empirical distribution $F_n$, which is not exactly the population distribution $F$. The effect of this problem is negligible in regular cases (when $T_n$ can be well approximated by a linear statistic), but leads to inconsistency of the bootstrap in other cases.

The symptom of the inconsistency problem in Example 2.4 is that $\nabla g(\bar{X}_n)$ is not necessarily equal to 0 when $\nabla g(\mu) = 0$. As a result, the expansion in Equation (2.26), compared to the expansion in Equation (2.25), has an extra term $n\nabla g(\bar{X}_n)^T(\bar{X}_n^* - \bar{X}_n)$ that does not converge to 0 fast enough. Therefore, the conditional distribution of $n(T_n^* - T_n)$ cannot mimic the distribution of $n(T_n - \theta)$.

**Example 2.5** (DasGupta (2008), Example 29.8)**.** This is a practical example of failure case 5). Let $X_1, \ldots, X_n \overset{iid}{\sim} U(0, \theta)$, and let $\widehat{\theta} = X_{(n)}$, $\widehat{\theta}^* = X_{(n)}^*$, $T_n = n(\theta - \widehat{\theta})$, $T_n^* = n(\hat{\theta} - \widehat{\theta}^*)$, where $X_{(n)}$ is the maximum of the sample $X_1, \ldots, X_n$ and $X_{(n)}^*$ is the maximum of a bootstrap sample $X_1^*, \ldots, X_n^*$. The ordinary nonparametric bootstrap will fail in this example, since $T_n^* \overset{\mathbb{P}}{\not\to} T_n$.

First, let us find the distribution of $T_n$. We know that the CDF of $\widehat{\theta}$ is

$$F_{\widehat{\theta}}(x) = \mathbb{P}(\widehat{\theta} \leq x) = [F_X(x)]^n = \left(\frac{x}{\theta}\right)^n$$

The CDF of $T_n$ can easily be found as follows:

$$\begin{aligned}
\mathbb{P}(T_n \leq x) &= \mathbb{P}\left(n(\theta - \widehat{\theta}) \leq x\right) \\
&= \mathbb{P}\left(\theta - \widehat{\theta} \leq \frac{x}{n}\right) \\
&= \mathbb{P}\left(\widehat{\theta} \geq \theta - \frac{x}{n}\right) \\
&= 1 - \mathbb{P}\left(\widehat{\theta} \leq \theta - \frac{x}{n}\right) \\
&= 1 - \left(\frac{\theta - x/n}{\theta}\right)^n \\
&= 1 - \left(1 - \frac{x/\theta}{n}\right)^n \\
&\xrightarrow{n\to\infty} 1 - e^{-x/\theta}
\end{aligned}$$

for $x \leq 0$, since $T_n \leq 0$. Thus, $T_n \sim \text{Exp}(\theta)$. Now let us find the distribution of $T_n^*$. Noting that $T_n^*$ is always $\geq 0$, for $x \geq 0$ we have

$$\begin{aligned}
\mathbb{P}(T_n^* \leq x) &\geq \mathbb{P}(T_n^* = 0) \\
&= \mathbb{P}(\widehat{\theta}^* = \hat{\theta}) \\
&= \mathbb{P}(\widehat{\theta}^* \geq \hat{\theta}) \\
&= 1 - \mathbb{P}(\widehat{\theta}^* < \hat{\theta}) \\
&= 1 - \left(\frac{n-1}{n}\right)^n \\
&\xrightarrow{n\to\infty} 1 - e^{-1}
\end{aligned}$$

For $x < \theta$, we have that $\mathbb{P}(T_n^* \leq x) > \mathbb{P}(T_n \leq x)$ asymptotically, since $1 - e^{-\frac{x}{\theta}} < 1 - e^{-1}$, and consequently the bootstrap estimator $H_{Boot} = \mathbb{P}(T_n^* \leq x)$ always overestimates $H_n = \mathbb{P}(T_n \leq x)$. Thus, $T_n^* \overset{\mathbb{P}}{\not\to} T_n$ and the bootstrap estimator $H_{Boot}$ is inconsistent for $H_n$.

**Example 2.6** (Andrews (2000))**.** This is a practical example of failure case 6). Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, 1), \quad \mu \geq 0$. The ML estimator for $\mu$ is $\hat{\mu}_n = max(0, \bar{X}_n)$. Then $\sqrt{n}(\hat{\mu}_n - \mu)$ is asymptotically distributed as:

$$\begin{cases} Z & \text{if } \mu > 0 \\ max(0, Z) & \text{if } \mu = 0 \end{cases}$$

where $Z \sim \mathcal{N}(0, 1)$. If $\mu > 0$, the maximum plays no role asymptotically, but if $\mu = 0$, the maximum will always play a role asymptotically. Then, if the true mean is $\mu = 0$, the asymptotic distribution is non-Gaussian and the bootstrap is inconsistent for $\theta$. Andrews also provides some solutions for this specific example, but states that there are no universal approaches and it has to be done on a case-by-case fashion.

### 2.3.1 m/n bootstrap to rectify bootstrap inconsistency

This technique consists in resampling $m$ out of $n$ observations from the empirical distribution function $F_n$, where $m < n$, and then use $T_m^* = T(X_1^*, \ldots, X_m^*)$ in place of $T_n = T(X_1^*, \ldots, X_n^*)$, where $X_1^*, \ldots, X_m^* \overset{iid}{\sim} F_n$. Bickel, Freedman, et al. (1981) studied this type of bootstrap estimator with $m$ as a function of $n$ or with $m$ varying freely. It can provide consistent bootstrap estimators in many cases where the inconsistency is caused by a lack of smoothness of $T$ (Shao and Tu, 1995).

If the $n$ out of $n$ ordinary bootstrap is already consistent, then there can still be $m$ out of $n$ schemes that are also consistent, with $m/n \to 0$ as $n \to \infty$. However, the $m$ out of $n$ scheme will perform somewhat worse than the ordinary bootstrap (DasGupta, 2008).

We now present some theorems from DasGupta (2008) that show that the $m/n$ bootstrap can solve the bootstrap inconsistency problem in a number of cases. Theorem 2.9 addresses the case when $\nabla g(\mu) = 0$ in Theorem 2.3. Theorem 2.10 addresses the case when $F$ has a discontinuity at $\theta$, i.e. $f(\theta-) \neq f(\theta+)$ in Theorem 2.4. Finally, Theorem 2.11 solves the inconsistency problem for extreme order statistics, like the case presented in Example 2.5. Proofs and more details can be found in Shao and Tu (1995).

**Definition 2.5.** Let $T$ be defined as for Equation (2.1). Let $X_1^*, \ldots, X_m^* \overset{iid}{\sim} F_n$, with $F_n$ the empirical distribution function of $X_1, \ldots, X_n$, defined as for Equation (2.1) as well. We define the $m/n$ bootstrap estimator as

$$H_{Boot,m,n}(x) = \mathbb{P}(T(X_1^*, \ldots, X_m^*) \leq x) \tag{2.28}$$

**Theorem 2.9.** *Let $X_1, \ldots, X_n \overset{iid}{\sim} F$, $X_i \in \mathbb{R}^p$, $p \geq 1$. Suppose $\mu = \mathbb{E}_F(X_1)$ and $\Sigma = \mathrm{Cov}_F(X_1)$ exist and are finite, and suppose $\Sigma$ is positive definite. Let $g : \mathbb{R}^p \to \mathbb{R}$ be such that $\nabla g(\mu) = 0$ and the Hessian matrix $\nabla^2 g(\mu) \neq 0$. Let $T_n = n(g(\bar{X}_n) - g(\mu))$ and $T_{m,n}^* = m(g(\bar{X}_m^*) - g(\bar{X}_n))$ and define $H_n(x) = \mathbb{P}_F(T_n \leq x)$ and $H_{Boot,m,n}(x) = \mathbb{P}^*(T_{m,n}^* \leq x)$. Here $\bar{X}_m^*$ denotes the mean of an iid sample of size $m = m(n)$ from $F_n$, where $m \to \infty$ as $n \to \infty$ and $m(n)$ means that $m$ is a function of $n$.*

*(a) If $m = o(n)$, then $K(H_{Boot,m,n}, H_n) \xrightarrow{\mathbb{P}} 0$.*

*(b) If $m = o(\frac{n}{\log\log n})$, then $K(H_{Boot,m,n}, H_n) \xrightarrow{a.s.} 0$.*

**Theorem 2.10.** *Let $X_1, \ldots, X_n \overset{iid}{\sim} F$, $X_i \in \mathbb{R}$. Let $q$ be a fixed constant, $0 < q < 1$, and let $\theta = F^{-1}(q)$ be the $q$-quantile of $F$ and $\widehat{\theta} = F_n^{-1}(q)$ be the sample $q$-quantile, with $F_n$ the empirical distribution function of $X_1, \ldots, X_n$. Suppose that $F$ is continuous at $\theta$ and that $f(\theta-)$ (left-hand derivative of $F$ at $\theta$) and $f(\theta-)$ (left-hand derivative of $F$ at $\theta$) exist, are positive and $f(\theta-) \neq f(\theta+)$. Let $T_n = \sqrt{n}(\widehat{\theta} - \theta)$, and $T_{m,n}^* = \sqrt{m}(\widehat{\theta}^* - \widehat{\theta})$, where $\widehat{\theta}^* = F_m^{*-1}$ denotes the $q$-quantile of a bootstrap sample of size $m$ from $F_n$, and define $H_n(x) = \mathbb{P}_F(T_n \leq x)$ and $H_{Boot,m,n}(x) = \mathbb{P}^*(T_{m,n}^* \leq x)$ . Then,*

*(a) If $m = o(n)$, then $K(H_{Boot,m,n}, H_n) \xrightarrow{\mathbb{P}} 0$.*

*(b) If $m = o(\frac{n}{\log\log n})$, then $K(H_{Boot,m,n}, H_n) \xrightarrow{a.s.} 0$.*

Results for the case where $f(\theta)$ exists but $f(\theta) = 0$ (which would lead to inconsistency of the bootstrap even if $f(\theta-) = f(\theta+)$) are given by Huang, Sen, and Shao (1996).

**Theorem 2.11.** *Let* $X_1, \ldots, X_n \overset{iid}{\sim} F$, $X_i \in \mathbb{R}$. *Suppose* $\theta = \theta(F)$ *is such that* $F(\theta) = 1$ *and* $F(x) < 1$ *for all* $x < \theta$. *Suppose, for some* $\delta > 0$, $\mathbb{P}_F(n^{1/\delta}(\theta - X_{(n)}) > x) \to e^{-(x/\theta)^\delta}$, *for all* $x$. *Let* $T_n = n^{1/\delta}(\theta - X_{(n)})$ *and* $T^*_{m,n} = m^{1/\delta}(X_{(n)} - X^*_{(m)})$, *and define* $H_n(x) = \mathbb{P}_F(T_n \le x)$ *and* $H_{Boot,m,n}(x) = \mathbb{P}^*(T^*_{m,n} \le x)$.

*(a) If* $m = o(n)$, *then* $K(H_{Boot,m,n}, H_n) \xrightarrow{\mathbb{P}} 0$.

*(b) If* $m = o(\frac{n}{\log\log n})$, *then* $K(H_{Boot,m,n}, H_n) \xrightarrow{a.s.} 0$.

The key question then is to choose the size $m$ of the new bootstrap samples. DasGupta (2008) states that this is a difficult question to answer, and that no precise prescriptions that have any sort of general optimality are possible. Then suggests a rule of thumb, to take $m \approx 2\sqrt{n}$.

**Example 2.2** (continued)**.** Let $\bar{X}^*_m$ and $T^*_m$ be the bootstrap analogs of $\bar{X}_n$ and $T_n$, respectively, based on a bootstrap sample of size $m$. Then,

$$T^*_m - T_n = \nabla g(\bar{X}_n)^T(\bar{X}^*_m - \bar{X}_n) + \frac{1}{2}(\bar{X}^*_m - \bar{X}_n)^T\nabla^2 g(\bar{X}_n)(\bar{X}^*_m - \bar{X}_n) + o_{\mathbb{P}}(m^{-1}) \quad a.s. \tag{2.29}$$

By Theorem 2.1 in Bickel et al. (1981), for almost all given sequences $X_1, X_2, \ldots$,

$$\frac{m}{2}(\bar{X}^*_m - \bar{X}_n)^T\nabla^2 g(\bar{X}_n)(\bar{X}^*_m - \bar{X}_n) \xrightarrow{D} \frac{1}{2}Z_\Sigma^T\nabla^2 g(\mu)Z_\Sigma \tag{2.30}$$

as $m \to \infty$ and $n \to \infty$. The expansion in Equation (2.29) still has the nonzero term $\nabla g(\bar{X}_n)^T(\bar{X}^*_m - \bar{X}_n)$, but it is of order $o_{\mathbb{P}}(m^{-1})$ almost surely if $m \log\log n/n \to 0$ since, by Equation (2.27), $\nabla g(\bar{X}_n) = \mathcal{O}(\sqrt{\log\log n/n})$. This proves that

$$T^*_m - T_n = \frac{1}{2}(\bar{X}^*_m - \bar{X}_n)^T\nabla^2 g(\bar{X}_n)(\bar{X}^*_m - \bar{X}_n) + o_{\mathbb{P}}(m^{-1}) \quad a.s. \tag{2.31}$$

as if we had taken $n = \infty$ or $\nabla g(\bar{X}_n) = 0$. By Equation (2.30) and Equation (2.31), the bootstrap estimator $H_{Boot,m,n}$ is strongly consistent if $m = m(n) \to \infty$ and $m \log\log n/n \to 0$. Similarly, it can be shown that $H_{Boot,m,n}$ is weakly consistent if $m = m(n) \to \infty$ and $m/n \to 0$, since $\nabla g(\bar{X}_n) = \mathcal{O}_{\mathbb{P}}(n^{-1/2})$.

**Example 2.3** (continued)**.** Let $X^*_1, \ldots, X^*_m \overset{iid}{\sim} F_n$, with $F_n$ the empirical distribution function of $X_1, \ldots, X_n$. Let $\widehat{\theta}^*_m = X^*_{(m)}$ be the maximum of $X^*_1, \ldots, X^*_m$, and let $T^*_m = n(\hat{\theta} - \widehat{\theta}^*_m)$. If $m/n \to 0$, then

$$
\begin{aligned}
\mathbb{P}(T_m^* \leq x) &\geq \mathbb{P}(T_m^* = 0) \\
&= \mathbb{P}(\widehat{\theta}_m^* = \hat{\theta}) \\
&= \mathbb{P}(\widehat{\theta}_m^* \geq \hat{\theta}) \\
&= 1 - \mathbb{P}(\widehat{\theta}_m^* < \hat{\theta}) \\
&= 1 - \left(\frac{n-1}{n}\right)^m \\
&= 1 - \left(1 - \frac{1}{n}\right)^{n\frac{m}{n}} \\
&\xrightarrow{n\to\infty} 1 - e^{-m/n} \\
&\xrightarrow{n\to\infty} 0
\end{aligned}
$$

which coincides with the result $\mathbb{P}(T_n \leq 0)$. Thus, the inconsistency of the bootstrap is rectified (Shao and Tu, 1995).

Example 3.8 from Shao and Tu (1995) shows how the $m/n$ bootstrap estimator can fix the bootstrap inconsistency of Example 2.3.

## 2.4 Interesting Remarks

In this section we try to compile a list of interesting remarks and statements from Shao and Tu (1995) about the bootstrap that could not fit in the previous sections, but are interesting enough to be included and mentioned in this work.

1) Whether the bootstrap is better than Edgeworth expansions in terms of estimating a tail probability depends on how far down the tail we want to go. Except for very extreme tail probabilities, the bootstrap estimator is better or equivalent to the Edgeworth expansion estimator.

2) The bootstrap estimator for the variance of an estimator $T_n$ tends to underestimate the variance of $T_n$, and it is not as efficient as the jackknife estimator when both estimators require relatively the same amount of computation. Hence, the bootstrap is not recommended if only the variance of $T_n$ needs to be computed.

3) The smoothed bootstrap is useful when the sample size $n$ is small. For large $n$, the smoothed bootstrap estimator improves the non-smoothed bootstrap estimator when the latter has a slow convergence rate (e.g., a rate much slower than $n^{-1/2}$), which occurs in the case of sample quantiles. However, smoothing usually increases the amount of computational work. More discussions can be found in De Angelis and Young (1992).

# Chapter 3

# Bootstrap Confidence Intervals

In this chapter we assume that bootstrap consistency holds, so we can construct bootstrap confidence intervals for $\theta$ (based on $\widehat{\theta}$) that are asymptotically consistent (see Vaart (1998) for a proof). The term "asymptotically" is used to stress the fact that bootstrap confidence intervals are only asymptotically correct. Sections 3.1 and 3.2 present a number of well-known bootstrap confidence intervals, according to their accuracy. Section 3.3 presents a method to achieve higher order accuracy for a bootstrap confidence interval of choice. An acronym for the name of every interval is displayed in parenthesis alongside the title of the section. Lastly, Section 3.4 introduces two broadly used R packages to implement the bootstrap and bootstrap confidence intervals.

In Chapter 2 an estimator has been generally noted as $T_n = T(X_1, \ldots, X_n)$, with $T$ a given functional. In this chapter, we stick to the most commonly used notation in the bootstrap literature, that is, $\theta$ to refer to the parameter of interest, $\widehat{\theta}$ for an estimator of $\theta$ and $\widehat{\theta}^*$ for the bootstrap estimator of $\widehat{\theta}$.

As mentioned in Chapter 1, in the bootstrap literature $\widehat{\theta}$ usually represents the value of the estimator $\widehat{\theta}$ evaluated at the sample data, i.e. $\widehat{\theta}(x_1, \ldots, x_n)$, as well as the estimator $\widehat{\theta}$ as a random variable, i.e. $\widehat{\theta}(X_1, \ldots, X_n)$, where $x_1, \ldots, x_n$ are the specific values drawn from $F$ and $X_1, \ldots, X_n$ are the respective random variables. Here we use $\hat{\theta}$ to refer to the former and $\widehat{\theta}$ to refer to the latter, trying to make a distinction by using different sizes of the "hat" symbol. If not specified, it should be clear from the context which one is being referred to.

Additionally, we use the same notation as in Chapter 1 for the distribution of the estimator $\widehat{\theta}$ ($G$), the exact bootstrap distribution of the bootstrap estimator $\widehat{\theta}^*$ ($\widehat{G}$) and the Monte Carlo bootstrap distribution of the bootstrap estimator $\widehat{\theta}^*$ ($\widehat{G}^*$). The latter is commonly referred to simply as bootstrap distribution, since the former is very rarely computed in practice.

## 3.1 First-Order Accurate Intervals

**Definition 3.1** (Efron and Tibshirani (1993))**.** A confidence interval is *first-order accurate* if, for its endpoints $(q_{\widehat{\theta}}(\alpha_1),\ q_{\widehat{\theta}}(\alpha_2))$, the actual one-sided rejection probabilities or one-sided non-coverage probabilities differ from the nominal values by $\mathcal{O}_{\mathbb{P}}(n^{-1/2})$. That is,

$$\mathbb{P}\left(\theta \leq q_{\widehat{\theta}}(\alpha)\right) = \alpha + \mathcal{O}_{\mathbb{P}}\left(n^{-1/2}\right) \tag{3.1}$$

First-order intervals are usually simple and fast to compute. On the flip side, they are normally less accurate than higher-order intervals. Next, we present a number of different first-order bootstrap confidence intervals.

### 3.1.1 *z* and *t* interval with bootstrap standard error (zB, tB)

These intervals combine classical statistical theory with bootstrapping. They are based on the $z$ or $t$ interval for estimating $\theta$, but the standard error of the estimator $\widehat{\theta}$ is estimated by using the standard error of $\widehat{\theta}^*$. That is, $\mathrm{se}(\widehat{\theta}) \approx \widehat{\mathrm{se}}^*(\widehat{\theta}^*)$. Thus, $z$ and $t$ $(1-\alpha)$-confidence intervals are

$$I_{zB} = \left[\hat{\theta} \pm \Phi^{-1}\left(1-\frac{\alpha}{2}\right) \cdot \widehat{\mathrm{se}}^*(\widehat{\theta}^*)\right] \tag{3.2}$$

$$I_{tB} = \left[\hat{\theta} \pm q_{t_{n-1}}\left(1-\frac{\alpha}{2}\right) \cdot \widehat{\mathrm{se}}^*(\widehat{\theta}^*)\right] \tag{3.3}$$

These intervals are always symmetric around $\hat{\theta}$. This is one of the reasons why these intervals are not transformation invariant. That is, if we want to construct a confidence interval for the transformed value $\phi = m(\theta)$, we have that

$$\left[\hat{\phi}^*_{\alpha/2},\ \hat{\phi}^*_{1-\alpha/2}\right] \neq \left[m\left(q_{\widehat{\theta}^*}\left(\alpha/2\right)\right),\ m\left(q_{\widehat{\theta}^*}(1-\alpha/2)\right)\right] \tag{3.4}$$

so we would have to bootstrap the new estimator $\widehat{\phi} = m(\widehat{\theta})$ and get bootstrap values $\hat{\phi}^*$ to construct a $z$ or $t$ interval with bootstrap SE for $\phi$.

There exists another version of the $z$ and $t$ interval with bootstrap standard error that corrects for the bias of $\widehat{\theta}$. This version uses the bootstrap bias as an estimation of the true bias of $\widehat{\theta}$, and then subtracts it from the original $z$ or $t$ interval. They take the form

$$I_{zB-BC} = \left[\hat{\theta} - \mathrm{bias}^*(\widehat{\theta}) \pm \Phi^{-1}\left(1-\frac{\alpha}{2}\right) \cdot \widehat{\mathrm{se}}^*(\widehat{\theta}^*)\right] \tag{3.5}$$

$$I_{tB-BC} = \left[\hat{\theta} - \mathrm{bias}^*(\widehat{\theta}) \pm q_{t_{n-1}}\left(1-\frac{\alpha}{2}\right) \cdot \widehat{\mathrm{se}}^*(\widehat{\theta}^*)\right] \tag{3.6}$$

where

$$\text{bias}^*(\widehat{\theta}) = \mathbb{E}^*(\widehat{\theta}^*) - \hat{\theta} \tag{3.7}$$

Hesterberg (2015) highlights that $z$ and $t$ intervals tend to be too narrow for small n, by a factor of $\sqrt{(n-1)/n}$ for the mean, what he calls *narrowness bias*. The reason goes back to the plug-in principle: the empirical CDF $F_n$ has variance $\text{Var}\,(X) = \widehat{\sigma}^2 = (1/n)\sum_{i=1}^{n}(X_i - \bar{X})^2$, not $s^2$. Then suggests some solutions to correct this .

### 3.1.2 Percentile interval (P)

Introduced by Efron (1979), this is the simplest type of "pure" bootstrap confidence interval. Commonly referred to simply as percentile interval, a $(1-\alpha)$-confidence Bootstrap percentile intervals are constructed by taking the $(\alpha/2)$- and $(1-\alpha/2)$-quantiles of the distribution of $\widehat{\theta}^*$, i.e. the distribution of bootstrap estimates $\hat{\theta}^{*b}$, $b = 1, \ldots, B$. That is,

$$I_P = \left[q_{\widehat{\theta}^*}\left(\frac{\alpha}{2}\right),\ q_{\widehat{\theta}^*}\left(1-\frac{\alpha}{2}\right)\right] \tag{3.8}$$

It suffers the same narrowness bias as the zB/tB interval and even more —for symmetric data it is like using $z_{\alpha/2}\,\hat{\sigma}/\sqrt{n}$ in place of $t_{\alpha/2,n-1}\,s/\sqrt{n}$. This extra bias $z_{\alpha/2}/t_{\alpha/2,n-1}$ is due to the unaccounted variability of $\text{Var}\,(X)$, since we are using $\bar{X}$ instead of $\mu$ to estimate it.

The percentile interval needs not be symmetric around $\widehat{\theta}$, unless the original population distribution $F$ is symmetric. This means that the *percentile take into account the asymmetry of the distribution of* $\widehat{\theta}$.

For certain parameters, there is a restriction on the values that the parameter can take. For instance, the values of the correlation coefficient must lie in the interval $[-1, 1]$. If a confidence procedure always produces intervals that fall within the allowable range, such intervals are called *range-preserving*. *The percentile interval is range-preserving*, since a) the plug-in estimate $\widehat{\theta}$ obeys the same restriction as $\theta$, and b) its endpoints are values of the bootstrap distribution $\widehat{G}^*$, which obey the same range restriction as $G$, the distribution of $\widehat{\theta}$ (Efron and Tibshirani, 1993). In contrast, zB or tB need not be range-preserving.

Another nice property of the percentile interval is that it is *transformation invariant* (Hesterberg, 2015). For instance, suppose we want to construct a confidence interval for $\phi = m(\theta)$, where $m$ is a monotone transformation. We do not need to generate new bootstrap replications $\hat{\phi}^* = m(\widehat{\theta}^*)$. We just need to transform the quantiles of $\widehat{\theta}^*$ to get the new interval:

$$\left[\hat{\phi}^*_{\alpha/2},\ \hat{\phi}^*_{1-\alpha/2}\right] = \left[m\left(q_{\widehat{\theta}^*}\left(\alpha/2\right)\right),\ m\left(q_{\widehat{\theta}^*}(1-\alpha/2)\right)\right] \tag{3.9}$$

The classical $z$ and $t$ intervals work well when the distribution of the estimator $\widehat{\theta}$ is normal, but not so well otherwise. An example from Chapter 13 in Efron and Tibshirani (1993) shows what happens when the distribution of $\widehat{\theta}$ is non-normal and the $z$ interval is used to construct a confidence interval. The estimated quantiles are off the true ones by quite some margin, but the bootstrap percentile quantiles are very close to the true ones. However, if a transformation $m$ is applied to $\widehat{\theta}$ to make it normally distributed, the quantiles are correctly estimated. Then the quantiles from the $z$ interval are correct and the bootstrap percentile quantiles are almost identical.

The bootstrap percentile interval for $\theta$ agrees well with a $z$ interval that is constructed on an appropriate transformation of $\theta$, and then mapped back to the $\theta$ scale. The difficulty in improving the $z$ interval (or $t$ interval) is to find an appropriate transformation for each parameter $\theta$ of interest. The following lemma states that the percentile interval achieves just that. So the advantage of the percentile interval over the $z$ interval is that it automatically finds that transformation and incorporates it to its quantiles, assuming such transformation exists.

**Lemma 3.1** (Percentile interval lemma (Efron and Tibshirani, 1993))**.** *Suppose there exists a monotonic transformation $\widehat{\phi} = m(\widehat{\theta})$, $\phi = m(\theta)$, that perfectly normalizes the distribution of the estimator $\widehat{\theta}$ for all $\theta$, that is,*

$$\widehat{\phi} \sim \mathcal{N}(\phi, c^2) \quad , \forall \theta \tag{3.10}$$

*for some constant standard deviation $c > 0$. Then the $(1-\alpha)$-confidence percentile interval for $\theta$ based on $\widehat{\theta}$ is the interval*

$$\Big[m^{-1}\Big(\hat{\phi} - c \cdot \Phi^{-1}(1-\alpha/2)\Big),\ m^{-1}\Big(\hat{\phi} - c \cdot \Phi^{-1}(\alpha/2)\Big)\Big] \tag{3.11}$$

**Remark.** Notice that the quantiles seem to be "reversed". Assuming that equation (3.10) holds, we have

$$\begin{aligned}
1-\alpha &= \mathbb{P}\left(\Phi^{-1}(\alpha/2) \le \frac{\hat{\phi}-\phi}{c} \le \Phi^{-1}(1-\alpha/2)\right) \\
&= \mathbb{P}\left(c \cdot \Phi^{-1}(\alpha/2) \le \hat{\phi}-\phi \le c \cdot \Phi^{-1}(1-\alpha/2)\right) \\
&= \mathbb{P}\left(-c \cdot \Phi^{-1}(\alpha/2) \ge \phi-\hat{\phi} \ge -c \cdot \Phi^{-1}(1-\alpha/2)\right) \\
&= \mathbb{P}\left(\hat{\phi} - c \cdot \Phi^{-1}(\alpha/2) \ge \phi \ge \hat{\phi} - c \cdot \Phi^{-1}(1-\alpha/2)\right)
\end{aligned}$$

Hence, $\phi \in \Big[\hat{\phi} - c \cdot \Phi^{-1}(1-\alpha/2),\ \hat{\phi} - c \cdot \Phi^{-1}(\alpha/2)\Big]$. Mapping back to the original estimator $\widehat{\theta}$ scale, we get the result in (3.11). Since the normal CDF $\Phi$ is symmetric, (3.11) is equivalent to $\Big[m^{-1}\Big(\hat{\phi} - c \cdot \Phi^{-1}(1-\alpha/2)\Big),\ m^{-1}\Big(\hat{\phi} + c \cdot \Phi^{-1}(1-\alpha/2)\Big)\Big]$.

**Remark.** Notice that Lemma 3.1 implies that the transformation $\theta \mapsto m(\theta)$ maps

$$\frac{\widehat{\theta} - \theta}{\text{se}(\widehat{\theta})} \sim ? \tag{3.12}$$

to

$$\frac{\widehat{\phi} - \phi}{c} \sim \mathcal{N}(0, 1) \tag{3.13}$$

so it is assuming that $\widehat{\theta}$ can be rearranged in a way such that it follows a normal distribution with zero mean and constant variance.

This lemma brings good news because it states we need not know what the correct transformation is, since the percentile interval will find it automatically. The bad news is that a transformation that normalizes the estimator $\widehat{\theta}$ might not always exist, and it is not possible to know when that is the case.

### 3.1.3 Expanded percentile interval (EP)

The percentile interval performs poorly in small samples, because of the narrowness bias and because it lacks a fudge factor to allow for variability of the standard error. The standard $t$ interval handles both, using $s$ instead of $\hat{\sigma}$ to avoid narrowness bias, and $t_{\alpha/2,\,n-1}$ instead of $z_{\alpha/2}$ to take into account the variability of $s$. It can also be interpreted as multiplying the length of a $z$ interval, $\bar{X} \pm z_{\alpha/2}\,\hat{\sigma}$, by a factor $a_{\alpha/2,\,n} = (t_{\alpha/2,\,n-1}/z_{\alpha/2})(s/\hat{\sigma})$, to provide better coverage.

Similarly, we can take the percentile interval and adjust it to provide better coverage *when the bootstrap distribution is normal*. However, if we just multiply the percentile interval quantiles by $a_{\alpha/2,\,n}$ it would not be transformation invariant. Instead, we can adjust the quantiles by finding an alternative $\alpha'$ value that gives the desired coverage $1 - \alpha$, preserving transformation invariance.

If the bootstrap distribution is approximately normal, then

$$q_{\widehat{\theta}^*}(\alpha/2) \approx \hat{\theta} + \Phi^{-1}(\alpha/2)\,\frac{\hat{\sigma}}{\sqrt{n}} \tag{3.14}$$

We want to find an adjusted value $\alpha'$ with

$$\begin{aligned} q_{\widehat{\theta}^*}(\alpha'/2) &\approx \hat{\theta} + \Phi^{-1}(\alpha'/2)\,\frac{\hat{\sigma}}{\sqrt{n}} \\ &= \hat{\theta} + t_{\alpha/2,\,n-1}\,\frac{s}{\sqrt{n}} \end{aligned} \tag{3.15}$$

This results in $\Phi^{-1}(\alpha'/2) = \sqrt{n/(n-1)}\, t_{\alpha/2,n-1}$, or equivalently

$$\alpha'/2 = \Phi\Big(\sqrt{n/(n-1)}\, t_{\alpha/2,\, n-1}\Big) \tag{3.16}$$

In summary, we have gone from

$$I_P = \left[q_{\widehat{\theta}^*}\left(\frac{\alpha}{2}\right),\ q_{\widehat{\theta}^*}\left(1-\frac{\alpha}{2}\right)\right]$$

to

$$I_{EP} = \left[q_{\widehat{\theta}^*}\left(\frac{\alpha'}{2}\right),\ q_{\widehat{\theta}^*}\left(1-\frac{\alpha'}{2}\right)\right] \tag{3.17}$$

where $I_{EP}$ gives the desired coverage $1-\alpha$.

This adjustment does not correct bias or skewness. It only counteracts the narrowness bias and provides a fudge factor for uncertain width. Although designed for the case in which the bootstrap distribution is approximately normal, it can also provide some degree of correction for non-normal bootstrap distributions. Hesterberg shows that it can also help for skewness to some extent, which improves on the performance of the percentile interval.

### 3.1.4 Reverse percentile interval (RP)

Also known as basic bootstrap interval, instead of estimating the distribution of $\widehat{\theta}$ to get confidence intervals for $\theta$, the reverse percentile interval is computed estimating the bootstrap distribution of $\widehat{\delta} := \widehat{\theta} - \theta$, assumed to be a pivot. Let us quickly remember the definition of a pivot.

**Definition 3.2.** A *pivot* $Z(X,\theta)$ is a function of the data and the unknown parameter $\theta$ such that, for all $\theta \in \Theta$, the distribution $G_Z$ of $Z(X,\theta)$ *does not* depend on the unknown parameter $\theta$. That is,

$$G_Z(z) := \mathbb{P}_\theta\Big(Z(X,\theta) \le z\Big) \tag{3.18}$$

*does not* depend on $\theta$.

Since the distribution of the pivot $\widehat{\delta} := \widehat{\theta} - \theta$ is unknown, it is estimated using the distribution of $\widehat{\delta}^* := \widehat{\theta}^* - \hat{\theta}$, for which we have that

$$
\begin{aligned}
1-\alpha &= \mathbb{P}(q_{\widehat{\theta}^*-\hat{\theta}}(\alpha/2) \leq \widehat{\theta}^* - \hat{\theta} \leq q_{\widehat{\theta}^*-\hat{\theta}}(1-\alpha/2)) \\
\textit{bootstrap consistency} \rightarrow &\approx \mathbb{P}(q_{\widehat{\theta}^*-\hat{\theta}}(\alpha/2) \leq \widehat{\theta} - \theta \leq q_{\widehat{\theta}^*-\hat{\theta}}(1-\alpha/2)) \\
&= \mathbb{P}(-q_{\widehat{\theta}^*-\hat{\theta}}(\alpha/2) \geq \theta - \widehat{\theta} \geq -q_{\widehat{\theta}^*-\hat{\theta}}(1-\alpha/2)) \\
&= \mathbb{P}(\widehat{\theta} - q_{\widehat{\theta}^*-\hat{\theta}}(\alpha/2) \geq \theta \geq \widehat{\theta} - q_{\widehat{\theta}^*-\hat{\theta}}(1-\alpha/2)) \qquad (3.19)
\end{aligned}
$$

Hence,

$$
I_{RP} = \left[\hat{\theta} - q_{\widehat{\theta}^*-\hat{\theta}}(1-\alpha/2),\ \hat{\theta} - q_{\widehat{\theta}^*-\hat{\theta}}(\alpha/2)\right] \qquad (3.20)
$$

By the equivariance of the quantile function we have

$$
q_{\widehat{\theta}^*-\hat{\theta}}(\alpha) = q_{\widehat{\theta}^*}(\alpha) - \hat{\theta} \qquad (3.21)
$$

Thus, a $(1-\alpha)$-confidence reverse percentile interval is constructed as

$$
I_{RP} = \left[2\hat{\theta} - q_{\widehat{\theta}^*}\left(1-\frac{\alpha}{2}\right),\ 2\hat{\theta} - q_{\widehat{\theta}^*}\left(\frac{\alpha}{2}\right)\right] \qquad (3.22)
$$

Equation (3.23) illustrates why this method is called *reverse* percentile interval:

$$
I_{RP} = \left[2\hat{\theta} - \underbrace{\left\{q_{\widehat{\theta}^*}(1-\alpha/2),\ q_{\widehat{\theta}^*}(\alpha/2)\right\}}_{\text{reverse percentile quantiles}}\right] \qquad (3.23)
$$

This interval is very popular and widely used. It is based on a mathematically sound derivation that seems to provide an interval with the desired coverage based on bootstrap consistency.

However, it has the same small-sample problems as the percentile interval. Moreover, it is asymmetrical in the wrong direction for skewed data, and also asymmetrical in the wrong direction for nonlinear transformations (Hesterberg, 2015).

Although using a pivot to construct confidence intervals is a good idea, $\widehat{\delta}$ is a wrong choice for a pivot because it is not even close to pivotal (Hesterberg, 2015). That is, $\widehat{\delta}$ is not really a pivot since its distribution depends on unknown parameters, namely $\theta$. In many cases, different values of $\theta$ will lead to different distributions of $\widehat{\theta}$, with different mean, variance, asymmetry, etc.

However, the reverse percentile interval lays the basis for the bootstrap $t$ interval, which is built on the same idea of using pivots and is a much better interval in many aspects.

### 3.1.5 Bias-corrected percentile interval (BCP)

Introduced by Efron (1981), and also known as BC interval, this is an improved version of the percentile interval, since it tries to take into account the bias of the estimator $\widehat{\theta}$. This interval assumes that there exists some monotonic transformation $\widehat{\phi} = m(\widehat{\theta})$, $\phi = m(\theta)$ that can achieve normality and constant standard error of the transformed estimator

$$\frac{\widehat{\phi} - \phi}{c} \sim \mathcal{N}(-z_0,\, 1) \tag{3.24}$$

where $c$ is the constant standard error and $z_0$ is a bias constant.

The bias $z_0$ arises from the following reasoning. If we assume $\widehat{\phi} \sim \mathcal{N}(\phi,\, c^2)$ for all $\phi = m(\theta)$ as hypothesized in (3.10), then $\widehat{\phi}^* \sim \mathcal{N}(\hat{\phi},\, c^2)$ (Section 11.3 of Efron and Hastie (2016)), and

$$\mathbb{P}^*(\widehat{\phi}^* \leq \hat{\phi}) = 0.50 \iff \mathbb{P}^*(\widehat{\theta}^* \leq \hat{\theta}) = 0.50 \tag{3.25}$$

That is, $\widehat{\theta}^*$ is median unbiased for $\hat{\theta}$, and likewise $\widehat{\theta}$ for $\theta$.

But sometimes $\widehat{\theta}$ is not median unbiased for $\theta$, and the distribution $G$ of $\widehat{\theta}$ is biased upwards (or downwards) relative to $\theta$. This, in turn, means that the distribution $\widehat{G}^*$ of $\widehat{\theta}^*$ is also biased upwards (or downwards) relative to the distribution $G$ of $\widehat{\theta}$. We can find out by checking if

$$\mathbb{P}^*(\widehat{\theta}^* \leq \hat{\theta}) = \int_{-\infty}^{\hat{\theta}} d\widehat{G}^* = 0.50 \tag{3.26}$$

Thus, if we have bootstrap estimates $\hat{\theta}^{*1}, \ldots, \hat{\theta}^{*B}$, we can estimate the proportion of estimates below $\hat{\theta}$ as

$$p_0 = \frac{\sum_{j=1}^{B} \mathbb{1}\{\hat{\theta}^{*j} \leq \hat{\theta}\}}{B} \tag{3.27}$$

Note that $\widehat{\theta}^*$ is biased upwards relative to $\hat{\theta}$ if $p_0 < 0.50$ and biased downwards if $p_0 > 0.50$. Finally, one can estimate the bias $z_0$ of (3.10) and correct it to yield (3.24) by computing

$$\hat{z}_0 = \Phi^{-1}\left(\frac{\sum_{j=1}^{B} \mathbb{1}\{\hat{\theta}^{*j} \leq \hat{\theta}\}}{B}\right) \tag{3.28}$$

The BC interval is constructed as

$$I_{BCP} = \left[ q_{\widehat{\theta}^*} \left( \Phi \left[ 2\hat{z}_0 + \Phi^{-1} \left( \frac{\alpha}{2} \right) \right] \right),\ q_{\widehat{\theta}^*} \left( \Phi \left[ 2\hat{z}_0 + \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right) \right] \right) \right] \tag{3.29}$$

Note that the interval corrects twice the bias. The reason is that, if $\widehat{\theta}$ has a bias for estimating $\theta$, so will have $\widehat{\theta}^*$ for estimating $\hat{\theta}$. So we need to correct for the estimator bias and the bootstrap bias.

From (3.24) we can see what the BC interval is equivalent to:

$$\begin{aligned}
1 - \alpha &= \mathbb{P} \left( \Phi^{-1}(\alpha/2) \leq z_0 + \frac{\widehat{\phi} - \phi}{c} \leq \Phi^{-1}(1 - \alpha/2) \right) \\
&= \mathbb{P} \left( -z_0 + \Phi^{-1}(\alpha/2) \leq \frac{\widehat{\phi} - \phi}{c} \leq -z_0 + \Phi^{-1}(1 - \alpha/2) \right) \\
&= \mathbb{P} \left( -cz_0 + c \cdot \Phi^{-1}(\alpha/2) \leq \widehat{\phi} - \phi \leq -cz_0 + c \cdot \Phi^{-1}(1 - \alpha/2) \right) \\
&= \mathbb{P} \left( cz_0 - c \cdot \Phi^{-1}(\alpha/2) \geq \phi - \widehat{\phi} \geq cz_0 - c \cdot \Phi^{-1}(1 - \alpha/2) \right) \\
&= \mathbb{P} \left( \widehat{\phi} + cz_0 - c \cdot \Phi^{-1}(\alpha/2) \geq \phi \geq \widehat{\phi} + cz_0 - c \cdot \Phi^{-1}(1 - \alpha/2) \right) \\
&= \mathbb{P} \left( (\widehat{\phi} + cz_0) + c \cdot \Phi^{-1}(1 - \alpha/2) \geq \phi \geq (\widehat{\phi} + cz_0) - c \cdot \Phi^{-1}(1 - \alpha/2) \right)
\end{aligned}$$

so we have that (3.29) is equivalent to

$$m^{-1} \left[ (\hat{\phi} + c\hat{z}_0) \pm c \cdot \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right) \right] \tag{3.30}$$

Allowing for the bias $z_0$ considerably improves the approximation made with the standard percentile interval (Efron, 1987). It also does not entail extra computational burden, in the sense that the computation of $\hat{z}_0$ is $\mathcal{O}(B)$ and it does not grow with $n$.

To finish off this section, we present Table 3.1 with a summary of some properties of first-order accurate bootstrap confidence intervals, and whether they are implemented in the R package *boot* used for the simulations in next chapter.

## 3.2 Second-Order Accurate Intervals

**Definition 3.3** (Efron and Tibshirani (1993))**.** A confidence interval is *second-order accurate* if, for its endpoints $(q_{\widehat{\theta}}(\alpha_1),\ q_{\widehat{\theta}}(\alpha_2))$, the actual one-sided rejection probabilities or one-sided non-coverage probabilities differ from the nominal values by $\mathcal{O}_{\mathbb{P}}(n^{-1})$. That is,

Table 3.1: Properties of first-order accurate bootstrap confidence intervals.

| Properties of 1st-order intervals | zB, tB | P | EP | RP | BCP |
|---|---|---|---|---|---|
| Suffers from narrowness bias | yes | yes | no | yes | yes |
| Corrects for bias | yes[1] | no | no | no | yes |
| Transformation invariant | no | yes | yes | no | yes |
| Range preserving | no | yes | yes | ?[2] | yes |
| Implemented in R package *boot* | yes | yes | no | yes | no |

[1] zB-BC and tB-BC.
[2] Unknown to the author of this thesis.

$$\mathbb{P}\left(\theta \leq q_{\widehat{\theta}}(\alpha)\right) = \alpha + \mathcal{O}_{\mathbb{P}}\left(n^{-1}\right) \tag{3.31}$$

Contrary to first-order intervals, second-order intervals are usually more complex to build and take longer to compute. However, their main advantage is that they are usually more accurate.

### 3.2.1 Bootstrap $t$ interval (BT)

Classical statistical theory says that if the population distribution $F$ is normal, then $\bar{X}$ and $s$ are independent, and the statistic $t = (\bar{X} - \mu)/(s/\sqrt{n})$ follows a Student's $t$ distribution with $n - 1$ degrees of freedom.

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}} \quad \sim \quad t_{n-1} \tag{3.32}$$

However, if $F$ is not normal, and say positively skewed, then $\bar{X}$ and $s$ are positively correlated, the correlation does not get smaller with large $n$ and the statistic $t$ *does not* follow a $t_{n-1}$ distribution, but some unknown distribution $D$ (Hesterberg, 2015).

Also known as studentised bootstrap (or less commonly percentile-$t$), the bootstrap $t$ tries to estimate that unknown distribution $D$ by means of the bootstrap distribution $D^*$. In general, for an unknown population distribution $F$ and for an estimator $\widehat{\theta}$ of $\theta$, the $t$ statistic is defined as

$$t = \frac{\widehat{\theta} - \theta}{\widehat{\text{se}}(\widehat{\theta})} \quad \sim \quad D \ = ? \tag{3.33}$$

The bootstrap $t$ estimates the distribution $D$ of $t$ by using

$$t^* = \frac{\widehat{\theta}^* - \hat{\theta}}{\widehat{\text{se}}^*(\widehat{\theta}^*)} \quad \sim \quad D^* \tag{3.34}$$

Then, assuming that bootstrap consistency holds, the distribution of $t^*$ converges to that of $t$. We can then perform a similar calculation as for the reverse percentile interval:

$$\begin{aligned}
1-\alpha &= \mathbb{P}(q_{t^*}(\alpha/2) \leq t^* \leq q_{t^*}(1-\alpha/2)) \\
\textit{boot. consist.} \rightarrow &\approx \mathbb{P}(q_{t^*}(\alpha/2) \leq t \leq q_{t^*}(1-\alpha/2)) \\
&= \mathbb{P}(q_{t^*}(\alpha/2)\,\widehat{\text{se}}(\widehat{\theta}) \leq \widehat{\theta} - \theta \leq q_{t^*}(1-\alpha/2)\,\widehat{\text{se}}(\widehat{\theta})) \\
&= \mathbb{P}(-q_{t^*}(\alpha/2)\,\widehat{\text{se}}(\widehat{\theta}) \geq \theta - \widehat{\theta} \geq -q_{t^*}(1-\alpha/2)\,\widehat{\text{se}}(\widehat{\theta})) \\
&= \mathbb{P}(\widehat{\theta} - q_{t^*}(\alpha/2)\,\widehat{\text{se}}(\widehat{\theta}) \geq \theta \geq \widehat{\theta} - q_{t^*}(1-\alpha/2)\,\widehat{\text{se}}(\widehat{\theta}))
\end{aligned} \tag{3.35}$$

Hence, a $(1-\alpha)$-confidence bootstrap $t$ interval is of the form

$$I_{BT} = \left[\hat{\theta} - q_{t^*}(1-\alpha/2)\,\widehat{\text{se}}(\widehat{\theta}),\ \hat{\theta} - q_{t^*}\left(\frac{\alpha}{2}\right)\,\widehat{\text{se}}(\widehat{\theta})\right] \tag{3.36}$$

In practice, the bootstrap $t$ interval picks up more asymmetry from the estimator $\widehat{\theta}$ than the percentile interval does, especially when the variance of $\widehat{\theta}$ depends on $\theta$ (Hesterberg, 2015). The bootstrap $t$ does this by using the studentised statistic $t^*$. Dividing by $\widehat{\text{se}}(\widehat{\theta}^*)$ aims to correct for the changing variance of $\widehat{\theta}$, since each $\hat{\theta}^*$ is brought to the same scale. Besides, the bootstrap $t$ interval also has no bias and does not suffer from narrowness bias (Hesterberg, 2015).

Because the $t^*$ statistic is studentised, the standard error of $\widehat{\theta}^*$ is needed for every bootstrap estimate $\hat{\theta}^{*b}$, $b = 1, \ldots, B$. For some cases the standard error can be derived theoretically, such as for the sample average, which is known to be $\sigma^2/n$. But usually it is unknown, and then an *iterated bootstrap* procedure is needed. A second bootstrap layer needs to be implemented, which generates second-layer bootstrap resamples $X^{**m}$, $m = 1, \ldots, M$, from each first-layer bootstrap sample $X^{*b}$, yielding second-layer bootstrap estimates $\hat{\theta}^{**m}$, $m = 1, \ldots, M$. Then we estimate the standard error of each $\hat{\theta}^{*b}$ by computing the standard deviation of its the second-layer bootstrap estimates $\hat{\theta}^{**m}$.

However, although being in general more precise than first-order intervals, it has some drawbacks worth noting. First, it may perform erratically for small $n$, especially when the standard error of $\widehat{\theta}^*$ cannot be estimated accurately (Efron and Tibshirani, 1993). Second, it is not range-preserving, as the values of the quantiles come from the distribution of $t^*$, which may not preserve the range of allowed values for $\widehat{\theta}$. And third, it is not transformation-invariant, and although it generally works well for location parameters, the practical difficulty lies in identifying the transformation $h(\cdot)$ that maps the problem to a location form (DiCiccio and Efron, 1996). Interestingly, Tibshirani (1988) proposes an "automatic" variance-stabilization procedure for the bootstrap $t$, which tries to find the transformation $h(\cdot)$ using the bootstrap itself.

### 3.2.2 BCa interval (BCa)

Introduced by Efron (1987), the BCa interval stands for *bias-corrected and accelerated* interval. The motivation for the development of the BCa interval stems from the usually erratic behaviour of the bootstrap $t$ interval in practice and the unsatisfactory coverage properties of the percentile interval. It is an improved and more general version of the percentile interval, with less erratic behaviour compared to the bootstrap $t$ and improved coverage properties, also achieving second-order accuracy (Efron and Tibshirani, 1993).

The underpinning model in which BCa intervals are based stems from that of BC intervals, as explained in Section 3.1.5. It makes a one further generalisation, assuming that there exists a monotonic transformation $\widehat{\phi} = m(\widehat{\theta})$, $\phi = m(\theta)$ that achieves normality but not necessarily with constant variance:

$$\frac{\widehat{\phi} - \phi}{c} \sim \mathcal{N}(-z_0\,\sigma_\phi,\, \sigma_\phi^2) \quad , \quad \sigma_\phi = 1 + a\phi \tag{3.37}$$

Notice that (3.24) is a special case of (3.37) when $a = 0$. The difference between (3.24) and (3.37) is greater than it seems. The hypothesized ideal transformation $m$ leading to (3.24) must be both *normalising* and *variance-stabilising*. Efron (1982) shows that normalisation and variance stabilisation are partially antagonistic goals in familiar families of distributions such as the Poisson and the binomial. Hence, it is no surprise that intervals based on (3.37 are usually more accurate than (3.24).

The BCa interval is constructed in a similar fashion as the percentile interval, but with modified left and right non-coverage probabilities $\alpha_1$ and $\alpha_2$. A $(1-\alpha)$-confidence interval is constructed as

$$I_{BCa} = \Big[q_{\widehat{\theta}^*}(\alpha_1),\; q_{\widehat{\theta}^*}(\alpha_2)\Big] \tag{3.38}$$

where

$$\begin{aligned}
\alpha_1 &= \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + \Phi^{-1}(\alpha/2)}{1 - \hat{a}\Big(\hat{z}_0 + \Phi^{-1}(\alpha/2)\Big)}\right) \\
\alpha_2 &= \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + \Phi^{-1}(1-\alpha/2)}{1 - \hat{a}\Big(\hat{z}_0 + \Phi^{-1}(1-\alpha/2)\Big)}\right)
\end{aligned} \tag{3.39}$$

with

$$\hat{z}_0 = \Phi^{-1}\left(\frac{\sum_{j=1}^{B} \mathbb{1}\{\hat{\theta}^{*j} \leq \hat{\theta}\}}{B}\right) \tag{3.40}$$

and, in the nonparametric bootstrap setting:

$$\hat{a} = \frac{1}{6} \frac{\sum_{i=1}^{n} U_i^3}{(\sum_{i=1}^{n} U_i^2)^{3/2}} \tag{3.41}$$

where $\alpha$ is the desired type I error, $\hat{\theta}^{*j}$ is the $j$th (first-layer) bootstrap estimate, $B$ is the number of bootstrap samples and $U_i$ is the empirical influence function of the estimator $\widehat{\theta} = T(F_n)$ evaluated at $X_i$

$$U_i = \lim_{\varepsilon \to 0} \frac{T((1-\varepsilon)F_n + \varepsilon \delta_{X_i}) - T(F_n)}{\varepsilon} \tag{3.42}$$

with $F_n$ the empirical distribution function of $X_1, \ldots, X_n$ and $\delta_x$ the "one-point probability distribution" putting probability 1 on $x$.

Quite a few formulas, right? Next we try to explain some of them. For more details, especially on the derivation of $\alpha_1$ and $\alpha_2$, we suggest the eager reader to read Efron (1987).

First, the estimate $\hat{z_0}$ is just an estimate of the bias of (3.37), just as (3.28) is an estimate of the bias of (3.24). Second, and as its name implies, $\hat{a}$ is an estimate of the parameter $a$ in (3.37). Efron (1987) shows that, in the parametric bootstrap setting, $\hat{a} = \frac{1}{6} \text{Skew}\left(\dot{\ell}_{\theta=\hat{\theta}}(\hat{\theta})\right)$ is an excellent estimate of $a$, where $\dot{\ell}_\theta(\hat{\theta}) = \frac{\partial}{\partial \theta} \log f_\theta(\hat{\theta})$, with $f_\theta$ the PDF of the estimator $\widehat{\theta}$. However, when using the nonparametric bootstrap we do not know $f_\theta$ since we do not assume any parametric distribution for the population $F$. Hence, we need to find a nonparametric way of estimating $\hat{a}$.

The skewness of a random variable $X \sim F$, with $\mu = \mathbb{E}(X)$ and $\sigma^2 = \text{Var}(X)$ is defined as

$$\text{Skew}(X) = \mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \frac{\mathbb{E}\left[(X-\mu)^3\right]}{\sigma^3} = \frac{\mathbb{E}\left[(X-\mu)^3\right]}{\mathbb{E}\left[(X-\mu)^2\right]^{3/2}} \tag{3.43}$$

This is where the empirical influence function kicks in. Efron (1987) uses the empirical influence function of $\widehat{\theta}$ to estimate $\text{Skew}\left(\dot{\ell}_{\theta=\hat{\theta}}(\hat{\theta})\right)$, resulting in the estimate $\hat{a}$ for $a$ in (3.41). DiCiccio and Efron (1996) suggest using the jackknife to calculate the $U_i$, calling it the jackknife influence function

$$U_i = (n-1)(\hat{\theta}_{(\cdot)} - \hat{\theta}_{(i)}) \tag{3.44}$$

where $\hat{\theta}_{(i)}$ is the estimate of $\theta$ based on the reduced sample $X_{(i)} = \{X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n\}$, and $\hat{\theta}_{(\cdot)} = \frac{1}{n} \sum_{i=1}^{n} \hat{\theta}_{(i)}$. There exist other methods to estimate $U_i$ that we do not explain here. The *boot* package in R, which we introduce in Section 3.4, has a function called *empinf* that computes the $U_i$ with different methods depending on the specific problem.

These methods include the jackknife, the infinitesimal jackknife, the positive jackknife and a regression-based estimation.

BCa intervals are the most general version of a type of models for confidence intervals. Notice that, if we let $a = 0$ in (3.37), we get the model (3.24) for BC intervals. Again, if we let $z_0 = 0$, we arrive at the model (3.13) for percentile intervals. Finally, if we let $m(\theta) = \theta$ in (3.10), we have that (3.12) follows a normal distribution, and we can construct $z$ or $t$ intervals for $\theta$ using a bootstrap estimate of the standard error of $\widehat{\theta}$. Hence, these four types of bootstrap intervals differ from each other by increasing levels of complexity, and thus different levels of generality in their application.

BCa intervals retain the advantages of percentile intervals, such as transformation invariance and the range-preserving property, but they also achieve second-order accuracy. However, they are not monotone in coverage level (DiCiccio and Efron, 1996).

### 3.2.3 ABC interval (ABC)

ABC intervals, standing for Approximate Bootstrap Confidence intervals, is a method of approximating the BCa interval endpoints analytically, without using any Monte Carlo bootstrap replications. The idea is to approximate BCa intervals without the high computational burden of sampling thousands of bootstrap replicates.

The ABC method works by approximating the bootstrap random sampling results by a Taylor series expansion. In turn, it requires that the statistic $\widehat{\theta} = T(X)$ is defined smoothly in $X$. An example of a non-smooth statistic is the sample mean, for which ABC intervals cannot be constructed. Moreover, the ABC method also requires the statistic $\widehat{\theta}$ to be represented in resampling form, i.e. $\widehat{\theta}^* = T(P^*)$, where $P^*$ is a resampling vector as described in Chapter 1, Section 1.2.

ABC intervals are both transformation-invariant and second-order accurate, just like BCa intervals. Table 14.2 of Efron and Tibshirani (1993) shows an example for which ABC intervals required only 3% of the computational effort needed for BCa intervals.

Given the computational power of computers nowadays, the ABC has lost relevance and is not commonly used as a bootstrap interval. On top of that, the mathematics of the ABC method are rather involved. Due to these reasons, we have decided not to explain ABC intervals more thoroughly. However, for the eager reader, we suggest reading Chapter 22 of Efron and Tibshirani (1993).

Like in the previous section, to finish off this section we present Table 3.2 with a summary of some properties of second-order accurate bootstrap confidence intervals, and whether they are implemented in the R package *boot* used for the simulations in next chapter.

Table 3.2: Properties of second-order accurate bootstrap confidence intervals.

| Properties of 2nd-order intervals | BT | BCa | ABC |
|---|---|---|---|
| Suffers from narrowness bias | no | yes | yes |
| Corrects for bias | yes[1] | yes | yes |
| Transformation invariant | no | yes | yes |
| Range preserving | no | yes | yes |
| Implemented in R package *boot* | yes | yes | no |

[1] Not explicitly, but it is unaffected by bias (Hesterberg, 2015).

## 3.3 Double Bootstrap

The double bootstrap is an iterated bootstrap procedure, in which bootstrap samples are iteratively bootstrapped. It is mainly used for two purposes: bias reduction in point estimation and reduction of coverage error in confidence intervals. Although it can be applied iteratively more than twice, it is commonly known as the double bootstrap as a double iteration is the most commonly used. A higher number of iterations is usually computationally unfeasible.

### 3.3.1 Double bootstrap for bias reduction

The bias of an estimator $\widehat{\theta}$ in estimating the parameter $\theta$ is defined as

$$\text{bias}(\widehat{\theta}) = \mathbb{E}(\widehat{\theta}) - \theta \tag{3.45}$$

If we knew the bias of $\widehat{\theta}$, then an unbiased estimate of $\theta$ could be constructed, namely $\hat{\theta} - \text{bias}(\widehat{\theta})$. We could, however, use the bootstrap to get an estimate of the bias of $\widehat{\theta}$. Let $X = \{X_1, \ldots, X_n\}$ be the original sample from the population. Assuming consistency of the bootstrap, it can be estimated by

$$\widehat{\text{bias}}(\widehat{\theta}) = \mathbb{E}^*(\widehat{\theta}^* \mid X) - \hat{\theta} \tag{3.46}$$

where $\hat{\theta}^*$ is the value of $\widehat{\theta}$ evaluated at a bootstrap sample $X^*$. Consequently, the bias-corrected bootstrap estimate of $\theta$ is

$$\widehat{\theta}_1 = \hat{\theta} - \widehat{\text{bias}}(\widehat{\theta}) = 2\hat{\theta} - \mathbb{E}^*(\widehat{\theta}^* \mid X) \tag{3.47}$$

where the subscript 1 means that the bias of $\widehat{\theta}$ has been corrected once.

We can keep using the bootstrap iteratively to reduce the bias of $\widehat{\theta}_1$. Let us change the standard notation for bootstrap samples so we avoid adding potentially infinite stars '*' for the nested bootstrap samples of $X$ and the nested bootstrap estimates of $\widehat{\theta}$. Let $X^{(1)}$ denote the new notation for a first-layer bootstrap sample $X^*$, $X^{(2)}$ for a second-layer bootstrap sample $X^{**}$, and so on. Then $X^{(k+1)}$ is defined as the $(k+1)$th-layer bootstrap sample, sampled from the $k$th-layer bootstrap sample $X^{(k)}$, $k \geq 0$. Consequently, let $X^{(0)}$ be the new notation for $X$. Similarly, let $\widehat{\theta}^{(1)}$ denote the new notation for $\widehat{\theta}^*$, and let $\hat{\theta}^{(k)}$ be the value of $\widehat{\theta}$ evaluated at $X^{(k)}$, where $\widehat{\theta}^{(0)}$ is the new notation for $\widehat{\theta}$. Martin (1990b) shows that the bootstrap estimator for $\theta$ after $j$ bias corrections can be computed as

$$\widehat{\theta}_j = \sum_{k=0}^{j} \binom{j+1}{k+1} (-1)^k \, \mathbb{E}^* \left( \widehat{\theta}^{(k)} \mid X \right) \tag{3.48}$$

The estimator $\widehat{\theta}_j$ has bias of order $n^{-(j+1)}$. In some instances, the limit $\widehat{\theta}_\infty = \lim_{j \to \infty} \widehat{\theta}_j$ of an infinite number of bias corrections exists (Martin, 1990a).

The value of $\mathbb{E}^*(\widehat{\theta}^{(k)} \mid X)$ can be approximated using a nested Monte Carlo simulation. For $k = 0$ ($\widehat{\theta}^{(0)} \equiv \widehat{\theta}$), resample $B$ times from $X^{(0)} \equiv X$ to obtain bootstrap samples $X^{(1)1}, \ldots, X^{(1)B}$. From each of these bootstrap samples $X^{(1)b}$, resample $B$ times to obtain a total of $B^2$ 2nd-layer bootstrap samples $X^{(2)1}, \ldots, X^{(2)B^2}$. Continue this process until all bootstrap samples $X^{(j)1}, \ldots, X^{(j)B^j}$ are obtained. Then:

$$\mathbb{E}^* \left( \widehat{\theta}^{(k)} \mid X \right) \approx \frac{1}{B^k} \sum_{i=1}^{B^k} \widehat{\theta}\left( X^{(k)i} \right) = \frac{1}{B^k} \sum_{i=1}^{B^k} \hat{\theta}^{(k)i}, \qquad k \geq 1 \tag{3.49}$$

That is, to estimate the expected value of $\widehat{\theta}^{(k)}$ given the data, we get bootstrap estimates $\hat{\theta}^{(k)}$ based on all the $k$th-layer bootstrap samples $X^{(k)i}$, $i = 1, \ldots, B^k$, and we average them all.

One clear drawback of this method is that it is computationally expensive, as $\mathcal{O}(B^j)$ resampling operations are required to compute $\mathbb{E}^*(\widehat{\theta}^{(j)} \mid X)$. Another thing to note is that the orders of bias are asymptotic. That means that when the sample size is small, high-order iterations can result in increased bias since the support of the distribution of $X^{(k)}$ shrinks as $k \to \infty$.

### 3.3.2 Double bootstrap for coverage correction

Also known as bootstrap calibration, this technique modifies the nominal level of a confidence interval (not necessarily a bootstrap interval) to try to make the true coverage

[0] Unknown.

get closer to the desired nominal coverage. Suppose we wish to construct a $\gamma$-confidence interval for a parameter $\theta$ based on some estimator $\widehat{\theta}$. Let $I^*(\gamma)$ be a nominal $\gamma$-level confidence bootstrap interval for the unknown parameter $\theta$ (could be a normal theory interval too). Denote the true coverage probability of $I^*(\gamma)$ by $\pi(\gamma)$, with $\pi(\gamma) = \mathbb{P}(\theta \in I^*(\gamma))$. The true coverage $\pi(\gamma)$ is often not exactly equal to the nominal coverage $\gamma$, but equal to

$$\pi(\gamma) = \mathbb{P}[\theta \in I^*(\gamma)] = \gamma + \Delta_n \tag{3.50}$$

with some approximation error $\Delta_n \to 0$ as $n \to \infty$. We could try to find a nominal coverage $\gamma'$ such that the true coverage of $I^*(\gamma')$ is exactly $\gamma$. That is

$$\pi(\gamma') = \mathbb{P}[\theta \in I^*(\gamma')] = \gamma \tag{3.51}$$

How do we find an estimate of $\gamma'$? We can use another level of bootstrap (double bootstrap) to get an estimate $\widehat{\gamma'}$ of $\gamma'$, which would yield

$$\pi(\widehat{\gamma'}) = \mathbb{P}[\theta \in I^*(\widehat{\gamma'})] = \gamma + \Delta'_n \tag{3.52}$$

where the new approximation error $\Delta'_n$ is typically smaller than $\Delta_n$.

### A second level of bootstrap

This part describes how to use a double bootstrap procedure to obtain an estimate of $\widehat{\gamma'}$ of $\gamma'$ based on Bühlmann and Mächler (2016). Assume the interval $I^*(\gamma)$ for $\theta$ is estimated using bootstrap samples $X_1^*, \ldots, X_M^*$. By taking $B$ bootstrap samples from each $X_i^*$ we get second-layer bootstrap samples $X_1^{**}, \ldots, X_B^{**}$, which we can use to construct a second-level confidence interval $I^{**}(\gamma)$ for $\widehat{\theta}$. An estimate of the true coverage of $I^{**}(\gamma)$ is

$$\hat{\pi}^*(\gamma) := \mathbb{P}^*[\widehat{\theta} \in I^{**}(\gamma)] \tag{3.53}$$

We are interested in finding a nominal coverage $\gamma'^*$ such that $\hat{\pi}^*(\gamma'^*) = \gamma$. Now, if we use

$$\gamma'^* = \hat{\pi}^{*-1}(\gamma) \tag{3.54}$$

we get that

$$\mathbb{P}^*[\widehat{\theta} \in I^{**}(\gamma'^*)] = \hat{\pi}^*(\hat{\pi}^{*-1}(\gamma)) = \gamma \tag{3.55}$$

is an exact confidence interval "in the bootstrap world" for the "parameter" $\widehat{\theta}$. Hence, the bootstrap estimate for the adjusted nominal coverage is $\widehat{\gamma'} = \gamma'^*$.

**Computation of bootstrap adjusted nominal level $\gamma'^*$**

1) Draw a bootstrap sample $X_i^*$ by resampling with replacement from the original data $X$. Then:

   a) Compute a $\gamma$-level bootstrap confidence interval $I_i^{**}(\gamma)$ for $\widehat{\theta}$ based on $X_i^*$. That is, generate second-level bootstrap samples $X_1^{**}, \ldots, X_B^{**}$ to get estimates $\widehat{\theta}_1^{**}, \ldots, \widehat{\theta}_M^{**}$ (here the subscript of $\widehat{\theta}$ denotes the value of the $i$-th $\widehat{\theta}(X_i^{**})$, and not the value of $\widehat{\theta}$ for an original sample with sample size $i$) and use them to construct $I_i^{**}(\gamma)$ using any of the methods for bootstrap confidence intervals described in Section 3.1 or 3.2.

   b) Evaluate whether the "parameter" $\widehat{\theta}$ in the "bootstrap world" is in $I_i^{**}(\gamma)$, i.e. compute $\mathbb{1}\{\widehat{\theta} \in I_i^{**}(\gamma)\}$.

2) Repeat step 1) $M$ times. Then use

$$p^*(\gamma) = \frac{1}{M} \sum_{i=1}^{M} \mathbb{1}\{\widehat{\theta} \in I_i^{**}(\gamma)\} \tag{3.56}$$

   as an approximation for $\hat{\pi}^*(\gamma)$.

3) Compute $p^*$ for different values of $\gamma$ to find $\gamma'^*$ with

$$p^*(\gamma'^*) = \gamma \tag{3.57}$$

   and use $\widehat{\gamma'} = \gamma'^*$. The search for $\gamma'^*$ is a zero finding problem, and can be done using a grid search or a bisection strategy.

There are a total number of $B \cdot M$ bootstrap samples to compute. Normally $M < B$, since the magnitude of $M$ only determines the approximation for computing the actual level $\mathbb{P}^*[\widehat{\theta} \in I^{**}(\gamma)]$ ($I^{**}$ being computed with $B$ bootstrap samples).

In regular cases, bootstrap coverage correction reduces coverage error in confidence intervals by a factor of order $n^{-1/2}$ for one-sided intervals (Martin, 1990a). Consequently, first-order intervals have coverage error reduced from $\mathcal{O}_{\mathbb{P}}(n^{-1/2})$ to $\mathcal{O}_{\mathbb{P}}(n^{-1})$, while second-order intervals have coverage error reduced from $\mathcal{O}_{\mathbb{P}}(n^{-1})$ to $\mathcal{O}_{\mathbb{P}}(n^{-3/2})$. However, Martin (1990b) states that "higher order iterations do not cause one-sided critical points to be third-order correct". This means that the reduction in coverage error remains the same even if we add higher order iterations to the bootstrap coverage correction algorithm.

### 3.3.3 Advantages and disadvantages of bootstrap coverage correction

Here we highlight some of the advantages and disadvantages of the bootstrap coverage correction procedure as presented in Martin (1990b).

Some of the advantages of bootstrap coverage correction are:

1) Transformation invariance. The iterated bootstrap method is invariant under monotone transformations provided the method used to construct the original interval has this property, such as the bootstrap percentile interval.

2) Coverage correction is not necessarily a nested bootstrap algorithm. If the original interval $I^*(\gamma)$ is a non-bootstrap interval, only a single level of bootstrapping is required.

3) Asymptotic high-order coverage accuracy in confidence intervals can be obtained by adjusting critical points using a one-term Edgeworth correction.

4) Coverage correction is available quite generally. Not only can be applied to bootstrap or normal theory intervals, but to other types of intervals.

On the other side, there are two major disadvantages to take into account:

1) Computationally very expensive. It requires $B \cdot M$ bootstrap sample for a single iteration. The magnitude of $B$ and $M$ for an accurate coverage correction is also an open problem.

2) Its performance with small sample size might be terrible. As the number of iterations $j$ of the algorithm grows, the number of distinct points resampled decreases rapidly, shrinking almost surely to one. Hence, the quality of bootstrap estimates degrades as $j \to \infty$, likely producing larger coverage errors rather than smaller. In many cases though, one application of the algorithm is enough to reduce coverage error considerably.

## 3.4 R Packages with Bootstrap Implementations

A quick search on CRAN (as of October 2019) yields 75 packages that have the term "bootstrap" in their description. They range from specific implementations of some bootstrap confidence intervals to general-purpose implementations, bootstrap for times series, bootstrap for likelihood-ratio tests, etc.

The most widely used packages for general-purpose implementations are the *boot* (Canty and Ripley, 2019) and the *bootstrap* package (original, from StatLib, and by Rob Tibshirani. R port by Friedrich Leisch., 2019). The *boot* package is based on Davison and Hinkley (1997), whereas the *bootstrap* package is based on Efron and Tibshirani (1993). The *bootstrap* package description says that "new projects should preferentially use the

recommended package 'boot' ". So, following that recommendation, the *boot* package is the one used by default in this thesis.

The *boot* package has a function called "boot.ci" which computes a number of bootstrap confidence intervals by default. These are: normal ($z$ interval with bootstrap SE and bias correction), basic (RP interval), student (bootstrap $t$ interval), percentile interval and BCa interval.

It is also worth noting that the *boot* package uses the "R-6" (type 6 in R) method of computing quantiles. That means, the number of bootstrap sample $B$ has to be chosen such that $r = (B + 1)\alpha$ is a whole number, where the $r$-th order statistic of the sample is the desired $\alpha$-quantile.

# Chapter 4

# Coverage Analysis of Bootstrap CIs for Common Estimators

In this chapter, we present the results from a coverage analysis of bootstrap confidence intervals via simulation studies. These simulations investigate the actual coverage that one gets for a parameter $\theta$ using some bootstrap confidence intervals for common estimators in different scenarios. The parameters studied were the mean, median and correlation coefficient, and the estimators the sample mean, sample median and sample Pearson correlation coefficient. The bootstrap confidence intervals analysed were those returned by the function *boot.ci* from the R package *boot*, plus a custom implementation of the double bootstrap for the percentile interval for coverage correction. We created the different scenarios by changing the population distribution $F$ to five different distributions, summarized in Table 4.1 and Table 4.2, and by varying the sample size to 10, 100 and 1000 samples.

Table 4.1: Distributions used in the simulations for the sample mean and sample median.

| Distribution | Parameters |
|---|---|
| Normal | $\mu = 0,\ \sigma^2 = 1$ |
| Lognormal[1] | $\mu = 1,\ \sigma^2 = 0.4^2$ |
| Exponential | $\lambda = 1$ |
| Student $t$ | $df = 5$ |
| Bimodal[2] | $N_1 \sim \mathcal{N}(-1.5,\ 1), \quad N_2 \sim \mathcal{N}(1.5,\ 1)$ |

[1] Parameters of untransformed normal distribution $\log(X) \sim \mathcal{N}(\mu, \sigma^2)$.
[2] Normal mixture of $N_1$ and $N_2$.

Table 4.2: Distributions used in the simulations for the sample Pearson correlation coefficient.

| Distribution | Parameters |
|---|---|
| Bivariate normal | $\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ |
| Bivariate Student $t$ | $df = \begin{pmatrix} 5 \\ 5 \end{pmatrix}$, $\Sigma^1 = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ |

[1] $\Sigma$ is the scale matrix, not the covariance matrix, which is equal to $\Sigma \cdot \frac{\nu}{\nu-2}$, with $\nu$ the degrees of freedom. This does not affect the correlation matrix, which equals $\Sigma$.

## 4.1 Noncoverage Plots to Assess the Performance of a Confidence Interval

Before explaining the setup and presenting the results, we first present and explain the coverage plots we have used to convey the results. These plots are not unique to bootstrap intervals, but can be used for other types of confidence intervals.

A confidence interval is a random interval that depends on the sample data $X_1, \ldots, X_n \sim F$. If we knew $F$, we could draw many random samples from $F$ and compute many confidence intervals for some parameter $\theta$ of interest. Figure 4.1 illustrates the process of generating a 100 confidence intervals for a parameter $\theta$ with 90% confidence level. The lines in red are confidence intervals that missed the true parameter $\theta$, and lines in blue do contain $\theta$. This is an example of a confidence interval with symmetric noncoverage around $\theta$. That means, the proportion of intervals that miss $\theta$ to the right is the same as the proportion of the intervals that miss $\theta$ to the left.

However, this need not be the case. Figure 4.2 shows an example of a confidence interval with asymmetric noncoverage around $\theta$. It can be seen clearly that the confidence interval tends to miss $\theta$ more often to the right than to the left.

We used one-sided noncoverage plots from Hesterberg (2015) as a base for our noncoverage plots. These plots show the one-sided noncoverage of a confidence interval against different sample sizes, plotted as points. We replaced the points by horizontal lines, which represent the magnitude of the median interval length. The size of these horizontal lines is symbolic, and it has been normalised such that, for a specific $n$, the interval with the maximum median length gets the widest horizontal line. The other intervals

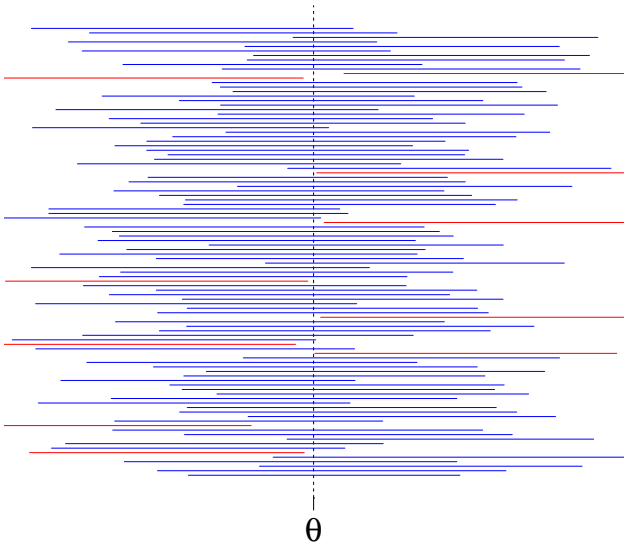


Figure 4.1: A 100 confidence intervals with 90% confidence level and symmetric noncoverage. Intervals that miss $\theta$ are marked in red. There are approximately the same number of intervals that miss $\theta$ to the left and to the right.

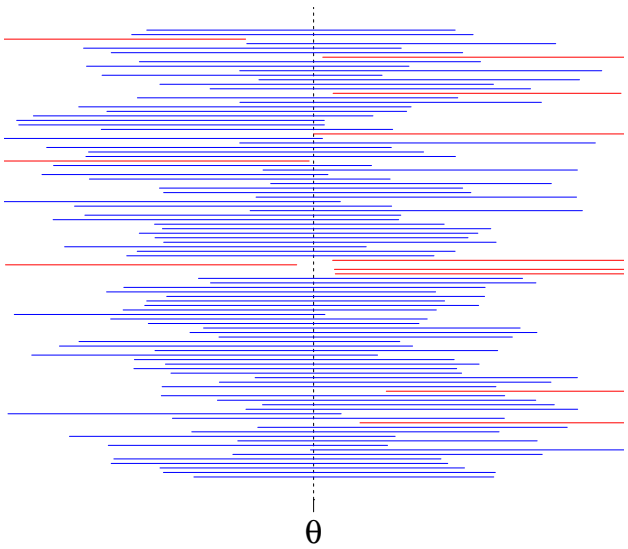


Figure 4.2: A 100 confidence intervals with 90% confidence level and asymmetric noncoverage. Intervals that miss $\theta$ are marked in red. There are more intervals that miss $\theta$ to the right than to the left.

have shorter horizontal lines in the same proportion as their median length is smaller to the maximum median length.

## 4.2 Simulation Setup

For the sample mean and sample median, we ran simulations for each distribution in Table 4.1. We had 5 simulation scenarios for the sample mean and 5 for the sample median. For the sample Pearson correlation coefficient, we had a total of 8 simulation scenarios. We had one simulation scenario for each distribution in Table 4.2, varying other two settings: the true value of the correlation coefficient $\rho$ to 0.5 and 0.9 and applying/not applying Fisher's $z$-transformation to the sample Pearson correlation coefficient. The former varies the asymmetry of the sample Pearson correlation coefficient (larger for larger $\rho$), whereas the latter help some bootstrap confidence intervals be more accurate, especially those that do not find normalising transformations automatically.

We varied the sample size of the original sample to 10, 100 and 1000. All bootstrap confidence intervals were calculated using the value $\alpha = 0.05$ of type I error as the nominal value.

For each simulation scenario we ran $R$ simulations. In each simulation, we computed the endpoints of $(1-\alpha)$-confidence bootstrap intervals, each constructed from $B = 14999$ bootstrap estimates $\widehat{\theta}^{*b}$, $b = 1, \ldots, B$. The value of $R$ differs depending on the scenario. For most scenarios, $R = 1000$ for all types of bootstrap confidence intervals, except the double bootstrap for the percentile interval, for which $R = 500$, due to time constraints (they take longer to compute and have to be computed separately). For the sample mean, we ran $R = 10000$ simulations, since its variance has a theoretical formula ($\sigma^2/n$) that can be implemented straight away and there is no need to use a second layer of bootstrapping to compute the variance of the sample mean. For all other cases, the variance of the estimator (sample median, sample Pearson correlation coefficient and transformed sample Pearson correlation coefficient) had to be estimated using a second layer of bootstrapping, drawing $M = 200$ second-layer bootstrap estimates $\widehat{\theta}^{**m}$, $m = 1, \ldots, M$.

All of this thesis' work was done using R 3.6.1 (R Core Team, 2019). Simulations were done using the *boot* package (v1.3-23; Canty and Ripley (2019)), for bootstrap computations. The package *nor1mix* (v1.3-0.; Mächler (2017)) was used to draw random samples from the bimodal normal mixture presented in Table 4.1. The package *mvtnorm* (v1.0-11; Genz, Bretz, Miwa, Mi, Leisch, Scheipl, and Hothorn (2019), Genz and Bretz (2009)) was used to draw random samples from both bivariate normal and Student $t$ distributions presented in Table 4.2. The packages *parallel* (R Core Team, 2019), *doParallel* (v1.0.15; Corporation and Weston (2019)) and *foreach* (v1.4.7; R Core Team (2019)) were used for the parallel computation of the simulations.

As mentioned above, the double bootstrap implementation (for coverage correction) for

the percentile interval was computed separately. This is because one has to compute several times a percentile interval ($L = 200$ times in our case) for different nominal coverages (nominal values of $\alpha$), and then pick the $\alpha$ whose corresponding actual coverage is closest to the desired $\alpha = 0.05$. This means that, while for the other bootstrap intervals $B$ computations of the estimator are needed, for the double bootstrap we need $a \cdot L \cdot B$ computations, where $a$ is the number of nominal coverages used. In our simulations, the nominal coverages we tried were $\boldsymbol{\alpha} = \{0.01, 0.03, 0.05, 0.07, 0.09\}$ and hence $a = 5$. If we were to compute all bootstrap intervals (including the double bootstrap implementation) in the same thread, it would be very inefficient since the double bootstrap would take much longer to compute.

In addition, we added an interpolation step after choosing the optimal value of $\alpha$. It tries to refine the optimal value of $\alpha$ based on the empirical coverages obtained. Let $e_{opt}$ denote the empirical coverage closest to $\alpha = 0.05$ associated with nominal coverage $\alpha_{opt}$. Let $e_{subopt}$ denote the empirical coverage closest to $\alpha = 0.05$ that leaves $\alpha = 0.05$ in between itself and $e_{opt}$, associated with nominal coverage $\alpha_{subopt}$. Then, the new optimal nominal coverage $\alpha_{new}$ can be calculated from the linear relationship between the nominal and empirical coverages:

$$\frac{\alpha_{new} - \alpha_{opt}}{0.05 - e_{opt}} = \frac{\alpha_{subopt} - \alpha_{opt}}{e_{subopt} - e_{opt}} \quad , \alpha_{opt} < \alpha_{subopt} \tag{4.1}$$

which results in

$$\alpha_{new} = \alpha_{opt} + \frac{\alpha_{subopt} - \alpha_{opt}}{e_{subopt} - e_{opt}}(0.05 - e_{opt}) \quad , \alpha_{opt} < \alpha_{subopt} \tag{4.2}$$

For $\alpha_{opt} > \alpha_{subopt}$, the same relationship holds but one has to swap $\alpha_{opt}$ for $\alpha_{subopt}$ and $e_{opt}$ for $e_{subopt}$.

Table 4.3 shows an example where a linear interpolation as described above can be applied. The green values are $\alpha_{opt} = 0.03$ and $e_{opt} = 0.045$, whereas the yellow values are $\alpha_{subopt} = 0.05$ and $e_{subopt} = 0.065$. Pluggin in these values in (4.2) gives $\alpha_{new} = 0.035$.

The whole idea about the above interpolating scheme is to avoid the computational burden of computing the empirical coverage for more nominal values. Instead of computing

Table 4.3: Example of values taken for linear interpolation between nominal and empirical coverages. The interpolated nominal coverage that gives an empirical coverage of $\alpha = 0.05$ is 0.035.

| Nominal coverage | 0.01 | 0.03 | 0.05 | 0.07 | 0.09 |
|---|---|---|---|---|---|
| Empirical coverage | 0.02 | 0.045 | 0.065 | 0.08 | 0.095 |

the empirical coverage for let's say $\boldsymbol{\alpha} = (0.01, 0.02, \ldots, 0.09)$, or even a more fine-grained $\boldsymbol{\alpha}$, we can get away with just computing $\boldsymbol{\alpha} = (0.01, 0.03, 0.05, 0.07, 0.09)$ and interpolate for values in between. Obviously it will not be as precise, but it is better than not interpolating at all.

## 4.3 Coverage of Bootstrap Confidence Intervals for the Sample Mean

In this section we study the coverage of some bootstrap confidence intervals for the sample mean. Let $X_1, \ldots X_n \sim F$, with $\mathbb{E}(X_1) = \mu$ and $\mathrm{Var}(X_1) = \sigma^2$. The sample mean $\bar{X}$ is an estimator for the mean $\mu$ of a distribution $F$, and it is defined as

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \tag{4.3}$$

If $F$ is a normal distribution, $\bar{X}$ is known to follow a normal distribution. But when $F$ is not a normal distribution, $\bar{X}$ does not follow a normal distribution. However, according to the CLT, under some mild conditions on $F$, $\bar{X}$ asymptotically follows a normal distribution:

$$\frac{(\bar{X} - \mu)}{\sigma/\sqrt{n}} \xrightarrow{D} \mathcal{N}(0,\, 1) \tag{4.4}$$

This convergence justifies in many cases the use of a normal approximation to construct confidence intervals for $\mu$. But what value of $n$ is large enough for the approximation to work well is an open question, and depends on many things. If $F$ is assumed to be a normal distribution and $\sigma^2$ is estimated with $s^2 = (n-1)^{-1} \sum (X_i - \bar{X})^2$, then (4.4) is known to follow a Student $t$ distribution with $n - 1$ degrees of freedom, for all $n \geq 2$. Apart from this case and another handful of known cases, the distribution of $\bar{X}$ is usually unknown. Nonparametric bootstrap confidence intervals can easily outperform the normal approximation (4.4) since they do not make assumptions on the distribution of $F$. They can also pick up asymmetry from the distribution of $\bar{X}$ which further improves the results.

It is in this context in which we want to evaluate the performance of bootstrap confidence intervals by using noncoverage plots as defined in Section 4.1. Figures 4.3 to 4.7 and tables 4.4 to 4.8 show the results of the simulations described in Section 4.2 for the sample mean.

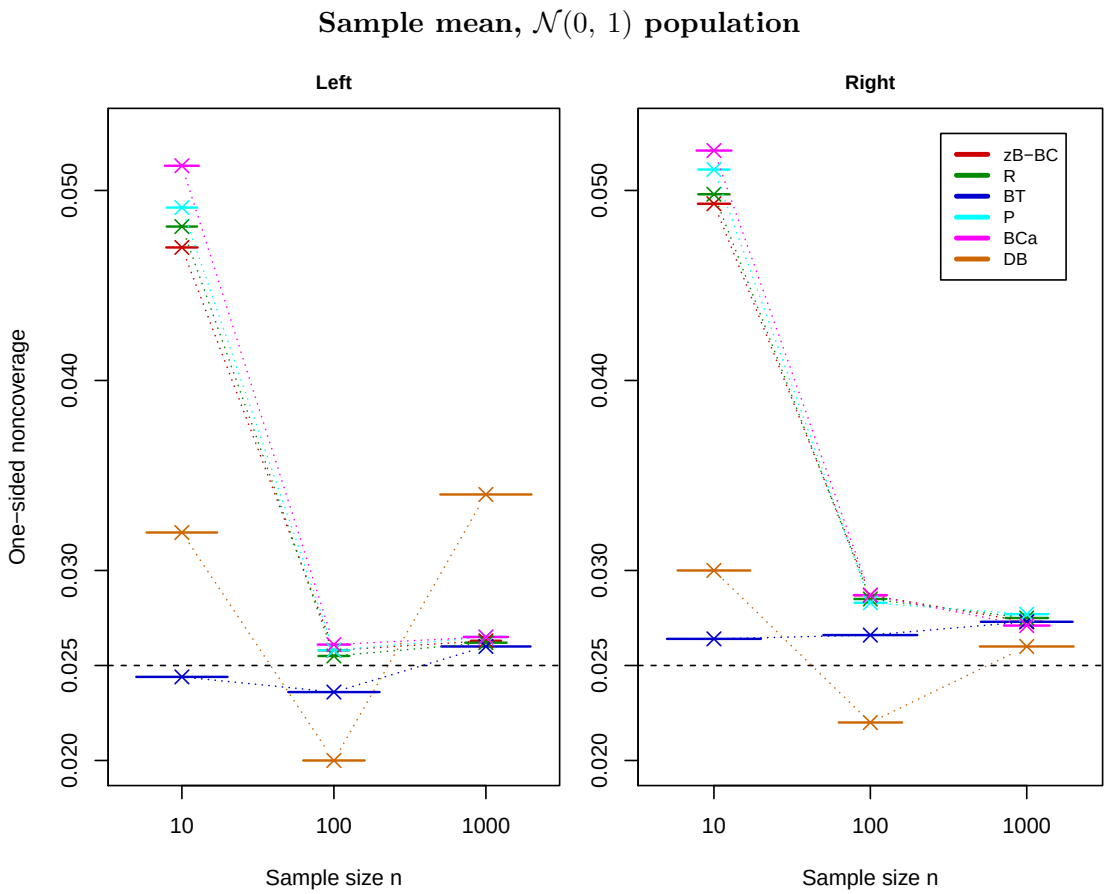


Figure 4.3: Sample mean of normal population: plot of noncoverage to the left (Left) and to the right (Right) of some bootstrap intervals.

Table 4.4: Sample mean of normal population: values of noncoverage to the left, noncoverage to the right and median length of some bootstrap intervals.

| | Sample mean, $\mathcal{N}(0,\ 1)$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Noncoverage left | | | Noncoverage right | | | Median length | | |
| | 10 | 100 | 1000 | 10 | 100 | 1000 | 10 | 100 | 1000 |
| zB-BC | 0.0470 | 0.0258 | 0.0263 | 0.0493 | 0.0287 | 0.0273 | 1.1323 | 0.3889 | 0.1238 |
| R | 0.0481 | 0.0255 | 0.0262 | 0.0498 | 0.0285 | 0.0275 | 1.1275 | 0.3890 | 0.1239 |
| BT | 0.0244 | 0.0236 | 0.0260 | 0.0264 | 0.0266 | 0.0273 | 1.4756 | 0.3960 | 0.1241 |
| P | 0.0491 | 0.0258 | 0.0265 | 0.0511 | 0.0283 | 0.0277 | 1.1275 | 0.3890 | 0.1239 |
| BCa | 0.0513 | 0.0261 | 0.0265 | 0.0521 | 0.0287 | 0.0271 | 1.1471 | 0.3892 | 0.1239 |
| DB | 0.0320 | 0.0200 | 0.0340 | 0.0300 | 0.0220 | 0.0260 | 1.3571 | 0.3925 | 0.1241 |

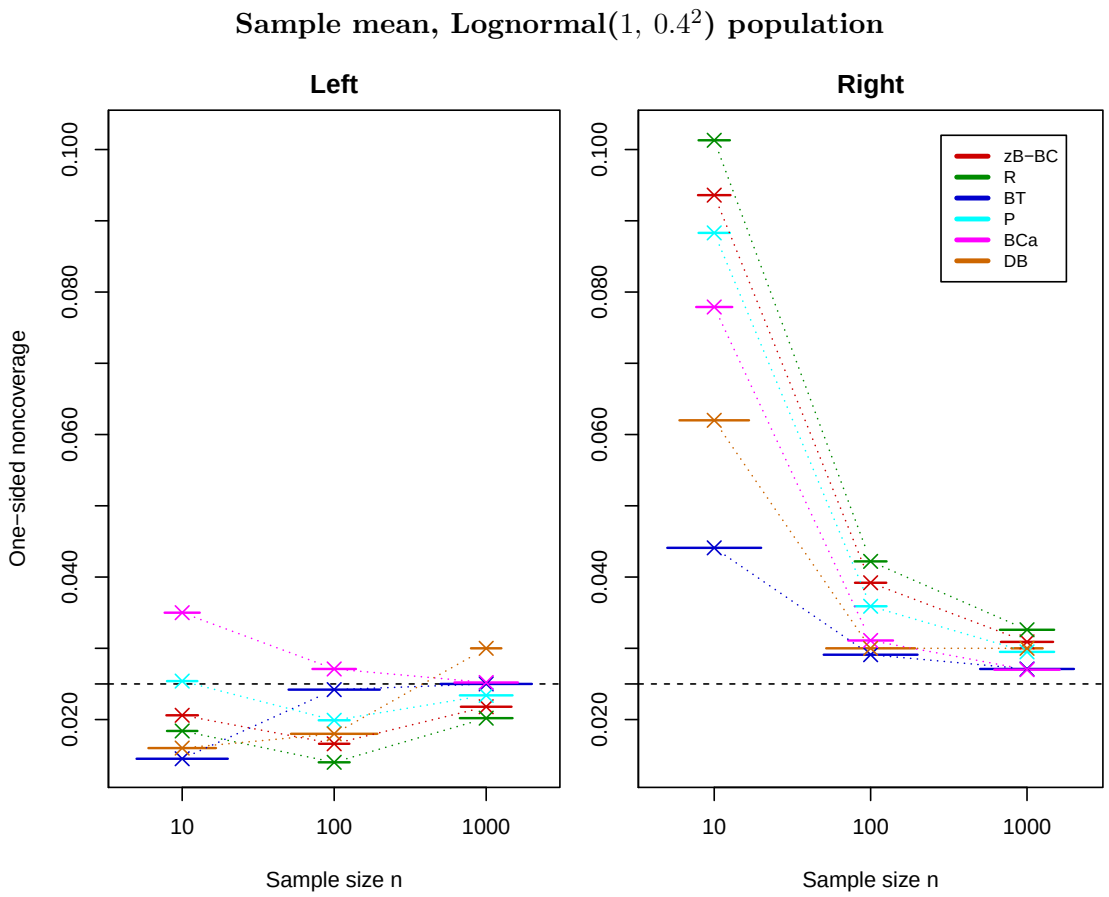


Figure 4.4: Sample mean of lognormal population: plot of noncoverage to the left (Left) and to the right (Right) of some bootstrap intervals.

Table 4.5: Sample mean of lognormal population: values of noncoverage to the left, noncoverage to the right and median length of some bootstrap intervals.

| | Sample mean, Lognormal(1, $0.4^2$) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Noncoverage left | | | Noncoverage right | | | Median length | | |
| | 10 | 100 | 1000 | 10 | 100 | 1000 | 10 | 100 | 1000 |
| zB-BC | 0.0206 | 0.0166 | 0.0218 | 0.0936 | 0.0392 | 0.0309 | 1.3046 | 0.4711 | 0.1517 |
| R | 0.0184 | 0.0140 | 0.0202 | 0.1013 | 0.0422 | 0.0326 | 1.2961 | 0.4712 | 0.1518 |
| BT | 0.0145 | 0.0242 | 0.0250 | 0.0441 | 0.0291 | 0.0271 | 1.7483 | 0.4851 | 0.1523 |
| P | 0.0254 | 0.0199 | 0.0234 | 0.0883 | 0.0359 | 0.0295 | 1.2961 | 0.4712 | 0.1518 |
| BCa | 0.0350 | 0.0271 | 0.0252 | 0.0779 | 0.0311 | 0.0270 | 1.3310 | 0.4741 | 0.1519 |
| DB | 0.0160 | 0.0180 | 0.0300 | 0.0620 | 0.0300 | 0.0300 | 1.5724 | 0.4840 | 0.1515 |

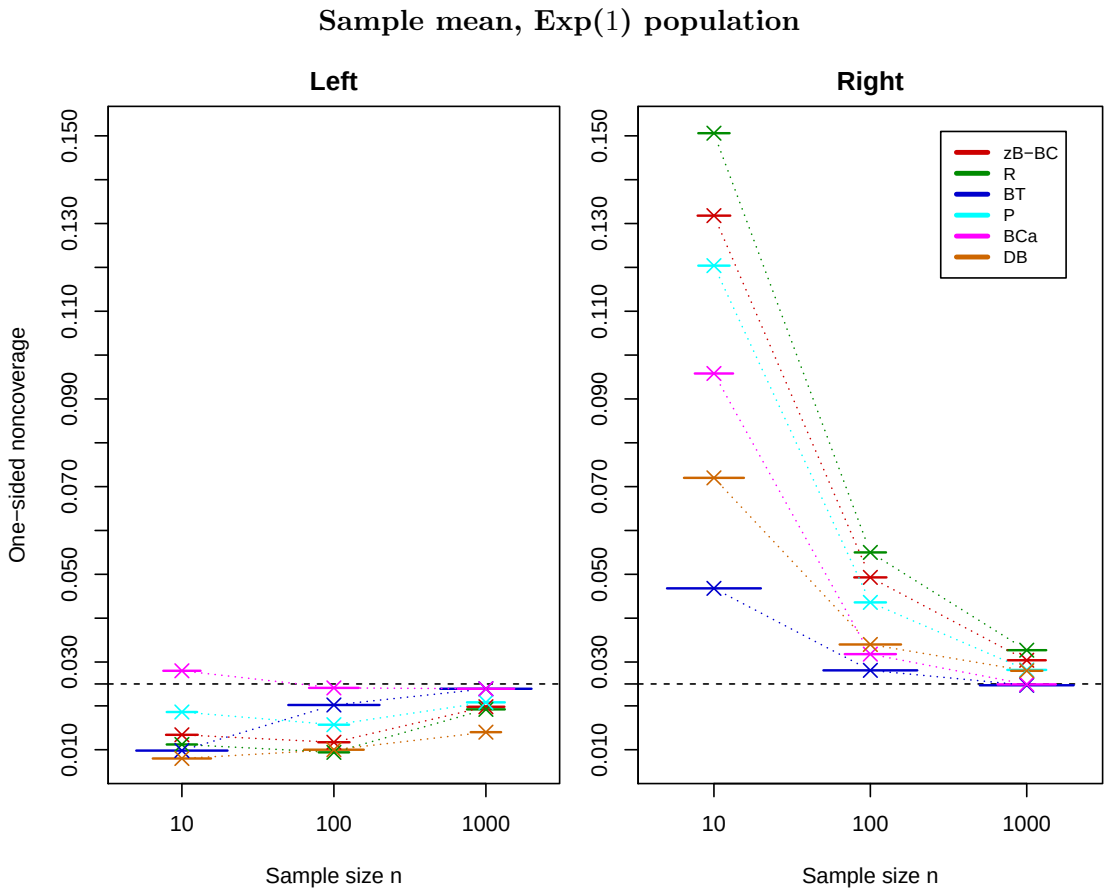


Figure 4.5: Sample mean of exponential population: plot of noncoverage to the left (Left) and to the right (Right) of some bootstrap intervals.

Table 4.6: Sample mean of exponential population: values of noncoverage to the left, noncoverage to the right and median length of some bootstrap intervals.

| | Sample mean, Exp(1) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Noncoverage left | | | Noncoverage right | | | Median length | | |
| | 10 | 100 | 1000 | 10 | 100 | 1000 | 10 | 100 | 1000 |
| zB-BC | 0.0134 | 0.0117 | 0.0198 | 0.1318 | 0.0493 | 0.0304 | 1.0126 | 0.3834 | 0.1237 |
| R | 0.0112 | 0.0094 | 0.0192 | 0.1506 | 0.0550 | 0.0327 | 1.0016 | 0.3831 | 0.1237 |
| BT | 0.0098 | 0.0202 | 0.0239 | 0.0468 | 0.0281 | 0.0247 | 1.5048 | 0.4023 | 0.1244 |
| P | 0.0186 | 0.0157 | 0.0208 | 0.1204 | 0.0436 | 0.0282 | 1.0016 | 0.3831 | 0.1237 |
| BCa | 0.0280 | 0.0241 | 0.0239 | 0.0958 | 0.0318 | 0.0249 | 1.0578 | 0.3892 | 0.1239 |
| DB | 0.0080 | 0.0100 | 0.0140 | 0.0720 | 0.0340 | 0.0280 | 1.2321 | 0.3923 | 0.1236 |

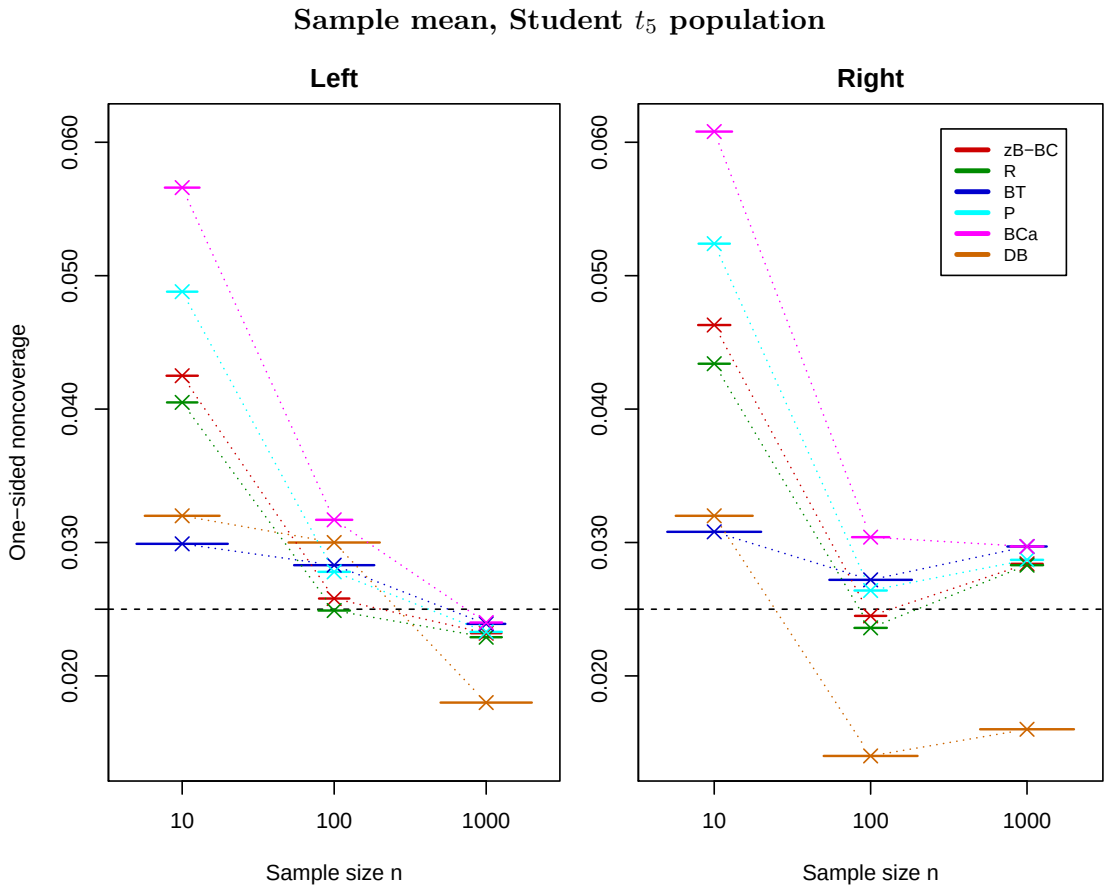


Figure 4.6: Sample mean of Student $t_5$ population: plot of noncoverage to the left (Left) and to the right (Right) of some bootstrap intervals.

Table 4.7: Sample mean of Student $t_5$ population: values of noncoverage to the left, noncoverage to the right and median length of some bootstrap intervals.

| | Sample mean, Student $t_5$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Noncoverage left | | | Noncoverage right | | | Median length | | |
| | 10 | 100 | 1000 | 10 | 100 | 1000 | 10 | 100 | 1000 |
| zB-BC | 0.0425 | 0.0258 | 0.0232 | 0.0463 | 0.0245 | 0.0284 | 1.3638 | 0.4934 | 0.1593 |
| R | 0.0405 | 0.0249 | 0.0229 | 0.0434 | 0.0236 | 0.0283 | 1.3577 | 0.4936 | 0.1593 |
| BT | 0.0299 | 0.0283 | 0.0239 | 0.0308 | 0.0272 | 0.0297 | 1.7784 | 0.5028 | 0.1596 |
| P | 0.0488 | 0.0278 | 0.0233 | 0.0524 | 0.0264 | 0.0287 | 1.3577 | 0.4936 | 0.1593 |
| BCa | 0.0566 | 0.0317 | 0.0240 | 0.0608 | 0.0304 | 0.0297 | 1.3874 | 0.4945 | 0.1593 |
| DB | 0.0320 | 0.0300 | 0.0180 | 0.0320 | 0.0140 | 0.0160 | 1.6648 | 0.5049 | 0.1615 |

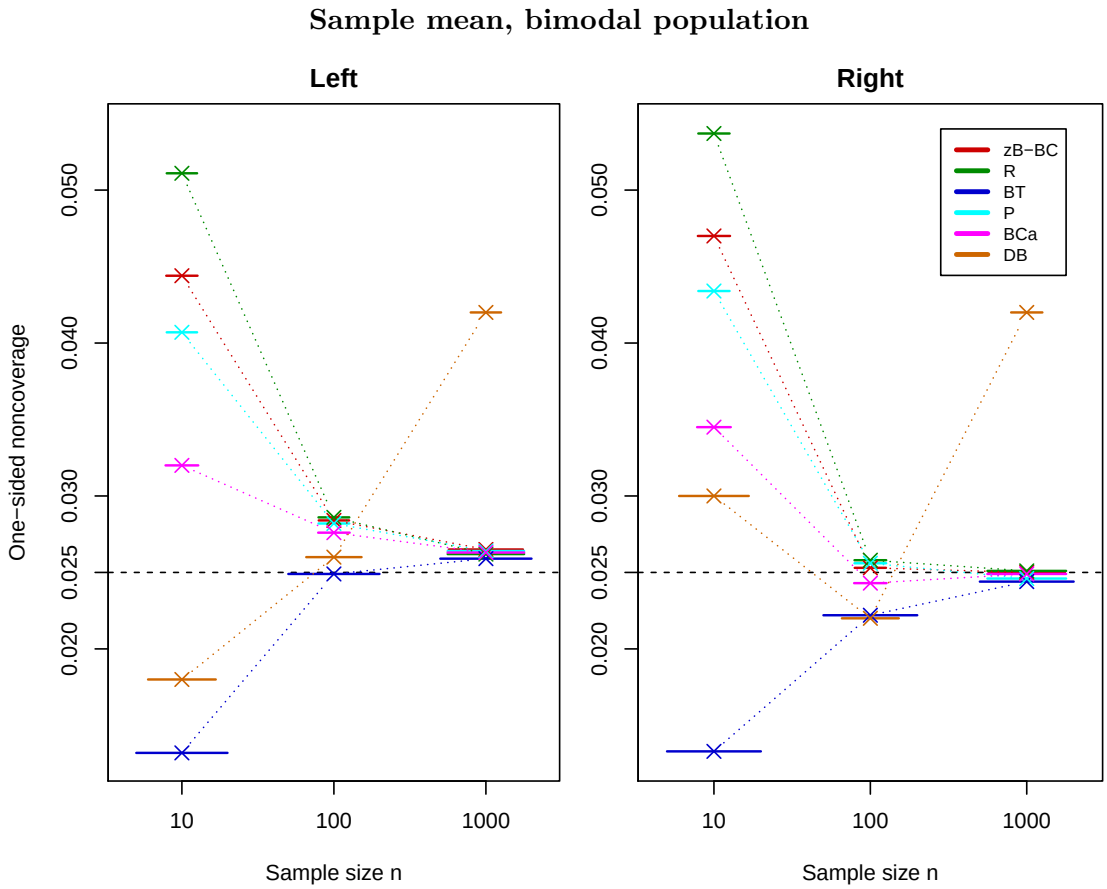


Figure 4.7: Sample mean of bimodal population: plot of noncoverage to the left (Left) and to the right (Right) of some bootstrap intervals.

Table 4.8: Sample mean of bimodal population: values of noncoverage to the left, noncoverage to the right and median length of some bootstrap intervals.

| | Sample mean, bimodal | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Noncoverage left | | | Noncoverage right | | | Median length | | |
| | 10 | 100 | 1000 | 10 | 100 | 1000 | 10 | 100 | 1000 |
| zB-BC | 0.0610 | 0.0530 | 0.0480 | 0.0930 | 0.0560 | 0.0400 | 3.1257 | 1.3576 | 0.4661 |
| R | 0.1240 | 0.0950 | 0.0550 | 0.1760 | 0.0900 | 0.0560 | 2.9726 | 1.3303 | 0.4619 |
| BT | 0.0500 | 0.0380 | 0.0380 | 0.0700 | 0.0410 | 0.0390 | 3.9890 | 1.5622 | 0.4848 |
| P | 0.0270 | 0.0270 | 0.0360 | 0.0360 | 0.0300 | 0.0270 | 2.9726 | 1.3303 | 0.4619 |
| BCa | 0.0180 | 0.0200 | 0.0310 | 0.0640 | 0.0400 | 0.0300 | 2.9720 | 1.3184 | 0.4620 |
| DB | 0.0240 | 0.0200 | 0.0360 | 0.0260 | 0.0220 | 0.0340 | 3.1352 | 1.3415 | 0.4557 |

## 4.4 Coverage of Bootstrap Confidence Intervals for the Sample Median

The median of a distribution $F$ is the value that leaves 50% of $F$ above it and 50% below it. The sample median is an estimator of the median of a distribution $F$, which is usually used as a measure of location. Usually it is also used as a robust estimator for the mean of a population.

Let $X_1, \ldots, X_n \sim F$, with $\mathbb{E}(X_1) = \mu$ and $\text{Var}(X_1) = \sigma^2$. The sample median of $X_1, \ldots, X_n$ is defined as

$$\widehat{\text{median}}(X_1, \ldots, X_n) = \begin{cases} X_{\left(\frac{n+1}{2}\right)} & \text{if } n \text{ odd} \\ \frac{X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n}{2}+1\right)}}{2} & \text{if } n \text{ even} \end{cases} \tag{4.5}$$

where $X_{(k)}$ is the $k$-th order statistic.

Like in the case of the sample median, there is a CLT expansion for the sample median that holds asymptotically. Under some regularity conditions, the sample median asymptotically follows a normal distribution:

$$\sqrt{n}\left(\widehat{\text{median}}(X_1, \ldots, X_n) - \text{median}(F)\right) \xrightarrow{D} \mathcal{N}\left(0, \frac{1}{4f^2(\text{median}(F))}\right) \tag{4.6}$$

where $f(\cdot)$ is the PDF of $F$.

This asymptotic result can be used to construct confidence intervals for the median of $F$. One could use $\widehat{\text{median}}(X_1, \ldots, X_n)$ as an estimate of $\text{median}(F)$ to compute the asymptotic variance in (4.6). However, unlike in the case of the sample mean, there is no exact distribution for $\widehat{\text{median}}(X_1, \ldots, X_n)$ for finite $n$ under any assumption of the distribution $F$.

Again, like in the previous section, we want to investigate the performance of bootstrap confidence intervals by using noncoverage plots as defined in Section 4.1. Figures 4.8 to 4.12 and tables 4.9 to 4.13 show the results of the simulations described in Section 4.2 for the sample median.

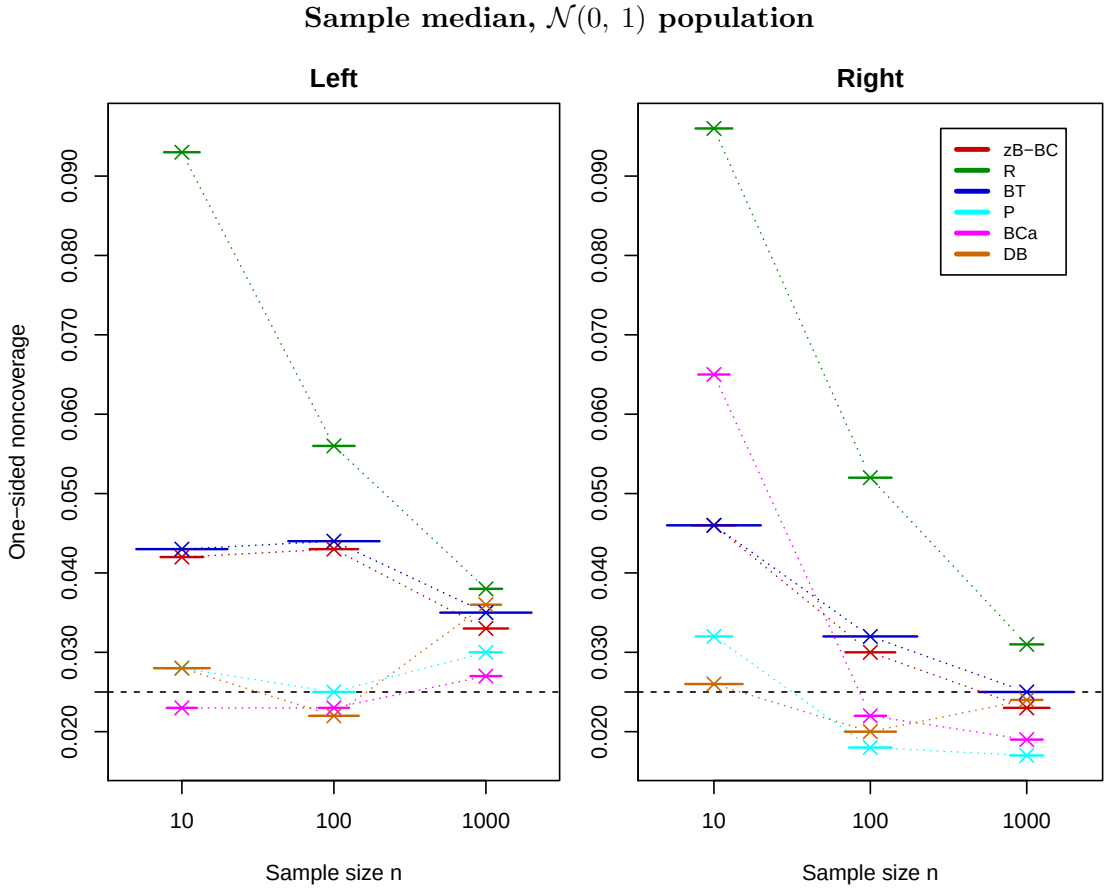


Figure 4.8: Sample median of normal population: plot of noncoverage to the left (Left) and to the right (Right) of some bootstrap intervals.

Table 4.9: Sample median of normal population: values of noncoverage to the left, noncoverage to the right and median length of some bootstrap intervals.

| | Sample median, $\mathcal{N}(0,\ 1)$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Noncoverage left | | | Noncoverage right | | | Median length | | |
| | 10 | 100 | 1000 | 10 | 100 | 1000 | 10 | 100 | 1000 |
| zB-BC | 0.0420 | 0.0430 | 0.0330 | 0.0460 | 0.0300 | 0.0230 | 1.4740 | 0.4833 | 0.1568 |
| R | 0.0930 | 0.0560 | 0.0380 | 0.0960 | 0.0520 | 0.0310 | 1.4377 | 0.4791 | 0.1559 |
| BT | 0.0430 | 0.0440 | 0.0350 | 0.0460 | 0.0320 | 0.0250 | 1.7235 | 0.5088 | 0.1604 |
| P | 0.0280 | 0.0250 | 0.0300 | 0.0320 | 0.0180 | 0.0170 | 1.4377 | 0.4791 | 0.1559 |
| BCa | 0.0230 | 0.0230 | 0.0270 | 0.0650 | 0.0220 | 0.0190 | 1.4103 | 0.4726 | 0.1558 |
| DB | 0.0280 | 0.0220 | 0.0360 | 0.0260 | 0.0200 | 0.0240 | 1.5413 | 0.4840 | 0.1558 |

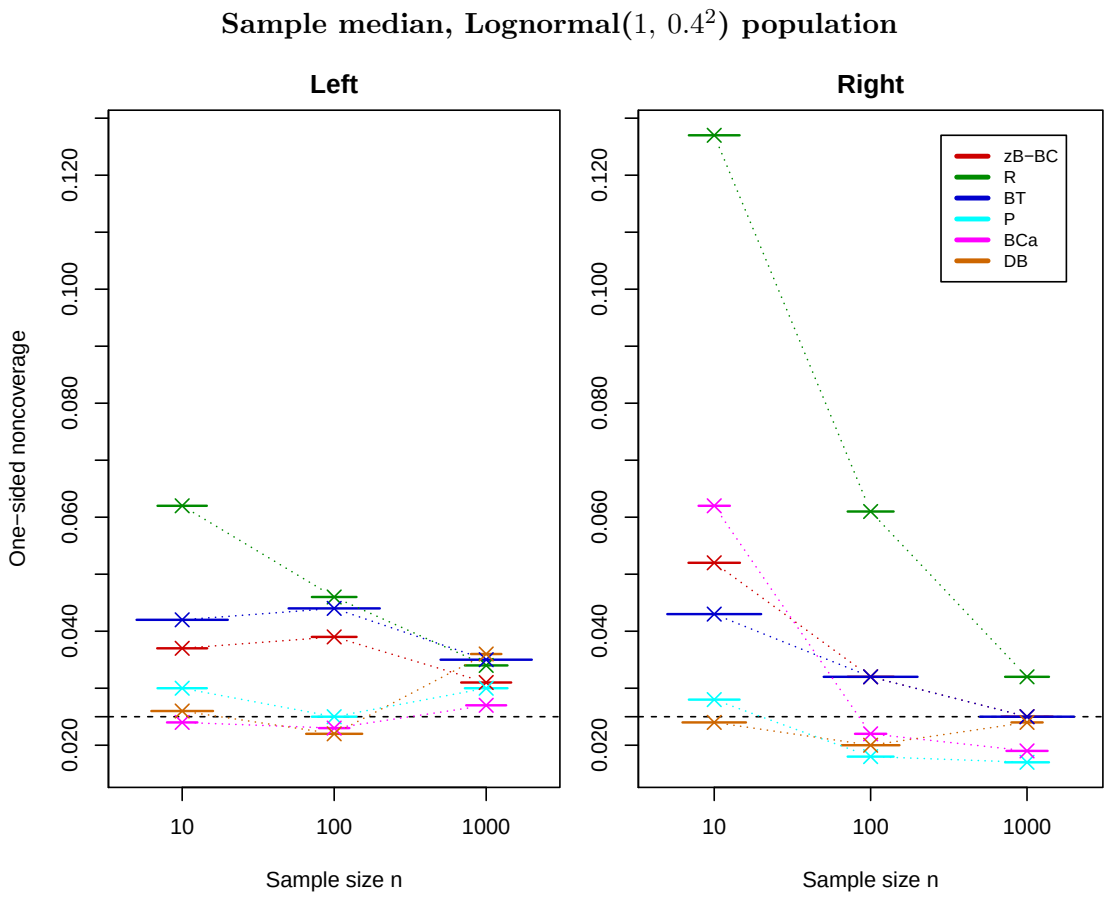


Figure 4.9: Sample median of lognormal population: plot of noncoverage to the left (Left) and to the right (Right) of some bootstrap intervals.

Table 4.10: Sample median of lognormal population: values of noncoverage to the left, noncoverage to the right and median length of some bootstrap intervals.

| | Sample median, Lognormal(1, $0.4^2$) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Noncoverage left | | | Noncoverage right | | | Median length | | |
| | 10 | 100 | 1000 | 10 | 100 | 1000 | 10 | 100 | 1000 |
| zB-BC | 0.0370 | 0.0390 | 0.0310 | 0.0520 | 0.0320 | 0.0250 | 1.6145 | 0.5233 | 0.1706 |
| R | 0.0620 | 0.0460 | 0.0340 | 0.1270 | 0.0610 | 0.0320 | 1.6099 | 0.5232 | 0.1700 |
| BT | 0.0420 | 0.0440 | 0.0350 | 0.0430 | 0.0320 | 0.0250 | 1.8878 | 0.5503 | 0.1742 |
| P | 0.0300 | 0.0250 | 0.0300 | 0.0280 | 0.0180 | 0.0170 | 1.6099 | 0.5232 | 0.1700 |
| BCa | 0.0240 | 0.0230 | 0.0270 | 0.0620 | 0.0220 | 0.0190 | 1.4846 | 0.5150 | 0.1698 |
| DB | 0.0260 | 0.0220 | 0.0360 | 0.0240 | 0.0200 | 0.0240 | 1.6922 | 0.5299 | 0.1689 |

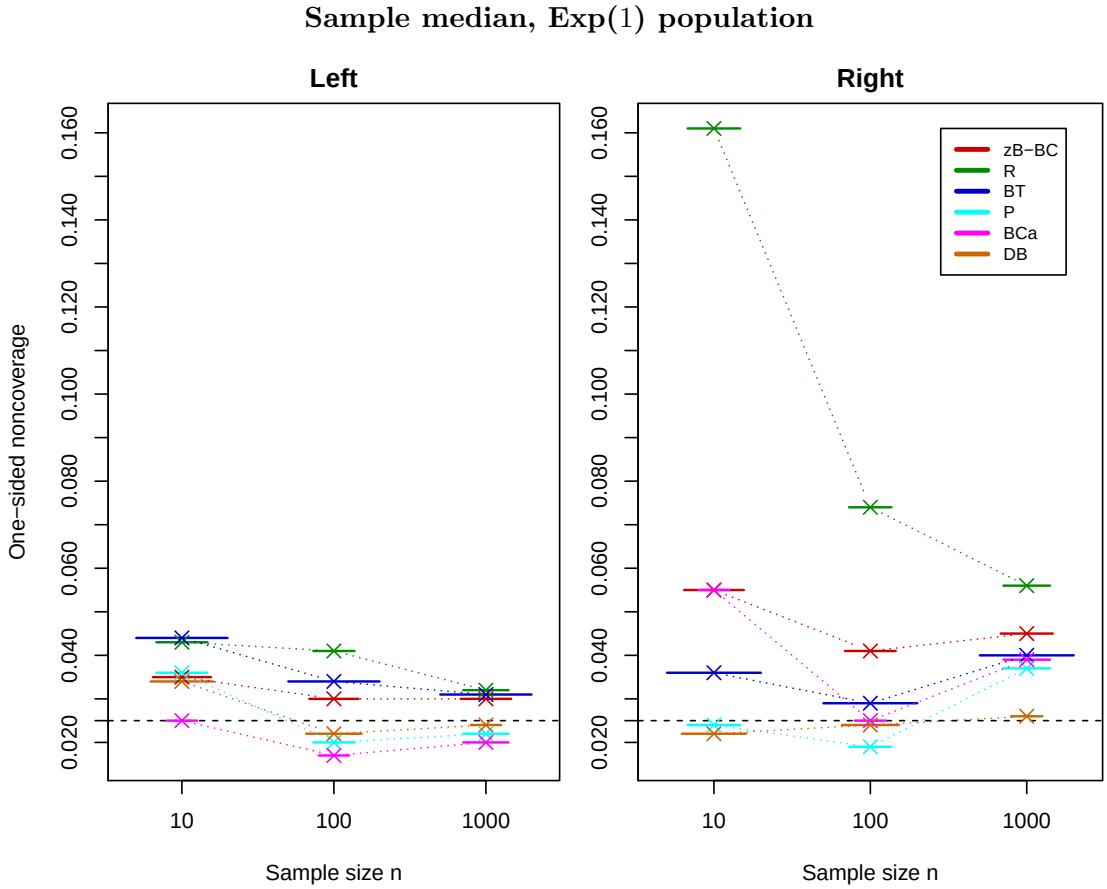


Figure 4.10: Sample median of exponential population: plot of noncoverage to the left (Left) and to the right (Right) of some bootstrap intervals.

Table 4.11: Sample median of exponential population: values of noncoverage to the left, noncoverage to the right and median length of some bootstrap intervals.

| | Sample median, Exp(1) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Noncoverage left | | | Noncoverage right | | | Median length | | |
| | 10 | 100 | 1000 | 10 | 100 | 1000 | 10 | 100 | 1000 |
| zB-BC | 0.0350 | 0.0300 | 0.0300 | 0.0550 | 0.0410 | 0.0450 | 1.2544 | 0.3910 | 0.1233 |
| R | 0.0430 | 0.0410 | 0.0320 | 0.1610 | 0.0740 | 0.0560 | 1.2110 | 0.3864 | 0.1230 |
| BT | 0.0440 | 0.0340 | 0.0310 | 0.0360 | 0.0290 | 0.0400 | 1.4604 | 0.4147 | 0.1257 |
| P | 0.0360 | 0.0200 | 0.0220 | 0.0240 | 0.0190 | 0.0370 | 1.2110 | 0.3864 | 0.1230 |
| BCa | 0.0250 | 0.0170 | 0.0200 | 0.0550 | 0.0250 | 0.0390 | 1.0824 | 0.3802 | 0.1230 |
| DB | 0.0340 | 0.0220 | 0.0240 | 0.0220 | 0.0240 | 0.0260 | 1.2818 | 0.3944 | 0.1222 |

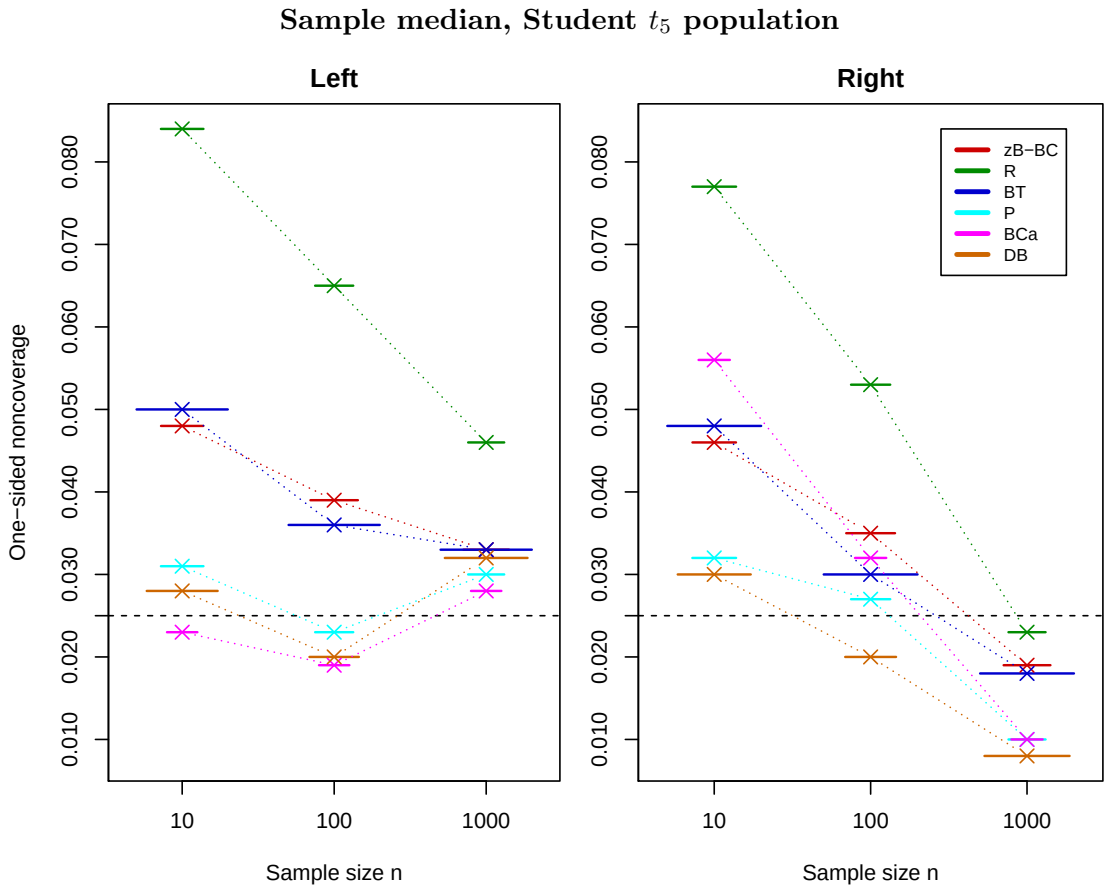


Figure 4.11: Sample median of Student $t_5$ population: plot of noncoverage to the left (Left) and to the right (Right) of some bootstrap intervals.

Table 4.12: Sample median of Student $t_5$ population: values of noncoverage to the left, noncoverage to the right and median length of some bootstrap intervals.

| | Sample median, Student $t_5$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Noncoverage left | | | Noncoverage right | | | Median length | | |
| | 10 | 100 | 1000 | 10 | 100 | 1000 | 10 | 100 | 1000 |
| zB-BC | 0.0480 | 0.0390 | 0.0330 | 0.0460 | 0.0350 | 0.0190 | 1.6261 | 0.5140 | 0.1630 |
| R | 0.0840 | 0.0650 | 0.0460 | 0.0770 | 0.0530 | 0.0230 | 1.6270 | 0.5076 | 0.1624 |
| BT | 0.0500 | 0.0360 | 0.0330 | 0.0480 | 0.0300 | 0.0180 | 1.8257 | 0.5439 | 0.1661 |
| P | 0.0310 | 0.0230 | 0.0300 | 0.0320 | 0.0270 | 0.0100 | 1.6270 | 0.5076 | 0.1624 |
| BCa | 0.0230 | 0.0190 | 0.0280 | 0.0560 | 0.0320 | 0.0100 | 1.5774 | 0.5023 | 0.1620 |
| DB | 0.0280 | 0.0200 | 0.0320 | 0.0300 | 0.0200 | 0.0080 | 1.7430 | 0.5153 | 0.1656 |

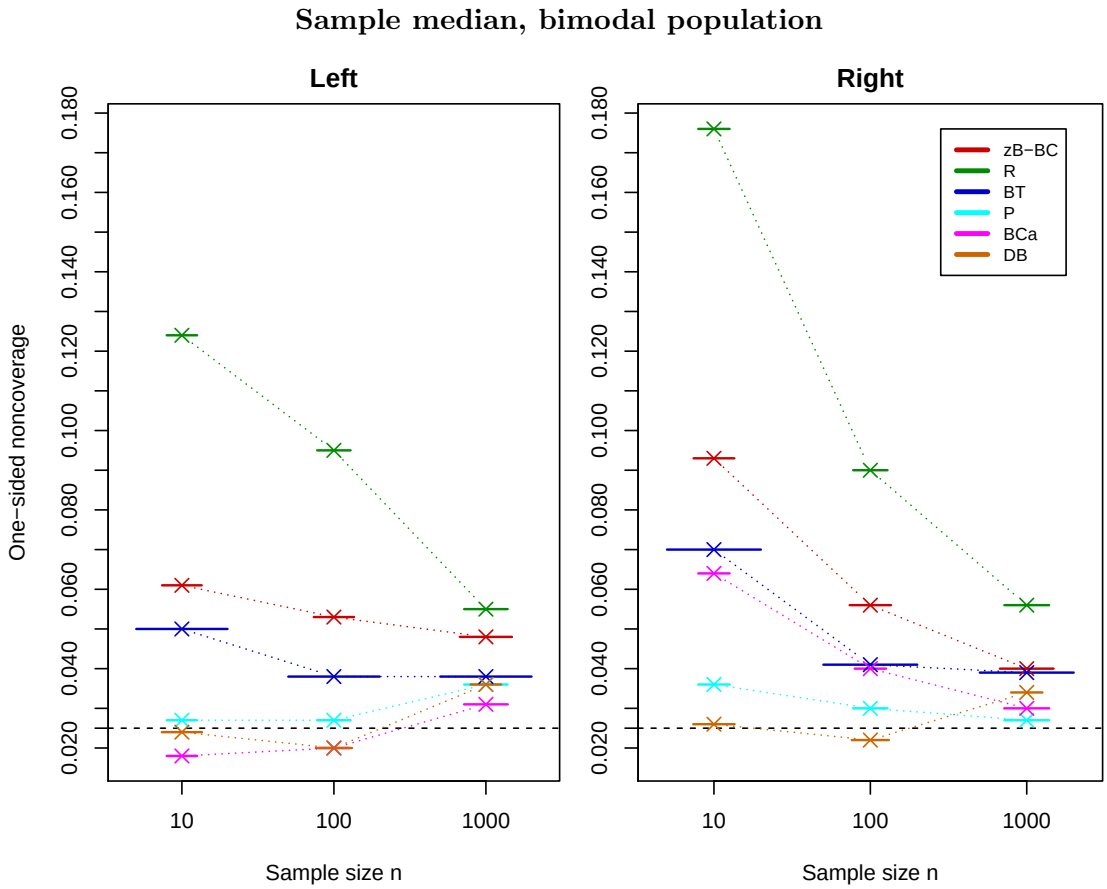


Figure 4.12: Sample median of bimodal population: plot of noncoverage to the left (Left) and to the right (Right) of some bootstrap intervals.

Table 4.13: Sample median of bimodal population: values of noncoverage to the left, noncoverage to the right and median length of some bootstrap intervals.

| | Sample median, bimodal | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Noncoverage left | | | Noncoverage right | | | Median length | | |
| | 10 | 100 | 1000 | 10 | 100 | 1000 | 10 | 100 | 1000 |
| zB-BC | 0.0610 | 0.0530 | 0.0480 | 0.0930 | 0.0560 | 0.0400 | 3.1257 | 1.3576 | 0.4661 |
| R | 0.1240 | 0.0950 | 0.0550 | 0.1760 | 0.0900 | 0.0560 | 2.9726 | 1.3303 | 0.4619 |
| BT | 0.0500 | 0.0380 | 0.0380 | 0.0700 | 0.0410 | 0.0390 | 3.9890 | 1.5622 | 0.4848 |
| P | 0.0270 | 0.0270 | 0.0360 | 0.0360 | 0.0300 | 0.0270 | 2.9726 | 1.3303 | 0.4619 |
| BCa | 0.0180 | 0.0200 | 0.0310 | 0.0640 | 0.0400 | 0.0300 | 2.9720 | 1.3184 | 0.4620 |
| DB | 0.0240 | 0.0200 | 0.0360 | 0.0260 | 0.0220 | 0.0340 | 3.1352 | 1.3415 | 0.4557 |

## 4.5 Coverage of Bootstrap Confidence Intervals for the Sample Correlation Coefficient

Correlation is a measure of the statistical association of two random variables. Arguably one of the most widely used estimators for the correlation of two random variables is the Pearson sample correlation coefficient. Let $X_1, \ldots, X_n$ be iid copies of $X \sim F_X$, with $\mathbb{E}(X) = \mu_X$ and $\text{Var}(X) = \sigma_X^2$, and let $Y_1, \ldots, Y_n$ be iid copies of $Y \sim F_Y$, with $\mathbb{E}(Y) = \mu_Y$ and $\text{Var}(Y) = \sigma_Y^2$. The Pearson correlation coefficient of $X$ and $Y$ is defined as

$$\rho = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{\mathbb{E}\left[(X - \mu_X)(Y - \mu_Y)\right]}{\sigma_X \sigma_Y} \tag{4.7}$$

We can substitute estimates of the variance of $X$ and $Y$ and their covariance based on $\{(X_1, Y_1), \ldots, (X_n, Y_n)\}$ to arrive at the sample Pearson correlation coefficient:

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})}} \tag{4.8}$$

Analogous to the sample mean and sample median, the sample Pearson correlation coefficient also admits a CLT expansion. Under some regularity conditions, it can be shown that $r$ asymptotically follows a normal distribution:

$$\sqrt{n}(r - \rho) \xrightarrow{D} \mathcal{N}(0,\ a^T \Sigma a) \tag{4.9}$$

where $a = \left(\frac{-\rho}{2\sigma_X^2}, \frac{-\rho}{2\sigma_Y^2}, \frac{1}{\sigma_X \sigma_Y}\right)^T$ and

$$\Sigma = \begin{pmatrix} \text{Cov}(X^2, X^2) & \text{Cov}(X^2, Y^2) & \text{Cov}(X^2, XY) \\ \text{Cov}(Y^2, X^2) & \text{Cov}(Y^2, Y^2) & \text{Cov}(Y^2, XY) \\ \text{Cov}(XY, X^2) & \text{Cov}(XY, Y^2) & \text{Cov}(XY, XY) \end{pmatrix} \tag{4.10}$$

Moreover, if it is assumed that the pairs $\{(X_1, Y_1), \ldots, (X_n, Y_n)\}$ come from a bivariate normal distribution $F_{XY}$, it can be shown that Fisher's $z$-transformation is a variance-stabilising transformation for (4.9). In that case, we have that

$$\sqrt{n}\left(\tanh^{-1}(r) - \tanh^{-1}(\rho)\right) \xrightarrow{D} \mathcal{N}(0,\ 1) \tag{4.11}$$

Beside these asymptotic results, there exists an exact formula for the PDF of $r$ when it is assumed that pairs $\{(X_1, Y_1), \ldots, (X_n, Y_n)\}$ come from a bivariate normal. There are different formulations for it, but we show next the formulation used by Efron and Hastie (2016):

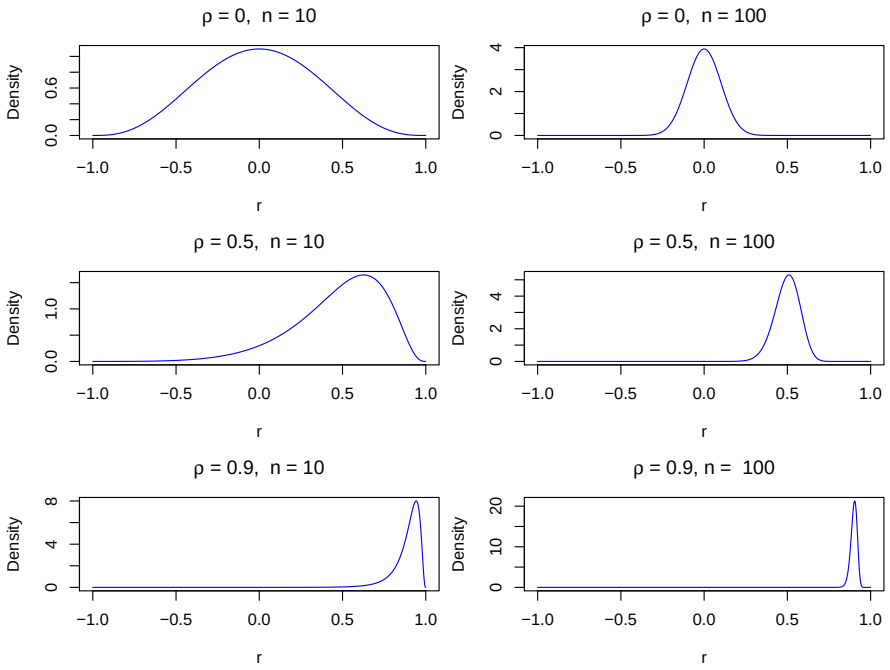


Figure 4.13: PDF of sample Pearson correlation coefficient $r$ for different values of $\rho$ and $n$.

$$f_\rho(r) = \frac{(n-2)(1-\rho^2)^{(n-1)/2}\,(1-r^2)^{(n-4)/2}}{\pi} \int_0^\infty \frac{dw}{\Big(\cosh(w) - r\rho\Big)^{n-1}} \tag{4.12}$$

Note that the true density of the estimator $r$ depends on the unknown value of $\rho$. When $\rho = 0$, the distribution of $r$ is symmetric, unimodal and with mean $\rho$. As $|\rho| \to 1$, the shape of the distribution of $r$ becomes increasingly skewed (positively for $\rho \to -1$, negatively for $\rho \to 1$), the variance of $r$ decreases, and when $|\rho| = 1$, $f_\rho(r)$ degenerates in a point mass probability distribution at $r = 1$ ($\rho = 1$) or $r = -1$ ($\rho = -1$). On top of the previous behaviour, the shape of $f_\rho(r)$ becomes increasingly normal as $n \to \infty$.

Now that we know enough about the behaviour of $r$, we want to investigate the performance of bootstrap confidence intervals by using noncoverage plots as defined in Section 4.1. Figures 4.14 to 4.17 and tables 4.14 to 4.17 show the results from the simulations described in Section 4.2 for the sample Pearson correlation coefficient and its transformed version, when the true correlation coefficient is $\rho = 0.5$. Figures 4.18 to 4.21 and tables 4.18 to 4.21 show the same results but for $\rho = 0.9$.

When Fisher's $z$-transformation has been applied, the size of median lengths in the tables refers to bootstrap intervals in the original scale, not in the transformed scale.

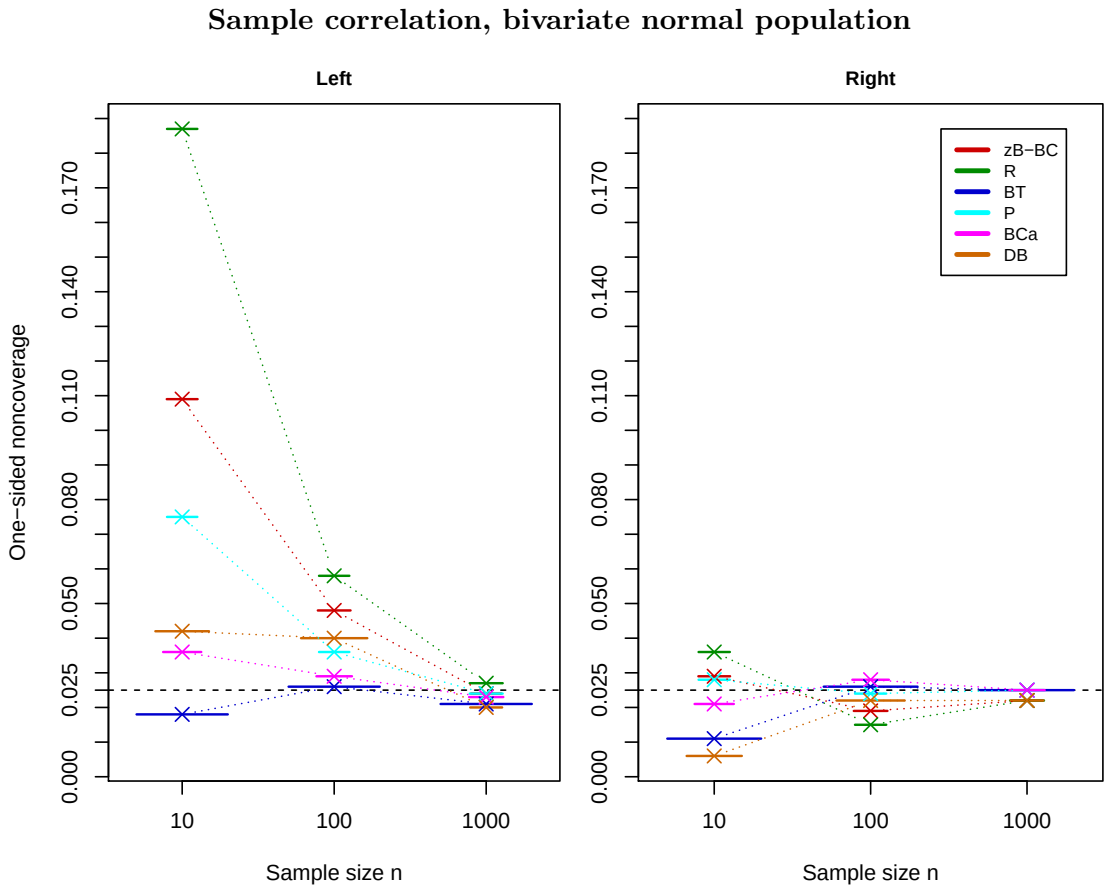


Figure 4.14: Sample correlation of bivariate normal population ($\rho = 0.5$): plot of noncoverage to the left (Left) and to the right (Right) of some bootstrap intervals.

Table 4.14: Sample correlation of bivariate normal population: values of noncoverage to the left, to the right and median length of some bootstrap intervals.

| | Sample correlation ($\rho = 0.5$), bivariate normal | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Noncoverage left | | | Noncoverage right | | | Median length | | |
| | 10 | 100 | 1000 | 10 | 100 | 1000 | 10 | 100 | 1000 |
| zB-BC | 0.1090 | 0.0480 | 0.0240 | 0.0290 | 0.0190 | 0.0220 | 0.9849 | 0.2915 | 0.0932 |
| R | 0.1870 | 0.0580 | 0.0270 | 0.0360 | 0.0150 | 0.0220 | 0.9796 | 0.2909 | 0.0932 |
| BT | 0.0180 | 0.0260 | 0.0210 | 0.0110 | 0.0260 | 0.0250 | 1.6715 | 0.3075 | 0.0938 |
| P | 0.0750 | 0.0360 | 0.0240 | 0.0280 | 0.0240 | 0.0250 | 0.9796 | 0.2909 | 0.0932 |
| BCa | 0.0360 | 0.0290 | 0.0230 | 0.0210 | 0.0280 | 0.0250 | 1.0680 | 0.2923 | 0.0932 |
| DB | 0.0420 | 0.0400 | 0.0200 | 0.0060 | 0.0220 | 0.0220 | 1.2440 | 0.3007 | 0.0932 |

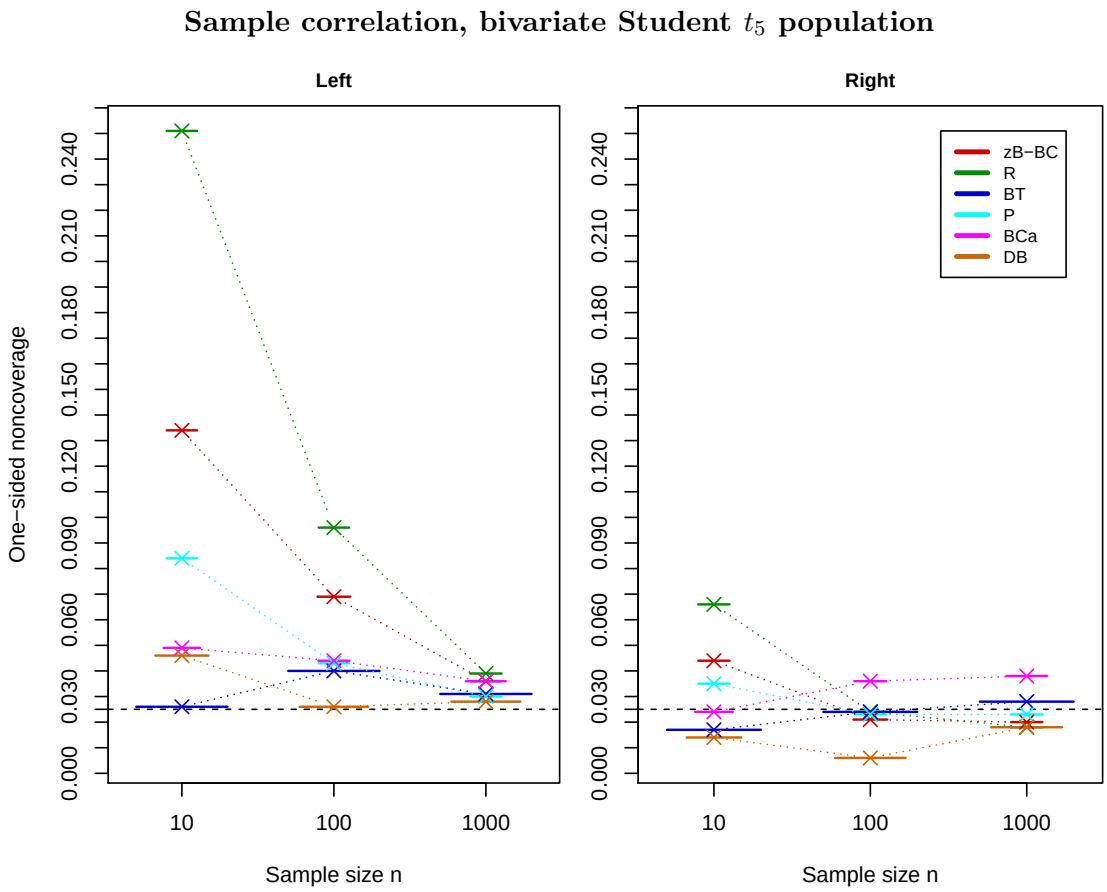


Figure 4.15: Sample correlation of bivariate Student $t_5$ population ($\rho = 0.5$): plot of noncoverage to the left (Left) and to the right (Right) of some bootstrap intervals.

Table 4.15: Sample correlation of bivariate Student $t_5$ population: values of noncoverage to the left, to the right and median length of some bootstrap intervals.

| | Sample correlation ($\rho = 0.5$), bivariate Student $t_5$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Noncoverage left | | | Noncoverage right | | | Median length | | |
| | 10 | 100 | 1000 | 10 | 100 | 1000 | 10 | 100 | 1000 |
| zB-BC | 0.1340 | 0.0690 | 0.0360 | 0.0440 | 0.0210 | 0.0200 | 1.0283 | 0.3567 | 0.1308 |
| R | 0.2510 | 0.0960 | 0.0390 | 0.0660 | 0.0240 | 0.0180 | 1.0279 | 0.3546 | 0.1309 |
| BT | 0.0260 | 0.0400 | 0.0310 | 0.0170 | 0.0240 | 0.0280 | 1.7529 | 0.4052 | 0.1350 |
| P | 0.0840 | 0.0430 | 0.0300 | 0.0350 | 0.0230 | 0.0230 | 1.0279 | 0.3546 | 0.1309 |
| BCa | 0.0490 | 0.0440 | 0.0360 | 0.0240 | 0.0360 | 0.0380 | 1.1023 | 0.3552 | 0.1315 |
| DB | 0.0460 | 0.0260 | 0.0280 | 0.0140 | 0.0060 | 0.0180 | 1.2996 | 0.3863 | 0.1335 |

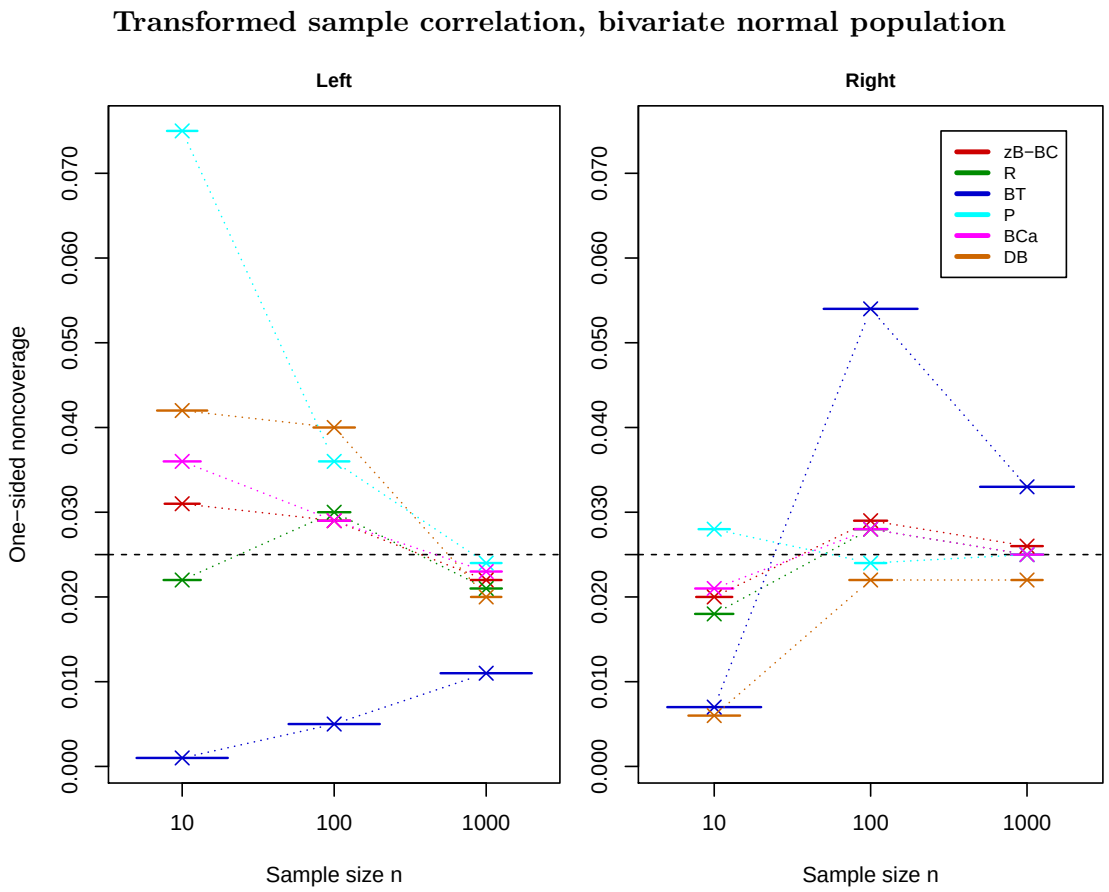


Figure 4.16: Transformed sample correlation of bivariate normal population ($\rho = 0.5$): plot of noncoverage to the left (Left) and to the right (Right) of some bootstrap intervals.

Table 4.16: Transformed sample correlation of bivariate normal population: values of noncoverage to the left, to the right and median length of some bootstrap intervals.

| | Transformed sample correlation ($\rho = 0.5$), bivariate normal | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Noncoverage left | | | Noncoverage right | | | Median length | | |
| | 10 | 100 | 1000 | 10 | 100 | 1000 | 10 | 100 | 1000 |
| zB-BC | 0.0310 | 0.0290 | 0.0220 | 0.0200 | 0.0290 | 0.0260 | 1.0424 | 0.2925 | 0.0932 |
| R | 0.0220 | 0.0300 | 0.0210 | 0.0180 | 0.0280 | 0.0250 | 1.0704 | 0.2927 | 0.0932 |
| BT | 0.0010 | 0.0050 | 0.0110 | 0.0070 | 0.0540 | 0.0330 | 1.7928 | 0.3451 | 0.0950 |
| P | 0.0750 | 0.0360 | 0.0240 | 0.0280 | 0.0240 | 0.0250 | 0.9796 | 0.2909 | 0.0932 |
| BCa | 0.0360 | 0.0290 | 0.0230 | 0.0210 | 0.0280 | 0.0250 | 1.0680 | 0.2923 | 0.0932 |
| DB | 0.0420 | 0.0400 | 0.0200 | 0.0060 | 0.0220 | 0.0220 | 1.2440 | 0.3007 | 0.0932 |

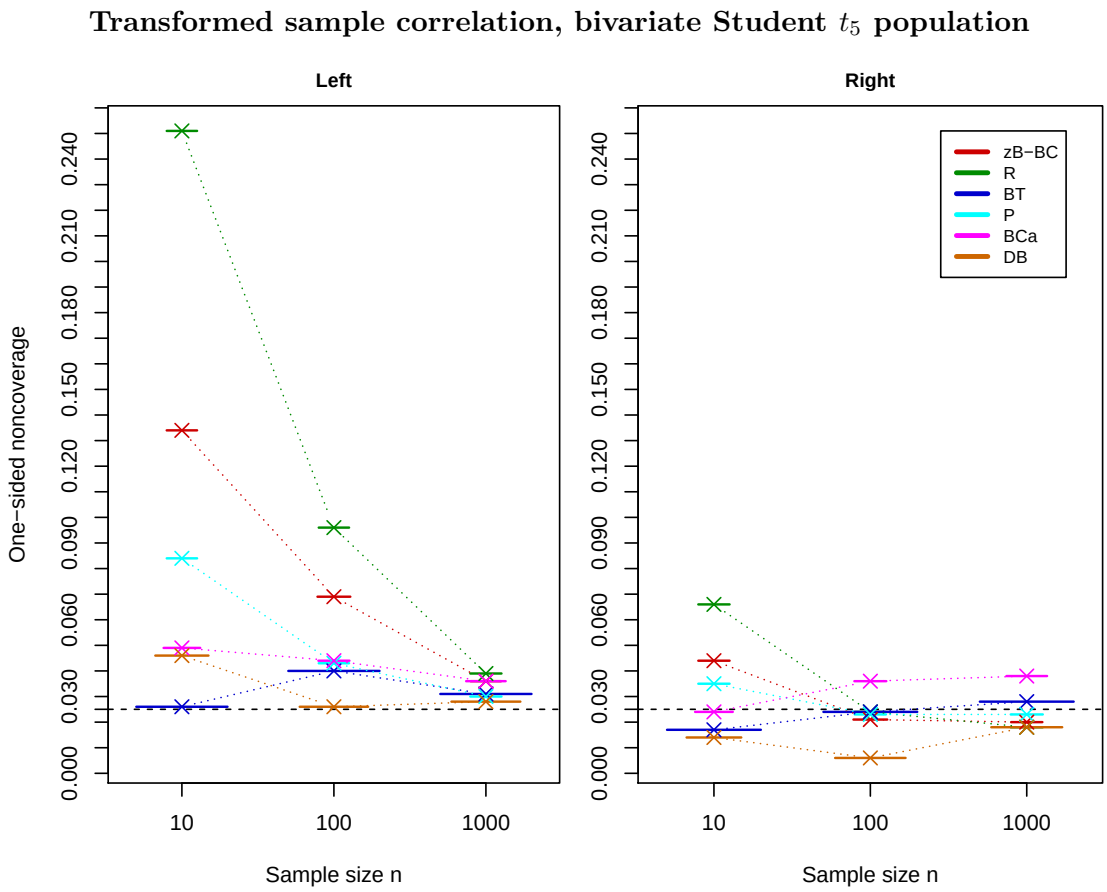


Figure 4.17: Transformed sample correlation of bivariate Student $t_5$ population ($\rho = 0.5$): plot of noncoverage to the left (Left) and to the right (Right) of some bootstrap intervals.

Table 4.17: Transformed sample correlation of bivariate Student $t_5$ population: values of noncoverage to the left, to the right and median length of some bootstrap intervals.

| | Transformed sample correlation ($\rho = 0.5$), bivariate Student $t_5$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Noncoverage left | | | Noncoverage right | | | Median length | | |
| | 10 | 100 | 1000 | 10 | 100 | 1000 | 10 | 100 | 1000 |
| zB-BC | 0.0460 | 0.0450 | 0.0270 | 0.0290 | 0.0310 | 0.0280 | 1.0668 | 0.3566 | 0.1309 |
| R | 0.0390 | 0.0450 | 0.0260 | 0.0420 | 0.0420 | 0.0250 | 1.0826 | 0.3560 | 0.1310 |
| BT | 0.0000 | 0.0040 | 0.0110 | 0.0080 | 0.0690 | 0.0550 | 1.8363 | 0.4728 | 0.1381 |
| P | 0.0840 | 0.0430 | 0.0300 | 0.0350 | 0.0230 | 0.0230 | 1.0279 | 0.3546 | 0.1309 |
| BCa | 0.0490 | 0.0440 | 0.0360 | 0.0240 | 0.0360 | 0.0380 | 1.1023 | 0.3552 | 0.1315 |
| DB | 0.0460 | 0.0260 | 0.0280 | 0.0140 | 0.0060 | 0.0180 | 1.2996 | 0.3863 | 0.1335 |

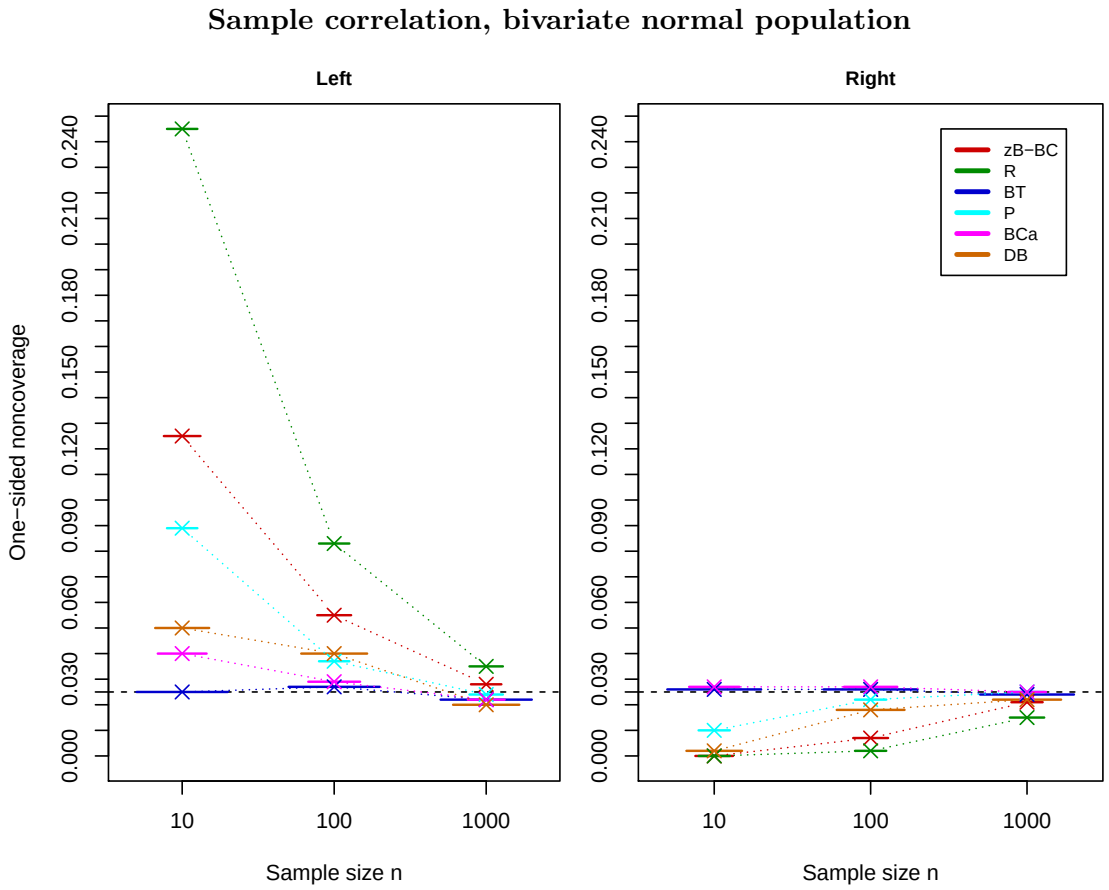


Figure 4.18: Sample correlation of bivariate normal population ($\rho = 0.9$): plot of noncoverage to the left (Left) and to the right (Right) of some bootstrap intervals.

Table 4.18: Sample correlation of bivariate normal population: values of noncoverage to the left, to the right and median length of some bootstrap intervals.

| | Sample correlation ($\rho = 0.9$), bivariate normal | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Noncoverage left | | | Noncoverage right | | | Median length | | |
| | 10 | 100 | 1000 | 10 | 100 | 1000 | 10 | 100 | 1000 |
| zB-BC | 0.1090 | 0.0480 | 0.0240 | 0.0290 | 0.0190 | 0.0220 | 0.9849 | 0.2915 | 0.0932 |
| R | 0.1870 | 0.0580 | 0.0270 | 0.0360 | 0.0150 | 0.0220 | 0.9796 | 0.2909 | 0.0932 |
| BT | 0.0180 | 0.0260 | 0.0210 | 0.0110 | 0.0260 | 0.0250 | 1.6715 | 0.3075 | 0.0938 |
| P | 0.0750 | 0.0360 | 0.0240 | 0.0280 | 0.0240 | 0.0250 | 0.9796 | 0.2909 | 0.0932 |
| BCa | 0.0360 | 0.0290 | 0.0230 | 0.0210 | 0.0280 | 0.0250 | 1.0680 | 0.2923 | 0.0932 |
| DB | 0.0420 | 0.0400 | 0.0200 | 0.0060 | 0.0220 | 0.0220 | 1.2440 | 0.3007 | 0.0932 |

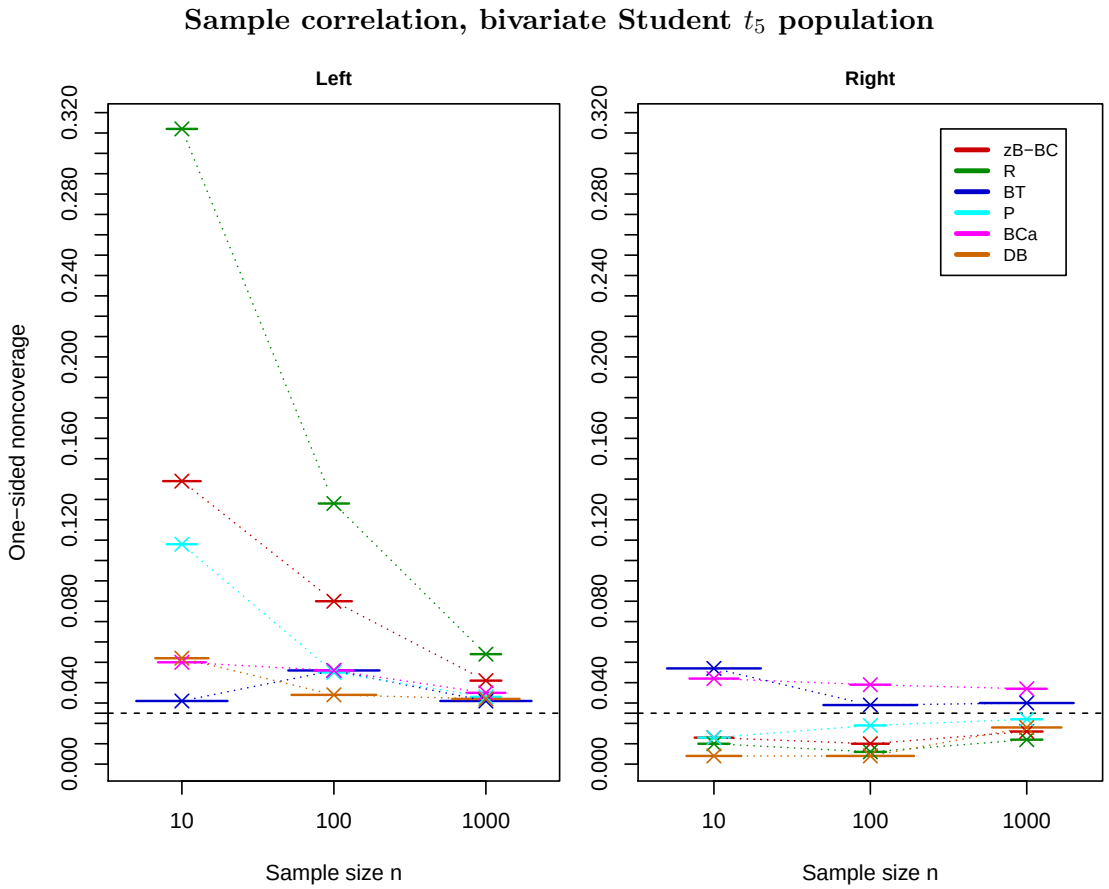


Figure 4.19: Sample correlation of bivariate Student $t_5$ population ($\rho = 0.9$): plot of noncoverage to the left (Left) and to the right (Right) of some bootstrap intervals.

Table 4.19: Sample correlation of bivariate Student $t_5$ population: values of noncoverage to the left, to the right and median length of some bootstrap intervals.

| | Sample correlation ($\rho = 0.9$), bivariate Student $t_5$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Noncoverage left | | | Noncoverage right | | | Median length | | |
| | 10 | 100 | 1000 | 10 | 100 | 1000 | 10 | 100 | 1000 |
| zB-BC | 0.1340 | 0.0690 | 0.0360 | 0.0440 | 0.0210 | 0.0200 | 1.0283 | 0.3567 | 0.1308 |
| R | 0.2510 | 0.0960 | 0.0390 | 0.0660 | 0.0240 | 0.0180 | 1.0279 | 0.3546 | 0.1309 |
| BT | 0.0260 | 0.0400 | 0.0310 | 0.0170 | 0.0240 | 0.0280 | 1.7529 | 0.4052 | 0.1350 |
| P | 0.0840 | 0.0430 | 0.0300 | 0.0350 | 0.0230 | 0.0230 | 1.0279 | 0.3546 | 0.1309 |
| BCa | 0.0490 | 0.0440 | 0.0360 | 0.0240 | 0.0360 | 0.0380 | 1.1023 | 0.3552 | 0.1315 |
| DB | 0.0460 | 0.0260 | 0.0280 | 0.0140 | 0.0060 | 0.0180 | 1.2996 | 0.3863 | 0.1335 |

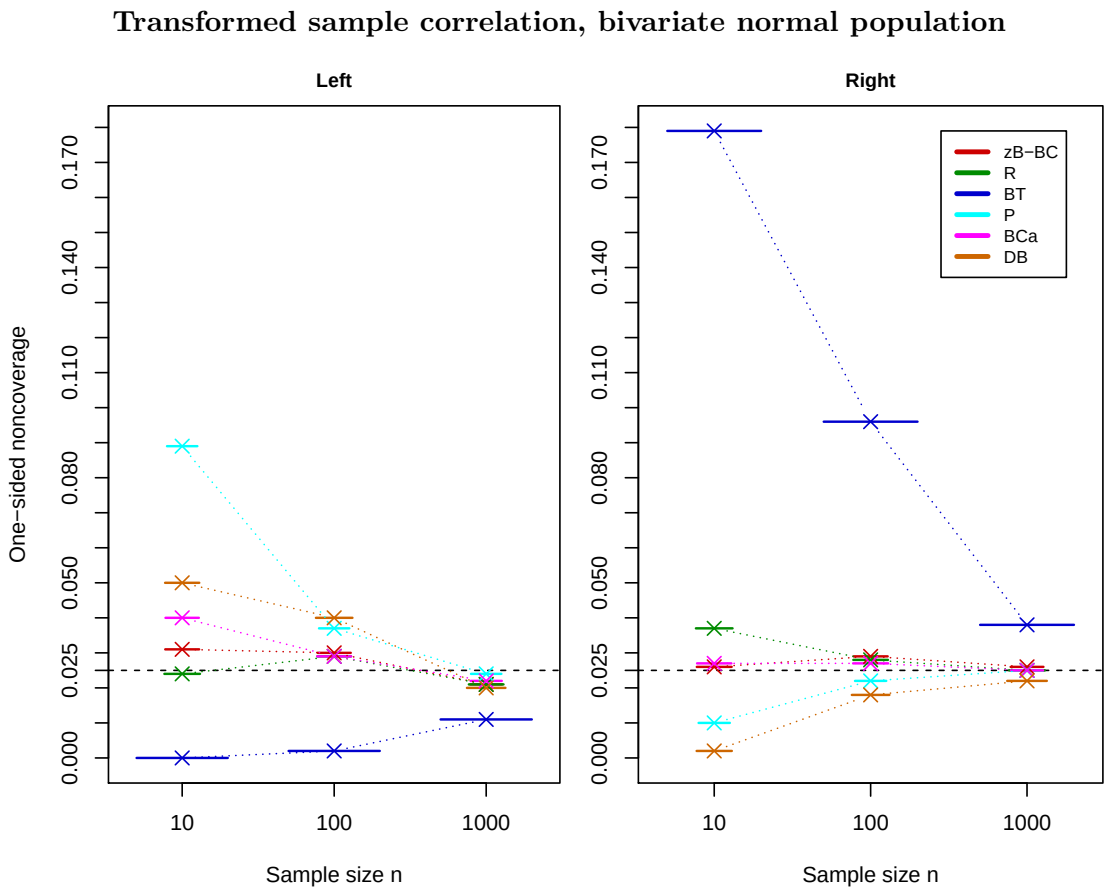


Figure 4.20: Transformed sample correlation of bivariate normal population ($\rho = 0.9$): plot of noncoverage to the left (Left) and to the right (Right) of some bootstrap intervals.

Table 4.20: Transformed sample correlation of bivariate normal population: values of noncoverage to the left, to the right and median length of some bootstrap intervals.

| | Transformed sample correlation ($\rho = 0.9$), bivariate normal | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Noncoverage left | | | Noncoverage right | | | Median length | | |
| | 10 | 100 | 1000 | 10 | 100 | 1000 | 10 | 100 | 1000 |
| zB-BC | 0.0310 | 0.0290 | 0.0220 | 0.0200 | 0.0290 | 0.0260 | 1.0424 | 0.2925 | 0.0932 |
| R | 0.0220 | 0.0300 | 0.0210 | 0.0180 | 0.0280 | 0.0250 | 1.0704 | 0.2927 | 0.0932 |
| BT | 0.0010 | 0.0050 | 0.0110 | 0.0070 | 0.0540 | 0.0330 | 1.7928 | 0.3451 | 0.0950 |
| P | 0.0750 | 0.0360 | 0.0240 | 0.0280 | 0.0240 | 0.0250 | 0.9796 | 0.2909 | 0.0932 |
| BCa | 0.0360 | 0.0290 | 0.0230 | 0.0210 | 0.0280 | 0.0250 | 1.0680 | 0.2923 | 0.0932 |
| DB | 0.0420 | 0.0400 | 0.0200 | 0.0060 | 0.0220 | 0.0220 | 1.2440 | 0.3007 | 0.0932 |

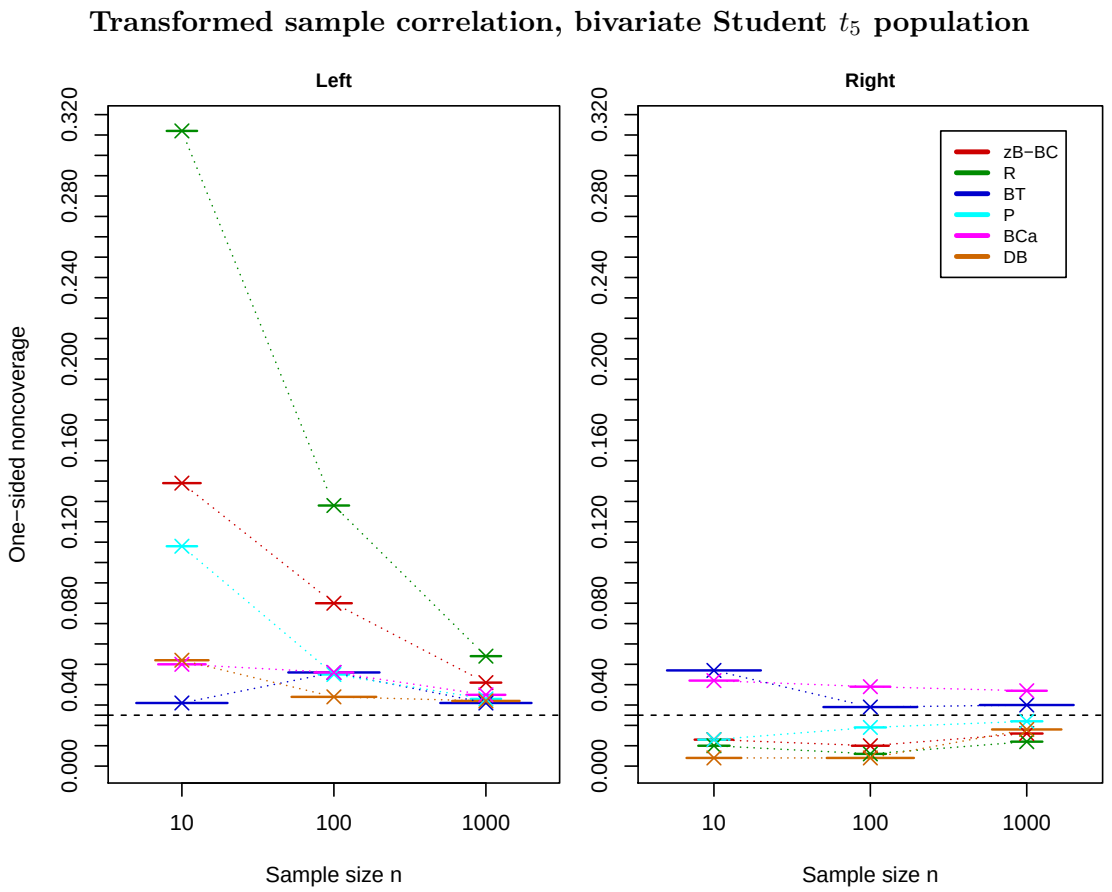


Figure 4.21: Transformed sample correlation of bivariate Student $t_5$ population ($\rho = 0.9$): plot of noncoverage to the left (Left) and to the right (Right) of some bootstrap intervals.

Table 4.21: Transformed sample correlation of bivariate Student $t_5$ population: values of noncoverage to the left, to the right and median length of some bootstrap intervals.

| | Transformed sample correlation ($\rho = 0.9$), bivariate Student $t_5$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Noncoverage left | | | Noncoverage right | | | Median length | | |
| | 10 | 100 | 1000 | 10 | 100 | 1000 | 10 | 100 | 1000 |
| zB-BC | 0.0460 | 0.0450 | 0.0270 | 0.0290 | 0.0310 | 0.0280 | 1.0668 | 0.3566 | 0.1309 |
| R | 0.0390 | 0.0450 | 0.0260 | 0.0420 | 0.0420 | 0.0250 | 1.0826 | 0.3560 | 0.1310 |
| BT | 0.0000 | 0.0040 | 0.0110 | 0.0080 | 0.0690 | 0.0550 | 1.8363 | 0.4728 | 0.1381 |
| P | 0.0840 | 0.0430 | 0.0300 | 0.0350 | 0.0230 | 0.0230 | 1.0279 | 0.3546 | 0.1309 |
| BCa | 0.0490 | 0.0440 | 0.0360 | 0.0240 | 0.0360 | 0.0380 | 1.1023 | 0.3552 | 0.1315 |
| DB | 0.0460 | 0.0260 | 0.0280 | 0.0140 | 0.0060 | 0.0180 | 1.2996 | 0.3863 | 0.1335 |

## 4.6 Ranking of Bootstrap Confidence Intervals

All the figures and tables in the previous sections give a lot of detail about the results of the simulations we ran. But one of the goals of these simulations is to know how the different bootstrap intervals behave in different scenarios. Hence, we decided to report a table with the performance ranking of every bootstrap confidence interval in each scenario.

For a specific scenario (one estimator + one population) we calculated a ranking for every $n$ for both noncoverage to the left and to the right. This yields a total of 6 rankings. We then summed them up to come up with a total sum of rankings, and computed the ranking of this sum. This last ranking is the one showed in Table 4.22. Ties are represented with an average of the corresponding positions if there was no tie. That is, if two intervals are tied at rank 3, if there was no tie they would be in ranks 3 and 4, so the average is 3.5. The best is represented with a 1 and the worst with a 6.

All rankings in Table 4.22 can be obtained from the tables with the values of noncoverage to the left and to the right if one follows the logic explained above.

Table 4.22: Performance ranking of some bootstrap confidence intervals that were tested in different scenarios. 1 is best, 6 is worst.

| Estimator | Population | zB-BC | R | BT | P | BCa | DB |
|---|---|---|---|---|---|---|---|
| Sample mean | Normal | 4 | 3 | 1 | 5 | 6 | 2 |
| | Lognormal | 5 | 6 | 1 | 3 | 2 | 4 |
| | Exponential | 5 | 6 | 2 | 3 | 1 | 4 |
| | Student $t$ | 3 | 1 | 2 | 4 | 6 | 5 |
| | Bimodal | 3.5 | 6 | 2 | 3.5 | 1 | 5 |
| Sample median | Normal | 4 | 6 | 5 | 3 | 1 | 2 |
| | Lognormal | 4 | 6 | 5 | 3 | 1 | 2 |
| | Exponential | 4 | 6 | 5 | 2 | 3 | 1 |
| | Student $t$ | 5 | 6 | 4 | 1 | 3 | 2 |
| | Bimodal | 5 | 6 | 4 | 1 | 3 | 2 |
| Sample correlation ($\rho = 0.5$) | Bivariate normal | 4 | 6 | 2.5 | 1 | 2.5 | 5 |
| | Bivariate Student $t$ | 5 | 6 | 1 | 2 | 4 | 3 |
| Transformed sample correlation ($\rho = 0.5$) | Bivariate normal | 4 | 3 | 6 | 2 | 1 | 5 |
| | Bivariate Student $t$ | 1 | 2 | 6 | 3 | 5 | 4 |
| Sample correlation ($\rho = 0.9$) | Bivariate normal | 5 | 6 | 1 | 3 | 2 | 4 |
| | Bivariate Student $t$ | 5 | 6 | 2 | 1 | 4 | 3 |
| Transformed sample correlation ($\rho = 0.9$) | Bivariate normal | 3 | 2 | 6 | 4 | 1 | 5 |
| | Bivariate Student $t$ | 2 | 1 | 6 | 3 | 5 | 4 |

# Chapter 5

# Summary and Conclusions

This thesis had mainly two goals: to present the fundamental bootstrap principles that make it work and to analyse how bootstrap confidence intervals perform in practice in common scenarios. The following list is a summary with the main things that have been presented:

- An introduction to the bootstrap at a level of an advance undergraduate or graduate student of statistics.
- A review of bootstrap consistency. Cases for which the bootstrap is known to always work have been presented. In a nutshell, it works for functions of the sample mean, sample quantiles and, more generally, estimators based on Hadamard-differentiable functionals.
- A review of bootstrap accuracy. Rates of convergence for different estimators have been presented, namely for the sample mean, for the $t$ statistic and for sample quantiles.
- A list of cases in which the bootstrap does not work. These are usually cases in which the regularity conditions needed for the CLT to work do not hold. A good number of examples are shown. Besides, a technique to theoretically solve bootstrap inconsistency is introduced, but the practical implementation remains an open problem.
- First and second order accurate bootstrap confidence intervals. Most of them are variations of the "purest" type of bootstrap confidence interval: the percentile interval.
- A technique called double bootstrap to improve bootstrap intervals. Can be used for bias reduction or for coverage correction.
- The two most widely used R packages to implement general-purpose bootstrap problems, namely *boot* and *bootstrap*.

- Results from coverage analyses of bootstrap intervals for the sample mean, sample median and sample correlation coefficient, displayed in noncoverage plots.
- A table with the ranking of bootstrap confidence intervals for the different scenarios simulated. These results must be taken with a pinch of salt, and the plots should be checked for more details.

We would also like to list some thoughts and conclusions learnt during the process of doing this thesis:

- If an estimator is computed based on some type of aggregation or summation of the sample data, it is likely that the bootstrap will work for it. An aggregation or summation implies that the estimator would be a function of the sample mean, and at some point the CLT would kick in and the estimator would follow an asymptotic normal distribution (if the required regularity conditions hold).
- Bootstrap confidence intervals are usually superior to normal approximations, even if both are first-order accurate. This is because bootstrap intervals do not assume normality and hence can pick up asymmetry from the sample data.
- Results of bootstrap $t$ intervals in different scenarios for the sample mean seem to agree with a remark from A.J.Canty, A.C. Davison and D.V. Hinkley in DiCiccio and Efron (1996). They point out that, for normal samples of size 20, the coverage of the bootstrap $t$ is usually better than the others by some margin. However, results in Chapter 4 suggest that the bootstrap $t$ achieves that at the expense of yielding wider confidence intervals.
- For estimators that take long to compute it might be wiser to stick to first-order bootstrap intervals, since the computational burden of a second-order bootstrap interval might be prohibitive. However, if the estimator happens to be smooth, we suggest using ABC intervals since they achieve second-order accuracy without the computational burden of other second-order bootstrap intervals.
- If we had to choose only one bootstrap confidence interval as the default to use in a general unknown situation, we would choose the BCa interval. The inherited advantage from the percentile interval of automatically finding normalising transformations for the estimator, plus its second-order accuracy justify our decision. When would we not choose it? As it can be seen in the results of Chapter 4 for the sample mean, when the population distribution is symmetric, other bootstrap intervals seem to work better. Besides, as pointed out in the previous bullet point, if the estimator takes long to compute, we would also choose a different bootstrap interval.

# Chapter 6

# Future Work

Due to time constraints, as it is natural to the nature of a master thesis, we did not have time to work on all the topics we would have liked to. We now list a number of interesting topics for future work, that we would have liked to have delved into:

- Learning how the m/n bootstrap technique cures the inconsistency of the bootstrap in practice. The technique proposes to draw bootstrap samples of size $m < n$, with $m$ fulfilling certain asymptotic conditions in relation to $n$. However, we did not get to learn how well it works in practice, and if there exist any publication on the matter that presents simulated results.
- Simulation of a robust estimator for the correlation coefficient, such as Spearman's or Kendall rank correlation coefficient.
- Run-time comparison of bootstrap confidence intervals, especially between the bootstrap $t$ and BCa intervals.
- Application of the bootstrap to dependent data. There is a variation of the bootstrap, called *block bootstrap*, which can be used when the data are correlated.
- Application of the bootstrap to (generalised) linear models. There are some variations of the bootstrap for these type of models, like the *residual bootstrap* (residuals are bootstrapped instead of the original sample) or the *wild bootstrap* (when heteroscedasticity is present).

# Bibliography


Andrews, D. W. (2000). Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. *Econometrica 68*(2), 399–405.

Athreya, K. et al. (1987). Bootstrap of the mean in the infinite variance case. *The Annals of Statistics 15*(2), 724–731.

Babu, G. J. (1984). Bootstrapping statistics with linear combinations of chi-squares as weak limit. *Sankhyā: The Indian Journal of Statistics, Series A*, 85–93.

Bickel, P. J., D. A. Freedman, et al. (1981). Some asymptotic theory for the bootstrap. *The annals of statistics 9*(6), 1196–1217.

Bühlmann, P. and M. Mächler (2016, October). Lecture notes in computational statistics. https://polybox.ethz.ch/index.php/s/q3NaRdR7K6Xrfkd/download. Seminar für Statistik, ETH Zürich.

Canty, A. and B. D. Ripley (2019). *boot: Bootstrap R (S-Plus) Functions*. R package version 1.3-23.

Corporation, M. and S. Weston (2019). *doParallel: Foreach Parallel Adaptor for the 'parallel' Package*. R package version 1.0.15.

DasGupta, A. (2008). *Asymptotic theory of statistics and probability*. Springer Science & Business Media.

Davison, A. C. and D. V. Hinkley (1997). *Bootstrap Methods and Their Applications*. Cambridge: Cambridge University Press. ISBN 0-521-57391-2.

De Angelis, D. and G. A. Young (1992). Smoothing the bootstrap. *International Statistical Review/Revue Internationale de Statistique*, 45–56.

DiCiccio, T. J. and B. Efron (1996). Bootstrap confidence intervals (with discussion). *Statistical science*, 189–212.

Efron, B. (1979, 01). Bootstrap methods: Another look at the jackknife. *Ann. Statist. 7*(1), 1–26.

Efron, B. (1981). Nonparametric standard errors and confidence intervals. *canadian Journal of Statistics 9*(2), 139–158.

Efron, B. (1982). Transformation theory: How normal is a family of distributions? *The Annals of Statistics*, 323–339.

Efron, B. (1987). Better bootstrap confidence intervals. *Journal of the American statistical Association 82*(397), 171–185.

Efron, B. (2003). Second thoughts on the bootstrap. *Statistical Science 18*(2), 135–140.

Efron, B. and T. Hastie (2016). *Computer Age Statistical Inference*, Volume 5. Cambridge University Press.

Efron, B. and R. J. Tibshirani (1993). *An Introduction to the Bootstrap*. CRC press.

Falk, M. and R.-D. Reiss (1989). Weak convergence of smoothed and nonsmoothed bootstrap quantile estimates. *The Annals of Probability*, 362–371.

Genz, A. and F. Bretz (2009). *Computation of Multivariate Normal and t Probabilities*. Lecture Notes in Statistics. Heidelberg: Springer-Verlag.

Genz, A., F. Bretz, T. Miwa, X. Mi, F. Leisch, F. Scheipl, and T. Hothorn (2019). *mvtnorm: Multivariate Normal and t Distributions*. R package version 1.0-11.

Gill, R. D., J. A. Wellner, and J. Præstgaard (1989). Non-and semi-parametric maximum likelihood estimators and the von mises method (part 1)[with discussion and reply]. *Scandinavian Journal of Statistics*, 97–128.

Giné, E. and J. Zinn (1989). Necessary conditions for the bootstrap of the mean. *The annals of statistics*, 684–691.

Hall, P. (1988). Rate of convergence in bootstrap approximations. *The Annals of Probability 16*(4), 1665–1684.

Hall, P. (1990). Asymptotic properties of the bootstrap for heavy-tailed distributions. *The Annals of Probability*, 1342–1360.

Hesterberg, T. C. (2015). What teachers should know about the bootstrap: Resampling in the undergraduate statistics curriculum. *The American Statistician 69*(4), 371–386.

Huang, J., P. Sen, and J. Shao (1996). Bootstrapping a sample quantile when the density has a jump. *Statistica Sinica*, 299–309.

J. Geyer, C. (2017, April). Stat 3701 lecture notes on the bootstrap. http://www.stat.umn.edu/geyer/3701/notes/bootstrap.pdf. University of Minnesota.

Johnson, N. J. (1978). Modified t tests and confidence intervals for asymmetrical populations. *Journal of the American Statistical Association 73*(363), 536–544.

Kleijnen et al. (1986). Testing the mean of an asymmetric population: Johnson's modified t test revisited. *Communications in Statistics-Simulation and Computation 15*(3), 715–732.

Knight, K. (1989). On the bootstrap of the sample mean in the infinite variance case. *The Annals of Statistics*, 1168–1175.

Liu, R. Y., K. Singh, and S.-H. Lo (1989). On a representation related to the bootstrap. *Sankhyā: The Indian Journal of Statistics, Series A*, 168–177.

Mächler, M. (2017). *nor1mix: Normal (1-d) Mixture Models (S3 Classes and Methods)*. R package version 1.2-3.

Martin, M. A. (1990a). On bootstrap iteration for coverage correction in confidence intervals. *Journal of the American Statistical Association 85*(412), 1105–1118.

Martin, M. A. (1990b). On the double bootstrap. In *Computing science and statistics*, pp. 73–78. Springer.

original, S., from StatLib, and by Rob Tibshirani. R port by Friedrich Leisch. (2019). *bootstrap: Functions for the Book "An Introduction to the Bootstrap"*. R package version 2019.6.

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Shao, J. and D. Tu (1995). *The jackknife and bootstrap*. Springer Science & Business Media.

Tibshirani, R. (1988). Variance stabilization and the bootstrap. *Biometrika 75*(3), 433–444.

Vaart, A. W. v. d. (1998). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

# Appendix A

# Tutorial: Bootstrap in R with the *boot* package

This tutorial tries to give a short introduction to the two main functions of the *boot* package: the *boot* function, to get bootstrap resamples, and the *boot.ci* function, to compute bootstrap confidence intervals. We use a two-sample $t$-test for the difference of means as a use case.

## A.1 Difference of Means

A common method to investigate the difference of two means is the two-sample $t$-test. The null hypothesis is that the two means are the same, and it is the status quo unless evidence from data says otherwise. There are slightly different variations of the $t$ statistic, depending on the assumptions on equal/unequal sample sizes and equal/unequal variances. We present the most general two-sample $t$-test, also known as *Welch's* $t$-test or simply *Welch*-test, which does not assume equal sample sizes nor equal variances.

Let $X_1, \ldots, X_n \overset{iid}{\sim} F_1$ and $Y_1, \ldots, Y_m \overset{iid}{\sim} F_2$, with $X_1 \in \mathbb{R}$ and $Y_1 \in \mathbb{R}$. Let $\mu_X = \mathbb{E}(X_1)$, $\mu_Y = \mathbb{E}(Y_1)$ be the means of $F_1$ and $F_2$ respectively, and let $\sigma_X^2 = \mathrm{Var}(X_1)$ and $\sigma_Y^2 = \mathrm{Var}(Y_1)$ be the variances of $F_1$ and $F_2$ respectively. Let $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$ and $s_X^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ be the sample mean and sample variance of $X_1, \ldots, X_n$ respectively, and let $\bar{Y}_m = m^{-1} \sum_{i=1}^m Y_i$ and $s_Y^2 = \frac{1}{m-1} \sum_{i=1}^m (Y_i - \bar{Y}_m)^2$ be the sample mean and sample variance of $Y_1, \ldots, Y_m$ respectively. For the two-sample *Welch*-test we have the following hypotheses:

$$H_0 : \mu_X - \mu_Y = 0 \quad vs \quad H_1 : \mu_X - \mu_Y \neq 0 \tag{A.1}$$

The $t$ statistic for the *Welch*-test is defined as

$$t = \frac{\bar{X}_n - \bar{Y}_m}{\sqrt{\widehat{\text{Var}}(\bar{X}_n - \bar{Y}_m)}} = \frac{\bar{X}_n - \bar{Y}_m}{\sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}}} \tag{A.2}$$

If $F_1$ and $F_2$ are normal distributions, $t$ is known to approximately follow a Student's $t$ distribution under $H_0$ with $\nu$ degrees of freedom, where $\nu$ is computed as

$$\nu = \frac{\left(\frac{s_X^2}{n} + \frac{s_Y^2}{m}\right)^2}{\frac{s_X^4}{n^2(n-1)} + \frac{s_Y^4}{m^2(m-1)}} \tag{A.3}$$

A two-sided *Welch*-test with a 95%-confidence level can be done in R very easily as follows:

```
set.seed(123)

## Settings
mu_x <- 0.1
sd_x <- 2
n <- 100

mu_y <- 0
sd_y <- 1
m <- 150

alpha <- 0.05

## Create X and Y
X <- rnorm(n, mean = mu_x, sd = sd_x)
Y <- rnorm(m, mean = mu_y, sd = sd_y)

## t-test
(tt <- t.test(x = X, y = Y, alternative = "two.sided",
              conf.level = 0.95, var.equal = FALSE))

##
##  Welch Two Sample t-test
##
## data:  X and Y
## t = 1.789, df = 135.68, p-value = 0.07584
```

```
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.03746002  0.74815974
## sample estimates:
##   mean of x   mean of y
##  0.28081182 -0.07453805
```

In this example we had two samples $X$ and $Y$ of size 100 and 150 respectively, $X$ with $\mu_X = 0.1$ and variance $\sigma_X^2 = 4$ and $Y$ with $\mu_Y = 0$ and $\sigma_Y^2 = 1$. The sample mean of $X$ is $\bar{X}_n = 0.2808118$, while the sample mean of $Y$ is $\bar{Y}_m = -0.074538$.

The R output shows that the confidence interval for the difference of means at a 95% confidence level is $[-0.03746, -0.03746]$. Since it contains the value of the null hypothesis, $\mu_X - \mu_Y = 0$, we cannot reject $H_0$ at the given confidence level.

Now let's use the bootstrap to compute confidence intervals for the difference of means. When using the *boot* package, we need to define the function of the statistic $\widehat{\theta}$ used as estimator for $\theta$, and then pass it as an argument to *boot*. This function has to accept two arguments: the first one is the original data; the second one is a vector of indices, and when used to subset the original data one gets a bootstrap sample. Let's define such function for the difference of means.

First, for the two-sample case we need to rearrange the data into a data frame:

```
groups <- as.factor(c(rep("X", n), rep("Y", m)))
df <- data.frame(obs = c(X, Y), group = groups)
```

so that we can define the function in terms of the data frame df:

```
diff_mean <- function(df, ind) {

  # Bootstrap resample
  X_star <- subset(df[ind, 1], df[ind, 2] == "X")
  Y_star <- subset(df[ind, 1], df[ind, 2] == "Y")

  # Bootstrap estimate of the difference of means
  diff_star <- mean(X_star) - mean(Y_star)

  # Bootstrap estimate of the variance of X_bar - Y_bar
  var_diff <- var(X_star)/n + var(Y_star)/m

  return(c(diff_star, var_diff))
}
```

Note that the function *diff_mean* returns the difference of means plus the estimated variance of the difference of means. The latter is used later to compute bootstrap $t$

confidence intervals (see Chapter 3).

Now we just need to load the *boot* package and call the *boot* function, passing to it the data *df*, the estimator function *diff_mean* and the number of bootstrap estimates. We additionally need to use the argument *strata*, and pass to it the factors associated to the observations. This will make sure that bootstrap resamples have the same proportion of observations from $X$ and $Y$ as in the original sample.

```
library(boot)
set.seed(1)
(bt <- boot(data = df, statistic = diff_mean,
            R = 1e3 - 1, strata = df[, 2]))

##
## STRATIFIED BOOTSTRAP
##
##
## Call:
## boot(data = df, statistic = diff_mean, R = 1000 - 1, strata = df[,
##     2])
##
##
## Bootstrap Statistics :
##       original        bias    std. error
## t1* 0.35534986 -0.0030370822 0.193355033
## t2* 0.03945362 -0.0002073271 0.004733292
```

The output shows three numbers for each "statistic" $t$, that is, for every element in the vector returned by *diff_mean*, in our case the sample mean and sample variance of the difference of means. *original* refers to the value of the estimator $\widehat{\theta}$, i.e. the value of the statistic supplied to *boot* evaluated at the original data, in our case $\bar{X}_n - \bar{Y}_m$ and $\frac{s_X^2}{n} + \frac{s_Y^2}{n}$. *bias* refers to the estimated bias of $\widehat{\theta}$, as defined in Equation (3.47). Lastly, *std. error* gives an estimation of the standard deviation of the bootstrap distribution $\widehat{\theta}^*$ of $\widehat{\theta}$, i.e. $\sqrt{\widehat{\mathrm{Var}}(\widehat{\theta}^*)}$.

The data is returned as a list object of class *boot* that we save in *bt* . The bootstrap estimates $\widehat{\theta}^{(*b)}$ of the difference of means and their estimated variances can be accessed at the list element *t*. It is a matrix with $R$ rows, each of which has the bootstrap estimates resulting from calling the supplied statistic function, in our case *diff_mean*.

```
head(bt$t)

##             [,1]       [,2]
## [1,] 0.18300767 0.05008876
## [2,] 0.37718256 0.03953191
## [3,] 0.74311139 0.04148653
```

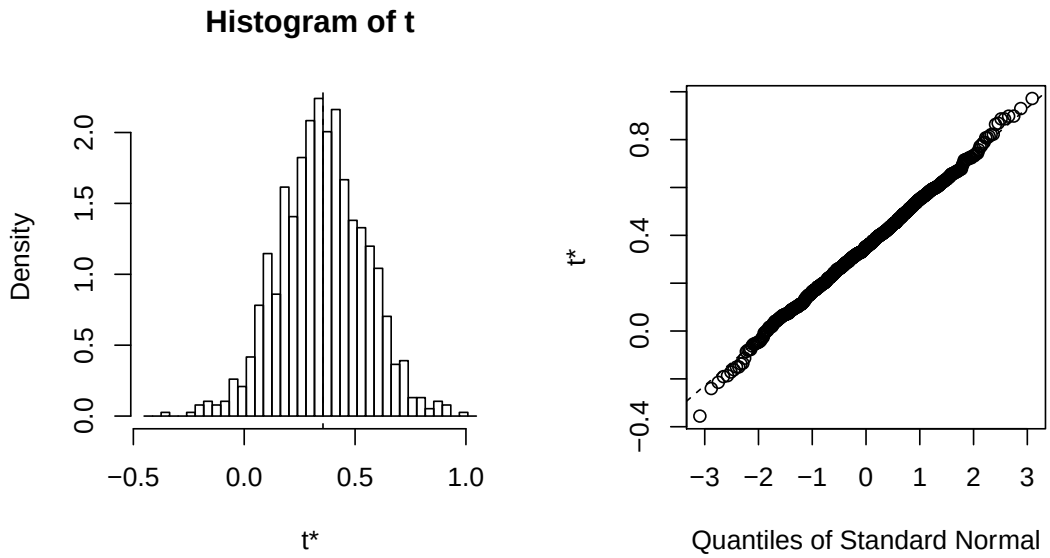


Figure A.1: Plot of boot object (1). Left: histogram of bootstrap estimates. Right: normal QQ plot of bootstrap estimates.

```
## [4,] 0.44728904 0.03508451
## [5,] 0.43608736 0.04253185
## [6,] 0.08218001 0.03798811
```

We can use the *plot* function on *bt*, which plots the bootstrap distribution of the difference of means and a normal QQ plot of such distribution. We specify the argument $index = 1$ to tell *plot.boot* to plot the first column of *bt$t*, although it is the default value:

```
plot(bt, index = 1)
```

We can do the same call but for $index = 2$ to get the same analysis for the bootstrap variance of the difference of means:

```
plot(bt, index = 2)
```

Note that the values called *original* in the *boot* output are drawn on the histograms with a dashed vertical line.

One attractive feature of the *boot* function is that it is prepared for parallel computation. One just needs to add a couple more arguments to the call. Setting *parallel = "multicore"* will use a fork of *ncpus* branches, and *parallel = "snow"* will use a Parallel Socket Cluster (PSOCK). In UNIX systems, both can be used, but in Windows only the PSOCK system can be used. There is a third argument, *cl*, to which we can pass our custom-made parallel

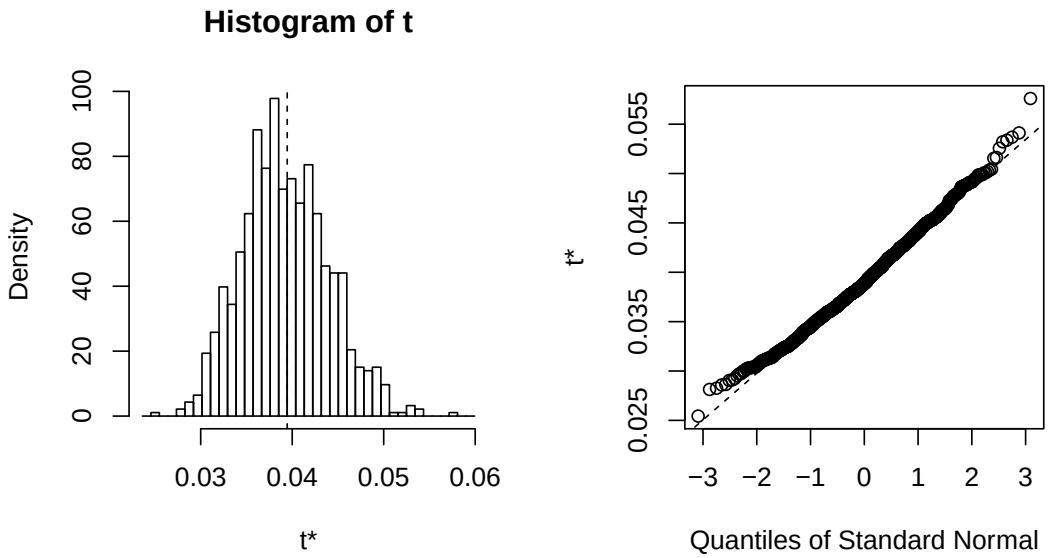


Figure A.2: Plot of boot object (2). Left: histogram of bootstrap estimates of the variance of the difference of means. Right: normal QQ plot of bootstrap estimates of the variance of the difference of means.

cluster. If no cluster is passed, *boot* will create one automatically, provided the other two arguments are specified.

```
bt <- boot(data = df, statistic = diff_mean,
           R = 1e3 - 1, strata = df[, 2],
           parallel = "multicore", ncpus = 6)
```

Once we have the bootstrap distribution of $\widehat{\theta}$, we can construct bootstrap confidence intervals for $\theta$. The function *boot.ci* from the *boot* package provides an easy way to compute five well-known bootstrap confidence intervals, all of which are explained in Chapter 3. The only one whose name does not fully match is the *Normal* interval, which is a normal interval with bias correction. One only needs to provide the function *boot.ci* with an object of class *boot* and a confidence level $1 - \alpha$:

```
(cis <- boot.ci(bt, conf = 1 - alpha))

## BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
## Based on 999 bootstrap replicates
##
## CALL :
## boot.ci(boot.out = bt, conf = 1 - alpha)
##
```

```
## Intervals :
## Level      Normal              Basic             Studentized
## 95%   (-0.0206,  0.7374 )   (-0.0137,  0.7507 )   (-0.0228,  0.7420 )
##
## Level     Percentile            BCa
## 95%   (-0.0400,  0.7244 )   (-0.0257,  0.7324 )
## Calculations and Intervals on Original Scale
```

All the computed bootstrap confidence intervals contain 0, which is the difference of means under $H_0$. Hence, we would not reject the null hypothesis $H_0$ at the 95% confidence level, whichever interval we decide to use.

## A.2 Nested *boot* Calls to Estimate the Variance of a Bootstrap Estimate

We have estimated the variance of $(\bar{X}_n^* - \bar{Y}_n^*)$ using an available theoretical result, because in a $t$-test it is assumed that the original populations are normally distributed. However, instead of using a theoretical result for the variance of the bootstrap estimate, we could also use a second bootstrap layer. This way we could get rid of the assumption of normality, by estimating the variance of $(\bar{X}_n - \bar{Y}_n)$ using another call to the *boot* function. This is called iterative or nested bootstrap. Note, however, that the variance of the bootstrap estimate is only needed for the studentized bootstrap (aka bootstrap $t$) confidence interval.

The statistic function that we have to pass to the *boot* function now also needs to compute the variance of $(\bar{X}_n^* - \bar{Y}_n^*)$. We estimate it with a second *boot* call, and then computing the variance of the second-layer bootstrap estiamtes.

```
diff_mean <- function (df, ind) {

  # Bootstrap resample
  X_star <- subset(df[ind, 1], df[ind, 2] == "X")
  Y_star <- subset(df[ind, 1], df[ind, 2] == "Y")

  # Original sample
  X <- subset(df[, 1], df[, 2] == "X")
  Y <- subset(df[, 1], df[, 2] == "Y")

  # Bootstrap estimate of the difference of means
  diff_star <- mean(X_star) - mean(Y_star)

  # Bootstrap estimate of the variance of X_bar* - Y_bar*
```

```
  boot_diff_star <- boot(data = df[ind, ], statistic = diff_layer2,
                         R = 200, strata = df[ind, 2])
  var_diff_star <- var(boot_diff_star$t)

  return(c(diff_star, var_diff_star))
}
```

The function $diff\_layer2$ is the second-layer bootstrap statistic function to estimate the second-layer bootstrap estimates of the difference of means. Note that despite being bootstrapped only $R = 200$ times, it turns out to be a large enough number to estimate accurately the variance of a random variable (Efron and Tibshirani, 1993).

```
diff_layer2 <- function (df2, ind2) {

  # Second-layer bootstrap resample
  X_star2 <- subset(df2[ind2, 1], df2[ind2, 2] == "X")
  Y_star2 <- subset(df2[ind2, 1], df2[ind2, 2] == "Y")

  # Second-layer bootstrap estimate of the difference of means
  diff_star2 <- mean(X_star2) - mean(Y_star2)

  return(diff_star2)
}
```

We can now call the *boot* function as before:

```
set.seed(1)
(bt <- boot(data = df, statistic = diff_mean,
            R = 1e3 - 1, strata = df[, 2]))

##
## STRATIFIED BOOTSTRAP
##
##
## Call:
## boot(data = df, statistic = diff_mean, R = 1000 - 1, strata = df[,
##     2])
##
##
## Bootstrap Statistics :
##       original       bias    std. error
## t1* 0.35534986 -0.003037082 0.193355033
## t2* 0.03987877 -0.001105656 0.006127428

#(bt <- boot(data = df, statistic = diff_mean,
```

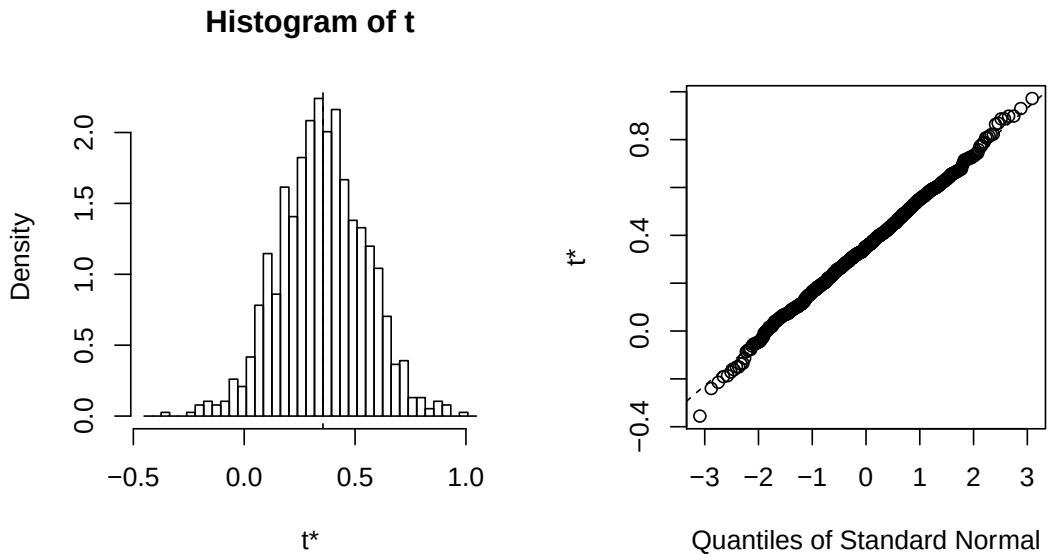


Figure A.3: Plot of boot object with double layer bootstrap. Left: histogram of bootstrap estimates. Right: normal QQ plot of bootstrap estimates.

```
#             R = 1e3 - 1, strata = df[, 2],
#             parallel = "multicore", ncpus = 6))
```

Similarly as before, we can plot the object *bt* to see the bootstrap distribution of the difference of means:

```
plot(bt)
```

Lastly, we can compute bootstrap confidence intervals by calling the aforementioned *boot.ci* function:

```
(cis <- boot.ci(bt, conf = 1 - alpha))
```

```
## BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
## Based on 999 bootstrap replicates
##
## CALL :
## boot.ci(boot.out = bt, conf = 1 - alpha)
##
## Intervals :
## Level      Normal              Basic             Studentized
## 95%   (-0.0206,  0.7374 )   (-0.0137,  0.7507 )   (-0.0207,  0.7443 )
##
```

```
## Level     Percentile            BCa
## 95%   (-0.0400,  0.7244 )   (-0.0257,  0.7324 )
## Calculations and Intervals on Original Scale
```

The reader can check that all intervals but the studentized interval are the same as previously, which is the only one affected by the computation of the variance of the bootstrap estimate.

As before, all the computed bootstrap confidence intervals contain 0, which is the difference of means under $H_0$. Hence, we would not reject the null hypothesis $H_0$ at the 95% confidence level, whichever interval we decide to use.

# Declaration of Originality

The signed declaration of originality is a component of every semester paper, Bachelor's thesis, Master's thesis and any other degree paper undertaken during the course of studies, including the respective electronic versions.

Lecturers may also require a declaration of originality for other written papers compiled for their courses.

---

I hereby confirm that I am the sole author of the written work here enclosed and that I have compiled it in my own words. Parts excepted are corrections of form and content by the supervisor .

**Title of work** (in block letters):

| |
|---|
| Review of bootstrap principles and coverage analysis of bootstrap confidence intervals for common estimators |

**Authored by** (in block letters):
*For papers written by groups the names of all authors are required.*

| **Name(s):** | **First name(s):** |
|---|---|
| Fuentes Pérez | Eufemiano |
| | |
| | |
| | |
| | |

With my signature I confirm that

- I have committed none of the forms of plagiarism described in the Citation etiquette information sheet.
- I have documented all methods, data and processes truthfully.
- I have not manipulated any data.
- I have mentioned all persons who were significant facilitators of the work .
- I am aware that the work may be screened electronically for plagiarism.
- I have understood and followed the guidelines in the document *Scientific Works in Mathematics*.

| **Place, date:** | **Signature(s):** |
|---|---|
| Zürich, December 17th 2019 |  |

*For papers written by groups the names of all authors are required. Their signatures collectively guarantee the entire content of the written paper.*