Bootstrap-based hypothesis testing

J.S. Allison, M.Sc.

Thesis submitted for the degree Philosophiae Doctor in Statistics at the North-West University

Promoter: Prof. J.W.H. Swanepoel

November 2008

Potchefstroom

ADDRESSING REVIEWER COMMENTS

BAD REVIEWS ON YOUR PAPER? FOLLOW THESE GUIDELINES AND YOU MAY YET GET IT PAST THE EDITOR:

Reviewer comment:

"The method/device/paradigm the authors propose is clearly wrong."

How NOT to respond:

✗ "Yes, we know. We thought we could still get a paper out of it. Sorry."

Correct response:

✓ "The reviewer raises an interesting concern. However, as the focus of this work is exploratory and not performance-based, validation was not found to be of critical importance to the contribution of the paper."

Reviewer comment:

"The authors fail to reference the work of Smith et al., who solved the same problem 20 years ago."

How NOT to respond:

✗ "Huh. We didn't think anybody had read that. Actually, their solution is better than ours."

Correct response:

✓ "The reviewer raises an interesting concern. However, our work is based on completely different first principles (we use different variable names), and has a much more attractive graphical user interface.

Reviewer comment:

"This paper is poorly written and scientifically unsound. I do not recommend it for publication."

How NOT to respond:

✗ "You #&@*% reviewer! I know who you are! I'm gonna get you when it's my turn to review!"

Correct response:

✓ "The reviewer raises an interesting concern. However, we feel the reviewer did not fully comprehend the scope of the work, and misjudged the results based on incorrect assumptions.

www.phdcomics.com

Reference: **www.phdcomics.com**

Summary

One of the main objectives of this dissertation is the development of a new method of evaluating the performance of bootstrap-based tests. The evaluation method that is currently in use in the literature has some major shortcomings, for example, it does not allow one to determine the robustness of a bootstrap estimator of a critical value. This is because the evaluation and the estimation are based on the same data. This traditional method of evaluation often leads to too optimistic probability of type I errors when bootstrap critical values are used.

We show how this new, more robust, method can detect defects of bootstrap estimated critical values which cannot be observed if one uses the current evaluation method. Based on the new evaluation method, some theoretical properties regarding the bootstrap critical value are derived when testing for the mean in a univariate population. These theoretical findings again highlight the importance of the two guidelines proposed by Hall and Wilson (1991) for bootstrap-based testing, namely that resampling must be done in a way that reflects the null hypothesis and bootstrap tests should be based on test statistics that are pivotal (or asymptotically pivotal).

We also developed a new nonparametric bootstrap test for Spearman's rho and, based on the results obtained from a Monte-Carlo study, we recommend that this new test should be used when testing for Spearman's rho. A semiparametric test based on copulas was also developed as a useful benchmark tool for measuring the performance of the nonparametric test.

Other research objectives of this dissertation include, among others, a brief overview of the nonparametric bootstrap and a general formulation of methods which can be used to apply the bootstrap correctly when conducting hypothesis testing.

Opsomming

Een van die vernaamste mikpunte van die proefskrif is die ontwikkeling van 'n nuwe metode om die gedrag van skoenlusgebaseerde toetse te evalueer. Die evalueringsmetode wat huidiglik in die literatuur gebruik word, besit ernstige tekortkominge. Dit kan byvoorbeeld nie aangewend word om die robuustheid van 'n skoenlusberamer van 'n kritiekewaarde te bepaal nie. Die rede hiervoor is dat die evaluering en beraming gebaseer is op dieselfde data. Die tradisionele metode van evaluering lei dikwels na 'n te optimistiese waarskynlikheid van tipe I fout, wanneer skoenlus kritiekewaardes gebruik word.

Daar word aangetoon hoe hierdie nuwe, meer robuuste, evalueringsmetode tekortkominge van skoenlus kritiekewaardes kan opspoor, wat nie vasgestel kan word as die huidige evalueringsmetode aangewend word nie. Gebaseer op hierdie nuwe metode, word daar teoretiese eienskappe vir die skoenlus kritiekewaarde afgelei, indien daar getoets word vir die gemiddeld in 'n eenveranderlike populasie. Hierdie teoretiese bevindinge beklemtoon weereens die belangrikheid van die twee riglyne wat Hall en Wilson (1991) voorgestel het, naamlik dat hersteekproefneming toegepas word op so 'n manier dat die nulhipotese gereflekteer word en skoenlustoetse gebaseer word op toetsstatistieke wat spilgroothede (of asimptotiese spilgroothede) is.

'n Nuwe effektiewe nie-parametriese skoenlustoets vir Spearman se rho is ontwikkel. 'n Semi-parametriese toets gebaseer op "copulas" is ook ontwikkel as 'n nuttige maatstaf om die prestasie van die nie-parametriese toets te meet.

Ander navorsingsdoelwitte sluit onder andere in, 'n kort oorsig van die nie-parametriese skoenlusmetode en 'n algemene formulering van metodes wat gebruik kan word om die skoenlusmetode reg toe te pas, wanneer hipotesetoetsing uitgevoer word.

Bedankings

Die skrywer wil hiermee graag die volgende bedankings doen:

- Prof. J.W.H. Swanepoel, vir sy leiding, insig, entoesiasme en voortgesette ondersteuning wat noodsaaklik was vir die voltooiing van hierdie studie.
- Leonard Santana en Gerhard Koekemoer vir waardevolle samesprekings en rekenaar hulp.
- My ouers, vir liefde, opvoeding, ondersteuning, belangstelling en bystand.
- My ouma, vir liefde, ondersteuning en intense belangstelling.
- My sussie, vir liefde en hulp met die inlees van die tabelle.

Vir geleenthede en die voorreg om hierdie taak te voltooi, bring ek dank aan God.

Contents

Chapter 1

Introduction

1.1 Overview

The basic objective of statistical analysis is "extracting all the information from the data" (Rao, 1989) to deduce properties about the population that generated the data. Most statistical analyses are based on functions of data, called statistics. Prior to obtaining data, uncertainty exists as to what value of any particular statistic will result. A statistic is, therefore, a random variable with a probability distribution, called the sampling distribution of the statistic. A central objective of statistical inference is to characterise this sampling distribution. Knowledge of this distribution enables, among other things, measurement of the precision and bias of the estimate, development of confidence intervals and the testing of hypotheses about the parameter being estimated.

The basic idea of any sort of hypothesis test is to compare the observed value of the test statistic with the distribution that it would follow if the null hypothesis were true (null distribution). The null hypothesis is then rejected if the observed value of the test statistic is sufficiently large/small relative to the null distribution.

In most cases, however, the null distribution will be unknown because it depends on the underlying population. We therefore have to compare the observed value of the test statistic with a distribution that is only approximately correct. Traditionally, the approximation used was based on asymptotic theory. However, hypothesis tests based on asymptotic theory may not give good results, especially in small samples. Many such examples exist, for example, Davidson and MacKinnon (1992) reported a simulation in which a version of the information matrix test rejects a true null hypothesis 99.9% of the time, even for large samples; Stewart (1997) and Dufour and Khalaf

(2002) found that in the case of multivariate regression models, standard asymptotic tests often over reject severely, while Boos and Brownie (1989) found that, when testing for homogeneity of variances, the asymptotic test does not give satisfactory results in terms of estimated sizes.

Another approach would be to estimate or approximate the sampling distribution of the statistic (or the null distribution in the case of hypothesis testing) from the observed data. Efron (1979) introduced a general resampling scheme, the "bootstrap", which can be used to do exactly this. Asymptotic theory indicates that bootstrap tests will generally perform better in finite samples than asymptotic tests, in the sense that the errors made based on the bootstrap tests will be of a lower order in the sample size (see, among others, Beran, 1988; Hall, 1992).
This resampling technique applied to hypothesis testing will form the cornerstone of this thesis: "Bootstrap-based hypothesis testing".

1.2 Objectives

The main objectives of this dissertation can be summarized as follows:

- Provide an overview of the nonparametric bootstrap.
- Present a general formulation of three methods which can be used to apply the bootstrap correctly when conducting hypothesis testing.
- Investigate the application of bootstrap-based testing for some common statistical problems.
- Develop a new method to evaluate the performance of bootstrap-based tests.
- Derive some theoretical properties regarding the bootstrap estimator of the critical value, when testing for the mean in a univariate population. This will be based on the new method to evaluate the performance of a bootstrap-based test.
- Develop a new semiparametric bootstrap test for Spearman's rho based on copulas.
- Develop a new nonparametric bootstrap test for Spearman's rho.

1.3 Thesis outline

Chapter 2 provides a brief overview of some recent developments concerning the nonparametric bootstrap methodology. Most topics will be discussed under the assumption of independent data.

The chapter concludes with a discussion on the application of the bootstrap to dependent data.

In Chapter 3 we discuss two methods to apply the bootstrap correctly to hypothesis testing. We will make use of some examples of common statistical problems to illustrate how these methods can be applied.

Chapter 4 introduces a new method which can be used to evaluate the performance of bootstrap-based tests. We also derive some theoretical properties regarding the bootstrap estimator of the critical value when testing for the mean in a univariate population. This will be based on the new method for evaluating the performance of a bootstrap-based test. The results of a Monte-Carlo study that substantiate the theoretical findings will also be presented.

In Chapter 5 we discuss how to resample residuals in order to apply the bootstrap correctly to hypothesis testing. The shortcomings associated with resampling incorrectly will be analysed by making use of a simple example.

Chapter 6 provides an overview of copulas, focusing especially on how to simulate from different families of copulas. The chapter concludes with a discussion on two measures of association, namely Spearman's rho and Kendall's tau.

In Chapter 7 we consider bootstrap-based testing for Spearman's rho. We propose two new tests for Spearman's rho: a semiparametric bootstrap test, based on copulas, and a nonparametric bootstrap test. In the last part of the chapter we propose a nonparametric bootstrap test to test whether the Spearman's rho of two bivariate populations are equal to one another.

Chapter 8 contains the results of numerous Monte-Carlo studies. The chapter concludes by looking at possible future research in the area of bootstrap-based testing.

Chapter 2

An overview of the bootstrap

This chapter provides a brief overview of some recent developments concerning the nonparametric bootstrap methodology, concentrating on basic ideas and applications rather than theoretical considerations. Topics include statistical error, confidence intervals, double bootstrapping, bootstrap calibration, bootstrap partial likelihood, bootstrapping complicated data sets and the modified bootstrap. The above topics will be discussed under the assumption of independent data. A major development in bootstrap methods has been their application to dependent data. Topics which will be discussed under this heading include the moving block bootstrap and the autoregressive sieve bootstrap.

2.1 Introduction

As was mentioned in Chapter 1, Efron (1979) introduced a very general resampling scheme (the "bootstrap") for estimating or approximating the sampling distributions of statistics. Efron and Tibshirani (1993) defined this technique as follows:

"A computer-based method for assigning measures of accuracy to statistical estimates."

An alternative definition would be: *A computer-based technique that enables one to estimate the distributional properties of a statistic.* The bootstrap is essentially a method that attempts to mimic the process of sampling from a population (like one does in Monte-Carlo simulations), by instead drawing samples from the observed sample data.

It has many attractive properties, especially for the statistical practitioner: it requires few assumptions, little modelling or analysis is required and it can be applied in an automatic way in a wide variety of situations. Efron an Tibshirani (1985) summarizes one of the most important benefits of the bootstrap methodology as follows:

"The bootstrap can answer questions which are too complicated for traditional statistical analysis."

2.2 Bootstrap estimate of standard error

Assume we have a random sample $\mathbf{X}_n = (X_1, X_2, \ldots, X_n)$ from an unknown distribution function (d.f.) F. Bootstrap methods depend on, what is referred to as, a bootstrap sample:
Let F_n be the empirical distribution function of $\mathbf{X}_n$ that places probability $1/n$ on each X_i, $i = 1, \ldots, n$. A bootstrap sample is defined as a random sample of size n drawn from F_n, say

$$\mathbf{X}_n^* = (X_1^*, X_2^*, \ldots, X_n^*).$$

The star notation indicates that $\mathbf{X}_n^*$ is not the same as the actual data set $\mathbf{X}_n$, but rather a randomised, or resampled version of $\mathbf{X}_n$. In other words, $\mathbf{X}_n^*$ is a random sample of size n, drawn with replacement from the "population" of n objects $(X_1, \ldots, X_n)$.
More formally, we write, for $j = 1, \ldots, n$,

$$P^*(X_j^* = X_i) = 1/n, \quad \text{for} \quad i = 1, \ldots, n,$$

where P^* denotes the conditional probability law of $\mathbf{X}_n^*$ given $\mathbf{X}_n$.
Suppose $\hat{\theta} = \hat{\theta}(X_1, X_2, \ldots, X_n)$ is some estimator of a parameter θ.
The standard error of $\hat{\theta}$ is

$$\sigma(F) = \{Var_F(\hat{\theta})\}^{1/2}$$

and the bootstrap estimate of $\sigma(F)$ is simply

$$\hat{\sigma} \equiv \sigma(F_n) = \{Var_*(\hat{\theta}^*)\}^{1/2}, \quad \text{where}$$

$\hat{\theta}^* = \hat{\theta}(X_1^*, \ldots, X_n^*)$ and Var_* denotes the variance under P^*.

One reason for the success of the bootstrap method is that a simple and accurate Monte-Carlo approximation can be given for $\hat{\sigma}$ (Efron, 1979):

i) For $b = 1, 2, \ldots, B$ (large), generate independent bootstrap samples from F_n:

$$\mathbf{X}_n^*(b) = (X_1^*(b), X_2^*(b), \ldots, X_n^*(b)).$$

ii) Calculate $\hat{\theta}^*(1), \hat{\theta}^*(2), \ldots, \hat{\theta}^*(B)$ (so-called bootstrap replications), where $\hat{\theta}^*(b) = \hat{\theta}(X_1^*(b), X_2^*(b), \ldots, X_n^*(b))$.

iii) Approximate $\hat{\sigma}$ by

$$\hat{\sigma}_B = \left\{ \frac{1}{B-1} \sum_{b=1}^{B} (\hat{\theta}^*(b) - \hat{\theta}^*(\cdot))^2 \right\}^{1/2} \text{, where}$$

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

Remarks:

(a) The strong law of large numbers implies that $\hat{\sigma}_B \to \hat{\sigma}$ a.s. as $B \to \infty$.

(b) Booth and Sarkar (1998) suggest that a choice of B between 200 and 800 is satisfactory for the estimation of standard errors and the construction of confidence intervals. With the powerful computers we have nowadays a safe choice to use is B=1000.

(c) Bootstrap estimators of other measures of statistical error (or accuracy), such as bias or prediction error, can be obtained in a similar manner.

(d) The bootstrap method, discussed above, is often called the nonparametric bootstrap.

(e) If $F(\cdot) = G(\cdot, \boldsymbol{\theta})$, with G a known cumulative distribution function (c.d.f.) and $\boldsymbol{\theta}$ a vector of unknown parameters, we can estimate $\boldsymbol{\theta}$ by their sample estimate, $\hat{\boldsymbol{\theta}}$, and generate bootstrap random samples $\mathbf{X}_n^* = (X_1^*, X_2^*, \ldots, X_n^*)$ from $G(\cdot, \hat{\boldsymbol{\theta}})$ and then continue as before. This is known as the parametric bootstrap.

2.3 The double bootstrap

The question of how accurate $\hat{\sigma}$ is (the bootstrap estimate of the standard error of $\hat{\theta}$, defined in Section 2.2) now arises. What is, for example, the standard error of the bootstrap standard error? The bootstrap can, once again, be applied to estimate this quantity. The standard error of the bootstrap estimate of standard error of $\hat{\theta}$ is denoted by:

$$\tau(F) = \{Var_F(\hat{\sigma})\}^{1/2}.$$

The bootstrap estimator is then simply

$$\hat{\tau} \equiv \tau(F_n) = \{Var_*(\hat{\sigma}^*)\}^{1/2},$$

where $\hat{\sigma}^* = \hat{\sigma}(X_1^*, \ldots, X_n^*)$.

The double bootstrap thus involves resampling resampled data, i.e., bootstrapping the bootstrap (see, e.g., Chapman and Hinkley, 1986). The following Monte-Carlo approximation can be given for $\hat{\tau}$:

i) Generate a bootstrap sample $X_1^*, \ldots, X_n^*$ from F_n:

   (a) Generate a bootstrap sample $X_1^{**}, \ldots, X_n^{**}$ from F_n^*, the e.d.f. of $X_1^*, \ldots, X_n^*$, and calculate

   $$\hat{\theta}^{**} = \hat{\theta}(X_1^{**}, \ldots, X_n^{**})$$

   Denote this by $\hat{\theta}^{**}(1)$.

   (b) Repeat step (a) C times independently, to obtain bootstrap replications

   $$\hat{\theta}^{**}(1), \ldots, \hat{\theta}^{**}(C).$$

   (c) Calculate

   $$\hat{\sigma}_C(1) = \left\{ \frac{1}{C-1} \sum_{c=1}^{C} \left(\hat{\theta}^{**}(c) - \frac{1}{C} \sum_{c=1}^{C} \hat{\theta}^{**}(c) \right)^2 \right\}^{1/2}.$$

ii) Repeat step i) B times independently, to obtain $\hat{\sigma}_C(1), \ldots, \hat{\sigma}_C(B)$.

iii) Calculate

$$\hat{\tau}_{B,C} = \left\{ \frac{1}{B-1} \sum_{b=1}^{B} \left(\hat{\sigma}_C(b) - \frac{1}{B} \sum_{b=1}^{B} \hat{\sigma}_C(b) \right)^2 \right\}^{1/2}.$$

Note that the strong law of large numbers implies that

$$\hat{\tau}_{B,C} \to \hat{\tau} \quad \text{a.s.} \quad \text{as} \quad B, C \to \infty.$$

Remarks:

(a) The double (or nested) bootstrap has been applied to various problems in the statistical literature (see, e.g., Tibshirani, 1988; Davison and Hinkley, 1997).

(b) Important applications include, among others, its application to the construction of "bootstrap calibration confidence intervals" (which will be discussed later) and the construction of "bootstrap partial likelihoods".

2.4 Partial likelihood approach

The bootstrap partial likelihood approach estimates the likelihood function of $\hat{\theta}$, an estimator for θ, using a double bootstrap procedure. In order to estimate this likelihood function Davison et al. (1992) proposed the following procedure:

i) Generate bootstrap samples $\mathbf{X}_n^*(1), \mathbf{X}_n^*(2), \ldots, \mathbf{X}_n^*(B)$, giving bootstrap replications $\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*$.

ii) From each of $\mathbf{X}_n^*(b), b = 1, \ldots, B$, generate C second stage bootstrap samples, giving second stage bootstrap replicates $\hat{\theta}_1^{**}(b), \ldots, \hat{\theta}_C^{**}(b)$.

iii) Calculate, e.g., the kernel density estimate

$$\hat{f}(t|\hat{\theta}_b^*) = \frac{1}{Ch} \sum_{c=1}^{C} k\left(\frac{t - \hat{\theta}_c^{**}(b)}{h}\right),$$

for $b = 1, \ldots, B$. Here k is a known symmetric density function and h the bandwidth.

iv) Evaluate $\hat{f}(\hat{\theta}|\hat{\theta}_b^*)$ for $b = 1, \ldots, B$.

v) $\hat{f}(\hat{\theta}|\hat{\theta}_b^*)$ provides an estimate of the likelihood of θ for a parameter value $\theta = \hat{\theta}_b^*$.

A smooth estimate of the likelihood of θ is then obtained by applying a scatter plot smoother to the pairs $\left(\hat{\theta}_b^*, \hat{f}(\hat{\theta}|\hat{\theta}_b^*)\right)$, $b = 1, \ldots, B$. This construction is called bootstrap partial likelihood because it estimates the likelihood of θ based on $\hat{\theta}$ rather than on the full data set $\mathbf{X}_n$. The interested reader is referred to Davison et al. (1992) for a more in depth discussion of this partial likelihood approach.

2.5 Estimation of sampling distributions

Consider the problem of estimating the sampling distribution of a random variable $R_n(\mathbf{X}_n; F)$:

$$H_F(x) = P_F(R_n(\mathbf{X}_n; F) \leq x), \quad x \in \mathbb{R}.$$

The bootstrap estimator of $H_F(x)$ is simply

$$\hat{H}(x) = H_{F_n}(x) = P^*(R_n(\mathbf{X}_n^*; F_n) \leq x).$$

Note: If, e.g., $R_n(\mathbf{X}_n; F) = \sqrt{n}(\bar{X}_n - \mu)/S_n(\mathbf{X}_n)$, with $\bar{X}_n$ and $S_n(\mathbf{X}_n)$ defined as the sample mean and sample standard deviation respectively, then the bootstrap statistic becomes

$$R_n(\mathbf{X}_n^*; F_n) = \sqrt{n}(\bar{X}_n^* - \bar{X}_n)/S_n(\mathbf{X}_n^*).$$

The Monte-Carlo approximation of $\hat{H}(x)$ is then simply:

$$\hat{H}_B(x) = \frac{1}{B}\sum_{b=1}^{B} I\left(R_n\left(\mathbf{X}_n^*(b); F_n\right) \leq x\right),$$

where $\mathbf{X}_n^*(1), \dots, \mathbf{X}_n^*(B)$ are independent bootstrap samples of size n drawn from F_n.

Remarks:

(a) The approximation $H_{F_n}(x) \approx H_F(x)$ is asymptotically ($n \to \infty$) valid in a large number of situations. This is usually established by proving theorems of the form:

$$\sup_{-\infty<x<\infty} |H_{F_n}(x) - H_F(x)| = o(1),$$

almost surely (or in probability). We shall also refer to this by saying that the bootstrap estimator is "first-order accurate". If $o(1)$ can be replaced by $o(n^{-1/2})$, the bootstrap estimator is said to be "second-order accurate". In this case the bootstrap approximation is better than the normal approximation, which is typically of the order $O(n^{-1/2})$.

(b) First- and second-order accuracy results have been proved for a large number of statistics in the literature, including L-estimators, M-estimators, U-statistics, nonparametric density and regression estimators, U-quantiles, empirical and quantile processes, and general classes of statistical functionals (see, e.g., Hall, 1992; Shao and Tu, 1995; Janssen, 1997; Jiménez-Gamero et al., 2003).

2.6 Bootstrap confidence intervals

Suppose that $X_1, X_2, \dots, X_n$ are i.i.d. random variables with c.d.f. F and $\hat{\theta} = \hat{\theta}(X_1, X_2, \dots, X_n)$ is an estimator for a parameter θ. Let $\hat{\sigma}$ be the estimated standard error of $\hat{\theta}$.
A $100(1-\alpha)\%$-confidence interval (c.i.) for θ based on the traditional standard normal method is:

$$I_{1-\alpha}^N = [\hat{\theta} - z(\alpha/2)\hat{\sigma}; \hat{\theta} + z(\alpha/2)\hat{\sigma}],$$

where $z(\alpha/2)$ is the $100(1-\alpha/2)\%$ percentile point of Φ (the standard normal distribution), e.g., if $1-\alpha = 0.95$, then $z(\alpha/2) = z(0.025) = 1.96$.
However, these standard intervals for θ can often be quite inaccurate, because they rely on the

central limit theorem and the conditions under which this theorem is valid. For small to moderate sample sizes,

$$|P(\theta \in I_{1-\alpha}^N) - (1-\alpha)| \text{ can be large.}$$

The bootstrap can be used to construct c.i.'s that frequently perform better, especially for small and moderate sample sizes.
Since the early 1980's, a bewildering array of methods for constructing bootstrap confidence intervals have been proposed (see, e.g., Hall, 1988, 1992; Swanepoel, 1990; Efron and Tibshirani, 1993; Shao and Tu, 1995; DiCiccio and Efron, 1996; Davison and Hinkley, 1997).
Carpenter and Bithell (2000) wrote an interesting paper addressing the questions of when bootstrap confidence intervals should be used, which method should be chosen and how it should be implemented.

Suppose F is unknown. The best-known procedures for constructing nonparametric bootstrap confidence intervals are:

1. percentile, bias-corrected percentile (BC), accelerated bias-corrected percentile (BC_a) and
2. bootstrap-t intervals.

We now provide a quick summary of these procedures (a more detailed discussion can be found, e.g., in the books by Efron and Tibshirani, 1993; Shao and Tu, 1995).

2.6.1 Percentile Intervals

Let $\hat{G}$ denote the c.d.f. of $\hat{\theta}^* = \hat{\theta}(X_1^*, \ldots, X_n^*)$, i.e.,

$$\hat{G}(t) = P^*(\hat{\theta}^* \leq t).$$

The percentile $100(1-\alpha)\%$ c.i.'s are given by

$$I_{1-\alpha} = \left[\hat{G}^{-1}(\alpha/2), \hat{G}^{-1}(1-\alpha/2)\right],$$

which can be approximated by the Monte-Carlo method as follows:

i) Draw B independent bootstrap samples of size n and calculate $\hat{\theta}_1^*, \hat{\theta}_2^*, \ldots, \hat{\theta}_B^*$.

ii) Calculate the corresponding order statistics $\hat{\theta}_{(1)}^* \leq \hat{\theta}_{(2)}^* \leq \cdots \leq \hat{\theta}_{(B)}^*$.

iii) Approximate $I_{1-\alpha}$ by

$$\hat{I}_{1-\alpha} = \left[\hat{\theta}^*_{(r)}, \hat{\theta}^*_{(s)}\right],$$

where $r = \lfloor B\alpha/2 \rfloor$, $s = \lfloor B(1-\alpha/2) \rfloor$ and $\lfloor x \rfloor$ denotes the largest integer less than or equal to x.

Remarks:

(a) An alternative percentile interval, in the literature known as the "basic c.i." (see, e.g., Davison and Hinkley, 1997), is given by

$$\hat{I}^0_{1-\alpha} = \left[2\hat{\theta} - \hat{\theta}^*_{(s)}, 2\hat{\theta} - \hat{\theta}^*_{(r)}\right].$$

(b) The BC percentile c.i.'s are merely adjustments of the percentile intervals and they attempt to eliminate the effects of the bias of the bootstrap distribution of $\hat{\theta}^*$. They can be calculated similarly as $\hat{I}_{1-\alpha}$, except that they make use of other values of r and s.

(c) The BC_a percentile method is an improved version of the BC percentile method. It incorporates both a bias and skewness correction. The BC_a percentile c.i.'s can also be calculated as $\hat{I}_{1-\alpha}$, using different values of r and s (these values are somewhat complex to compute).

2.6.2 Bootstrap-t Intervals

A $100(1-\alpha)\%$ two-sided symmetric bootstrap-t c.i. is given by

$$\left[\hat{\theta} - q(F_n)\hat{\sigma}, \hat{\theta} + q(F_n)\hat{\sigma}\right],$$

where $q(F_n)$ is defined by

$$P^*\left(|\hat{\theta}^* - \hat{\theta}|/\hat{\sigma}^* \leq q(F_n)\right) \approx 1 - \alpha.$$

$q(F_n)$ can be approximated by obtaining B independent bootstrap replications

$$T^*_n(b) = |\hat{\theta}^*(b) - \hat{\theta}|/\hat{\sigma}^*(b), \quad b = 1, \ldots, B$$

and then finding the $\lfloor B(1-\alpha) \rfloor$-th smallest among the $T^*_n(b)$'s.

Remarks:

(a) Two-sided equal-tailed and one-sided bootstrap-t c.i.'s can be derived similarly.

(b) Both the BC_a percentile interval and the bootstrap-t interval are second-order accurate; that is, their coverage probabilities differ from the nominal $1-\alpha$ level by only $O(n^{-1})$, instead of $O(n^{-1/2})$, which is usually achieved by the standard c.i.'s based on quantiles of the standard normal distribution (Hall, 1988).

(c) However, these procedures also have some drawbacks. The BC_a procedure depends on some tuning parameter a that has to be estimated satisfactorily. The performance of the bootstrap-t procedure is highly dependent on the quality of the estimator $\hat{\sigma}$. For nonlinear statistics the derivation of a good estimator $\hat{\sigma}$ can be problematic.

(d) In view of (c) above, the use of the simple percentile method for small and moderate sample sizes is encouraged, provided these intervals are calibrated.

2.7 Bootstrap calibration

Bootstrap calibration confidence intervals were first proposed by Beran (1987) and Loh (1987) and have become very popular in recent years. The basic idea is to improve the original c.i. $I_{1-\alpha}$ by adjusting its nominal level $1-\alpha$ through a double bootstrap. This is accomplished as follows: Recall that $\hat{G}(t) = P^*(\hat{\theta}^* \leq t)$, hence

$$\begin{aligned} P_F(\theta \in I_{1-\alpha}) &= P_F\left(|2\hat{G}(\theta) - 1| \leq 1 - \alpha\right) \\ &=: \Pi_F(1-\alpha). \end{aligned}$$

Suppose

$$\Pi_F(1-\alpha) \neq 1-\alpha,$$

then choose a λ, $0 < 1 - \alpha + \lambda < 1$, such that

$$\Pi_F(1-\alpha+\lambda) = 1-\alpha.$$

Note that λ is unknown, since

$$\lambda = \Pi_F^{-1}(1-\alpha) - (1-\alpha).$$

The bootstrap estimator of λ is simply

$$\hat{\lambda} = \Pi_{F_n}^{-1}(1-\alpha) - (1-\alpha)$$

and the adjusted interval is, therefore,

$$I_{1-\alpha+\hat{\lambda}} = \left[\hat{G}^{-1}\left(\frac{\alpha-\hat{\lambda}}{2}\right), \hat{G}^{-1}\left(1 - \frac{\alpha-\hat{\lambda}}{2}\right)\right].$$

Remarks:

(a) Booth and Hall (1994) provide a discussion of the Monte-Carlo approximation of $I_{1-\alpha+\hat{\lambda}}$.

(b) Hall and Martin (1989) cautioned that bootstrap calibration of percentile-method intervals has no role to play in quantile problems and thus cannot be used to improve coverage accuracy.

2.8 Bootstrapping complicated data sets

The bootstrap can be applied to much more complicated situations. A regression model is a familiar example of a complicated data structure. Consider, for example, the following model:

$$Y_i = g(\boldsymbol{\beta}; \mathbf{x}_i) + \varepsilon_i, \quad i = 1, \ldots, n.$$

Here, $\boldsymbol{\beta}$ is a $(p \times 1)$ vector of unknown parameters; for each i, $\mathbf{x}_i$ is a $(p \times 1)$ observed vector of covariates and g is a known function. The ε_i's are i.i.d. random errors with unknown c.d.f. F such that $E(\varepsilon_i) = 0, \quad i = 1, \ldots, n$.

Suppose that $\boldsymbol{\beta}$ is estimated by $\hat{\boldsymbol{\beta}}$ (e.g., the least-squares estimator). The bootstrap can be applied to approximate the sampling distribution of

$$\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}\left((Y_1, \mathbf{x}_1), \ldots, (Y_n, \mathbf{x}_n)\right)$$

as follows:

(1) Let F_n be the e.d.f. of the centered residuals, defined for $i = 1, \ldots, n$, by

$$\hat{\varepsilon}_i = Y_i - g(\hat{\boldsymbol{\beta}}; \mathbf{x}_i) - \frac{1}{n}\sum_{j=1}^{n}\{Y_j - g(\hat{\boldsymbol{\beta}}; \mathbf{x}_j)\}.$$

(2) Generate i.i.d. bootstrap residuals $\varepsilon_1^*, \varepsilon_2^*, \ldots, \varepsilon_n^*$ from F_n.

(3) Calculate bootstrap observations

$$Y_i^* = g(\hat{\boldsymbol{\beta}}; \mathbf{x}_i) + \varepsilon_i^*, \quad i = 1, \ldots, n.$$

(4) Approximate the sampling distribution of $\hat{\boldsymbol{\beta}}$ by the bootstrap distribution of

$$\hat{\boldsymbol{\beta}}^* = \hat{\boldsymbol{\beta}}\left((Y_1^*, \mathbf{x}_1), \ldots, (Y_n^*, \mathbf{x}_n)\right).$$

Remarks:

(a) The above resampling scheme is called "bootstrapping residuals" or "resampling residuals" in the literature.

(b) If the covariates $\mathbf{x}_i$ are random, researchers often apply the so-called "pairs bootstrap", which is less dependent on the underlying model assumption than the bootstrap based on residuals: Bootstrap data $(Y_1^*, \mathbf{x}_1^*), \ldots, (Y_n^*, \mathbf{x}_n^*)$ are generated by simple random sampling with replacement from $(Y_1, \mathbf{x}_1), \ldots, (Y_n, \mathbf{x}_n)$ and the sampling distribution of $\hat{\boldsymbol{\beta}}$ is approximated by the bootstrap distribution of

$$\hat{\boldsymbol{\beta}}^* = \hat{\boldsymbol{\beta}}\left((Y_1^*, \mathbf{x}_1^*), \ldots, (Y_n^*, \mathbf{x}_n^*)\right).$$

(c) The case where the ε_i's are *heteroscedastic errors* will be discussed in Chapter 5.

2.9 The modified bootstrap

Bickel and Freedman (1981) provided counter-examples to show where the standard (naive) bootstrap fails (i.e., the bootstrap estimators are not first-order accurate). Examples include degenerate U-statistics, extreme order statistics and spacings of the observations.

Swanepoel (1986) showed how these counter-examples can be mended by introducing the "modified bootstrap" or "m-out-of-n bootstrap", as it is more popularly known.

For any random variable $R_n(\mathbf{X}_n; F)$, the modified bootstrap consists of approximating the sampling distribution of $R_n(\mathbf{X}_n; F)$ under F by the bootstrap distribution of $R_m(\mathbf{X}_m^*; F_n)$ under F_n, i.e.,

$$P^* \left(R_m(\mathbf{X}_m^*; F_n) \leq x\right) \approx P_F \left(R_n(\mathbf{X}_n; F) \leq x\right),$$

where $\mathbf{X}_m^* = (X_1^*, \ldots, X_m^*)$, for some suitable choice of the bootstrap sample size m.

Since 1986 several new cases have been reported, illustrating the failure of the naive bootstrap. However, in each case the "m-out-of-n bootstrap" led to consistent (first-order accurate) bootstrap estimators, emphasising the ongoing success of the methodology.

Some of the above-mentioned cases include applications of the naive bootstrap to:

- the mean in the infinite variance case (Athreya, 1987; Knight, 1989);
- extreme order statistics (Deheuvels et al., 1993);
- the Cramèr-von Mises goodness-of-fit test statistic with doubly censored data (Bickel and Ren, 1995);
- unstable first-order autoregressive processes (Basawa et al., 1991; Datta, 1996; Heimann and Kreiss, 1996);
- critical branching processes with immigration (Sriram, 1994);
- estimation of the distribution of the Studentized mean (Hall and LePage, 1996);
- a sample quantile when the density has a jump (Huang et al., 1996);
- confidence intervals for endpoints of a c.d.f. (Athreya and Fukuchi, 1997);
- the maximum of a stationary process (Athreya et al., 1999); and

Recent papers on the modified bootstrap caused a revival of this idea:

- Chung and Lee (2001) applied the modified bootstrap to correct coverage error in the construction of bootstrap confidence bounds. They showed that the coverage error of a standard bootstrap percentile method confidence bound, which is of order $O(n^{-1/2})$ typically, can be reduced to $O(n^{-1})$ by use of an optimal bootstrap sample size m in the modified bootstrap. The authors also conducted a simulation study to illustrate their findings, which suggest that the modified bootstrap method yields intervals of shorter length and greater stability compared to competitors of similar coverage accuracy (see also Lee, 1999).

- Janssen et al. (2001), and also Janssen et al. (2002), showed that, compared to the standard (naive) bootstrap, the modified bootstrap provides faster consistency rates for the bootstrap distributions of U-quantiles and Kaplan-Meier quantiles (comprehensive surveys of bootstrapping U-statistics and bootstrapping in survival analysis were written by Janssen (1997) and Veraverbeke (1997), respectively).

  The results of Chung and Lee (2001) and those of Janssen et al. (2001), and also Janssen et al. (2002), illustrate that the modified bootstrap is useful, not only in cases where the standard bootstrap fails, but also in situations where it is valid.

Remarks:

(a) It is well-known that the ordinary delete-1 jackknife fails (i.e., it is not asymptotically consistent) when estimating, for example, the variance of a sample quantile (Efron, 1979). It is known (see, e.g., Wu, 1986) that by carefully choosing d, a delete-d jackknife estimator overcomes some of the deficiencies of the ordinary jackknife.

(b) Unlike the delete-d jackknife, however, which suffers from a combinatoric explosion of computation with increasing d, the modified ("m-out-of-n") bootstrap is just the opposite. The smaller the resample size m, the easier it is to resample and to compute.

Further Remarks:

- The choice of m is crucial. In order to make the modified bootstrap accessible for the statistical practitioner, a data-dependent rule to choose m is very important.

- Research on this topic has only started recently. Data-based choices of m were proposed, for example, by Sakov (1998); Götze and Račkauskas (1999) and Sakov and Bickel (2000).

2.10 The smoothed bootstrap

The e.d.f. F_n is a discrete distribution function and this seems undesirable when dealing with a continuous c.d.f. F. Efron (1979) suggested the smoothed bootstrap:
Instead of resampling from F_n, resample from a smoothed version of F_n denoted by $\hat{F}_n$. This smoothed distribution function can be expressed as

$$\hat{F}_n = K_n * F_n,$$

where $K_n(\cdot)$ is a sequence of continuous distribution functions and $*$ denotes convolution. We have

$$\hat{F}_n(x) = \int_{-\infty}^{+\infty} K_n(x-y)dF_n(y).$$

$K_n(\cdot)$ is mostly a **kernel** sequence:

$$K_n(t) = K(t/h),$$

where K is a known continuous distribution function (e.g., $K = \Phi$) and $h = h_n$ is a bandwidth. In this case

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right).$$

A bootstrap sample $Y_1^*, \ldots, Y_n^*$ from $\hat{F}_n$ can be obtained in a simple way:
Let $X_1^*, \ldots, X_n^*$ be i.i.d. rv's with c.d.f. F_n and let $R_1, \ldots, R_n$ be i.i.d. rv's with c.d.f. K. If $(X_1^*, \ldots, X_n^*)$ and $(R_1, \ldots, R_n)$ are independent, we may take

$$Y_i^* = X_i^* + hR_i, \quad i = 1, \ldots, n.$$

Remarks:

(a) Romano (1988) showed that if a parameter θ, which can be viewed as a functional $T(f)$ of the density f, is to be estimated, the standard bootstrap can fail (i.e., it is inconsistent) unless the resampling is done from $\hat{F}_n$.

(b) Hall and DiCiccio (1989) showed that in estimating the variance of a sample quantile, the rate of convergence of the relative error can be improved by using a smoothed bootstrap instead of the standard bootstrap (see also the review paper by De Angelis and Young, 1992).

(c) The correct choice of the bandwidth h is crucial. In order to make the smoothed bootstrap more accessible for the statistical practitioner, a reliable data-dependent bandwidth is of the utmost importance.

(d) Polansky (2001) derived a bandwidth selector for the smoothed bootstrap applied to construct one-sided percentile confidence intervals.

2.11 Bootstrapping dependent data

A major development of bootstrap methods since the mid-1980's has been their application to dependent data. Two well-known methods relating to dependent data will now be discussed, namely the Moving Block Bootstrap and the Autoregressive Sieve Bootstrap. In Section 2.11.3 a comparison is made between these two procedures.
Throughout the discussion, we will assume that $X_1, X_2, \ldots$ is a sequence of strictly stationary random variables (or vectors).

2.11.1 The Moving Block Bootstrap(MBB)

The *MBB* was suggested by Künsch (1989) and is implemented as follows:

(1) Define blocks $\mathcal{B}_j = (X_j, \ldots, X_{j+\ell-1})$, for $j = 1, \ldots, N$, where $N = n - \ell + 1$ and $1 \leq \ell \leq n$ denotes the block size.

(2) Let $b = \lfloor n/\ell \rfloor$. Select a random sample $\mathcal{B}_1^*, \ldots, \mathcal{B}_b^*$ from $\{\mathcal{B}_1, \ldots, \mathcal{B}_N\}$.

(3) Arrange the components of $\mathcal{B}_1^*, \ldots, \mathcal{B}_b^*$ into a sequence.

(4) This yields $n_1 = b\ell$ bootstrap observations $\mathbf{X}_{n_1}^* = (X_1^*, X_2^*, \ldots, X_{n_1}^*)$. Note that $n_1/n \to 1$ as $n \to \infty$.

Remarks:

(a) It has been shown that the *MBB* is asymptotically valid for a wide range of statistics and a wide range of data generating models, as long as they are short-range dependent (Bühlmann, 1995; Bühlmann and Künsch, 1995).

(b) Götze and Künsch (1996) as well as Lahiri (1996) proved that the *MBB* applied to certain statistics is *second-order* accurate.

(c) The correct choice of the block length ℓ is crucial and requires careful consideration.

Data-based choice of block length

Hall et al. (1995) derived a simple rule for choosing ℓ data-dependently. They consider the performance of the MBB with different block lengths for subsamples of length $m < n$, yielding an optimal block length $\hat{\ell}_m$. The estimated optimal block length is then derived with a Richardson extrapolation adjusting to the original sample size n :

$$\hat{\ell}_n = (n/m)^{1/k}\hat{\ell}_m, \quad \text{where}$$

$k = 3$ when estimating bias or variance of $\hat{\theta}_n$, $k = 4$ for estimating the distribution function of $(\hat{\theta}_n - \theta)/\hat{\sigma}_n$ and $k = 5$ for estimating the distribution function of $\mid \hat{\theta}_n - \theta \mid /\hat{\sigma}_n$. The method described above is not fully data-driven, since m is another tuning constant. Moreover, the behaviour of bootstrapped nonlinear statistics $\hat{\theta}_n$ for small m is unsatisfactory.

When the *MBB* is applied to estimate the bias or standard error of a statistic, Bühlmann and Künsch (1999) propose a fully data-driven procedure for the selection of the block length ℓ. It is based on an equivalence of ℓ to the inverse of the bandwidth of a lag weight estimator of the spectral density at zero. The procedure can easily be implemented and performs at least as well as the procedure of Hall et al. (1995).

2.11.2 The Autoregressive Sieve Bootstrap

The Autoregressive Sieve Bootstrap method was originally proposed by Swanepoel and Van Wyk (1986). We will now provide a short discussion on this method:

Let $\{X_j, -\infty < j < \infty\}$ denote a strictly stationary, invertible linear time series:

$$X_j = \mu + \sum_{i=0}^{\infty} \alpha_i \varepsilon_{j-i} \ , \tag{2.1}$$

for constants μ, α_i and i.i.d. rv's $\{\varepsilon_i\}$ with $E(\varepsilon_i) = 0$, $i = 1, \ldots, n$.

Inverting (2.1), we obtain an $AR(\infty)$-process

$$X_j - \mu = \sum_{i=1}^{\infty} \beta_i (X_{j-i} - \mu) + \varepsilon_j, \tag{2.2}$$

for constants β_i.

The basic idea of the $AR(\infty)$-sieve bootstrap is to approximate the $AR(\infty)$-model in (2.2) with a $AR(p)$ model:

$$X_j - \mu = \sum_{i=1}^{p} \beta_i (X_{j-i} - \mu) + \varepsilon_j.$$

Choose an estimate $\hat{p}$ of p by, e.g., the *AIC* model selection procedure with Gaussian innovations (Shibata, 1980 has shown optimality of the AIC for prediction in $AR(\infty)$-models). Some researchers recommend the use of the *AICC* criterion, which is a bias-corrected version of *AIC* (Hurvich and Tsai, 1989). One then proceeds further by applying classical "resampling residuals." (See also Bühlmann, 1995, 1997, 1998; Bickel and Bühlmann, 1999; Choi and Hall, 2000.)

2.11.3 Comparison between the MBB and the $AR(\infty)$-sieve bootstrap

- The $AR(\infty)$-sieve bootstrap yields bootstrap pseudo-data $\mathbf{X}_n^*$ that are conditionally (given $\mathbf{X}_n$) stationary.
- Sieve bootstrap samples do not exhibit any of the artefacts that typically appear in series generated by the *MBB*, as the result of joining together randomly selected blocks of data.
- Since the $AR(\infty)$-sieve bootstrap does not corrupt second-order properties, it may be used in a double-bootstrap form and potentially leads to higher-order accuracy. For example, the $AR(\infty)$-sieve double bootstrap can be employed to calibrate a basic percentile method confidence interval. This gives second-order accuracy, without requiring variance estimation of the underlying statistic (Choi and Hall, 2000). Moving block double bootstrapping does not seem promising since the block bootstrap in the first iteration corrupts dependence of the data where blocks join (Bühlmann, 2002).
- The $AR(\infty)$-sieve bootstrap adapts to the degree of dependence: its accuracy improves as the degree of dependence decreases. This is not the case with the block bootstrap (Bühlmann, 2002).
- Finally, empirical results of many authors in the literature show that the $AR(\infty)$-sieve bootstrap seems generally less sensitive to selection of a model in the sieve (i.e., the choice of p) than the blockwise bootstraps to the block length ℓ.

2.12 Further topics

In recent years the bootstrap has become an active and broad topic for research and application. Significant research areas where the bootstrap found extensive application (which have not been dealt with in this chapter) are, among others, efficient bootstrap simulations, survey sampling, nonparametric curve estimation, sequential analysis, directional data, categorical data, Bayesian inference, discriminant analysis and nonparametric autoregression.

In the next chapter the application of the bootstrap to hypothesis testing will be discussed.

Chapter 3

Two methods to apply the bootstrap to hypothesis testing

In this chapter we will discuss two methods to apply the bootstrap to hypothesis testing. The first method involves transforming the data in order to "mimic" H_0 (or, in the case of power calculations, H_A). The second method involves keeping the data values fixed, but instead changing the probabilities on each data value in order to conform to H_0 or H_A.

3.1 Introduction

In Section 2.6 a brief overview of bootstrap confidence intervals was provided. The construction of bootstrap confidence intervals are linked to the execution of bootstrap hypothesis tests because of the duality between confidence intervals and hypothesis tests, i.e., the null hypothesis is rejected if the hypothesized value under the null hypothesis lies outside the confidence interval. Shao and Tu (1995), however, provided a few reasons why it is important to consider bootstrap hypothesis testing seperately:

"Firstly, finding a test directly is much easier than getting a test through constructing a confidence interval, which is impossible in some cases. Secondly, the test obtained directly may be better since they usually take account of the special nature of the hypothesis."

MacKinnon (2002) also argues that there are many more ways to construct bootstrap confidence intervals (e.g., percentile intervals, bootstrap-t intervals, etc.) than there are to perform bootstrap-based tests. As a result, these alternative ways to compute bootstrap confidence intervals may lead to different results and can be confusing. Hall and Wilson (1991) highlighted two important guidelines for bootstrap hypothesis testing. Their first guideline states that, when one wants to estimate the critical value, resampling must be done in a way that reflects the null hypothesis. This must be done even if the data were generated from a distribution specified by the alternative hypothesis. This guideline is crucial to the success of bootstrap hypothesis testing and have been mentioned by, among others, Young (1986); Beran (1988); Hinkley (1988); Fisher and Hall (1990); Westfall and Young (1993) and most recently by Martin (2007). In some tests, however, it is not so easy to "mimic" the null hypothesis when resampling, and careful thought should be given as to how this resampling might take place. In their second guideline, Hall and Wilson (1991) recommended that bootstrap hypothesis tests should be based on test statistics that are (asymptotically) pivotal. The importance of using pivotal statistics in the bootstrap was considered by, among others, Beran (1987, 1988) and Hall (1992).

Although there is a very large literature on bootstrapping in Statistics, only a small proportion of it is devoted to bootstrap-based testing. The focus, however, is usually on estimating bootstrap standard errors and constructing bootstrap confidence intervals. The books by Westfall and Young (1993); Efron and Tibshirani (1993) and Davison and Hinkley (1997) cover bootstrap testing in some detail. Hypothesis testing based on the bootstrap has also been discussed by several authors in Econometrics (see, e.g., Horowitz, 2001; MacKinnon, 2002; Park, 2003; Davidson and MacKinnon, 2004).

In a very recent article, Martin (2007) investigated bootstrap hypothesis testing for some common statistical problems. One of the key things that he remarked on was that the bootstrap estimate of power has not been a focus of any previous studies of bootstrap testing. In order to use the bootstrap to estimate the power of a test, at a specific alternative, resampling must be done in a way that reflects the alternative hypothesis.

Two methods which can be used to apply the bootstrap to hypothesis testing will be discussed in this chapter. These methods will be referred to as the *transformation method* and the *exponentially tilted version of the e.d.f.* A general formulation for the transformation method will be provided, as well as an algorithm for obtaining the bootstrap critical value and another algorithm for using the bootstrap to estimate the power of a test at a specific alternative. The chapter concludes with an overview of the exponentially tilted version of the e.d.f.

3.2 Transformation method

3.2.1 General formulation

Let $\mathbf{X}_n = (X_1, X_2, \ldots, X_n)$ be a random sample from some unknown distribution F.
Consider, without loss of generality, the right-sided alternative hypothesis:

$$H_0 : \theta(F) = \theta_0 \quad \text{vs.} \quad H_A : \theta(F) > \theta_0, \tag{3.1}$$

where the parameter $\theta(F)$ is some functional of F.
The test rejects H_0 if and only if

$$T_n(\mathbf{X}_n) \geq C_n(\alpha), \text{ where}$$

$$P_{H_0}(T_n(\mathbf{X}_n) \geq C_n(\alpha)) \cong \alpha. \tag{3.2}$$

$T_n(\mathbf{X}_n)$ is an appropriate test statistic, $C_n(\alpha)$ is the critical value and α is the nominal significance level of the test.

Remarks:

(a) The critical value $C_n(\alpha)$ is unknown, since F is unknown.

(b) We wish to estimate $C_n(\alpha)$ by the bootstrap estimator $C_n(\alpha; \mathbf{X}_n)$, defined in (3.3).

When the bootstrap is applied, the following bootstrap sample is obtained: $\mathbf{X}_n^* = (X_1^*, X_2^*, \ldots, X_n^*)$, where the components of $\mathbf{X}_n^*$ are i.i.d. drawn from F_n, the e.d.f. of $\mathbf{X}_n$. In the bootstrap world (for fixed $\mathbf{X}_n$) we would like to have that $\theta(F_n) = \theta_0$. However, this is seldom, if ever the case, hence we need to transform $X_1, X_2, \ldots, X_n$.
Denote the transformed variables by

$$V_i^0 = V_i^0(\mathbf{X}_n; \theta_0), \quad i = 1, 2, \ldots, n.$$

The bootstrap random sample is now given by $\mathbf{V}_n^{0*} = (V_1^{0*}, V_2^{0*}, \ldots, V_n^{0*})$ drawn from G_n, the e.d.f. of $\mathbf{V}_n^0 = (V_1^0, V_2^0, \ldots, V_n^0)$. The transformed variables $V_i^0(\mathbf{X}_n; \theta_0)$, $\;i = 1, 2, \ldots, n$, are chosen such that $\theta(G_n) = \theta_0$, in the bootstrap world.
The bootstrap estimator $C_n(\alpha; \mathbf{X}_n)$ is now chosen such that (see (3.2))

$$P^*_{H^*_0}\left(T_n(\mathbf{V}^{0*}_n) \geq C_n(\alpha; \mathbf{X}_n)\right) \cong \alpha. \tag{3.3}$$

The critical value $C_n(\alpha; \mathbf{X}_n)$ can be approximated by $\hat{C}_n(\alpha; \mathbf{X}_n)$ using the following Monte-Carlo algorithm:

i) Obtain your first bootstrap sample $\mathbf{V}^{0*}_n$ and calculate $T_n(\mathbf{V}^{0*}_n)$. Denote this by T^*_1.

ii) Independently repeat step i) a number B times to obtain B bootstrap replications $T^*_1, T^*_2, \ldots, T^*_B$.

iii) Obtain the order statistics $T^*_{(1)} \leq T^*_{(2)} \leq \cdots \leq T^*_{(B)}$.

iv) $\hat{C}_n(\alpha; \mathbf{X}_n) = T^*_{(\lfloor B(1-\alpha) \rfloor)}$.

The bootstrap p-value is given by

$$P_{boot} = P^*_{H^*_0}(T_n(\mathbf{V}^{0*}_n) \geq T_n(\mathbf{X}_n)),$$

which can be approximated by

$$\hat{P}_{boot} = \frac{1}{B} \sum_{b=1}^{B} I(T^*_b \geq T_n(\mathbf{X}_n)). \tag{3.4}$$

Bootstrap estimate of power

Again, consider testing the hypothesis stated in (3.1):

$$H_0 : \theta(F) = \theta_0 \quad \text{vs.} \quad H_A : \theta(F) > \theta_0.$$

The following procedure can be used to estimate the power of the test (by making use of the bootstrap), at a specific alternative $H_A : \theta(F) = \theta_A$:

a) Obtain $C_n(\alpha; \mathbf{X}_n)$, as described above.

b) Since $\theta(F_n)$ is hardly ever equal to θ_A (in the bootstrap world), we need to transform $\{X_i,\ i = 1, 2, \ldots, n\}$.

   Denote the transformed variables by

   $$V^A_i = V^A_i(\mathbf{X}_n; \theta_A),\ i = 1, 2, \ldots, n.$$

   The bootstrap random sample is given by $\mathbf{V}^{A*}_n = (V^{A*}_1, V^{A*}_2, \ldots, V^{A*}_n)$ obtained from H_n, the e.d.f. of $\mathbf{V}^A_n = (V^A_1, V^A_2, \ldots, V^A_n)$. The transformed variables $V^A_i(\mathbf{X}_n; \theta_A),\ i = 1, 2, \ldots, n$, are chosen such that $\theta(H_n) = \theta_A$, in the bootstrap world.

c) The estimated power of the test, at the specific alternative, is then given by

$$P^A_{boot} = P^*_{H^*_A}(T_n(\mathbf{V}^{A*}_n) \geq C_n(\alpha; \mathbf{X}_n)).$$

It is possible to approximate P^A_{boot} by $\hat{P}^A_{boot}$ using the following Monte-Carlo algorithm:

i) Calculate $\hat{C}_n(\alpha; \mathbf{X}_n)$, as previously discussed.

ii) Obtain your first bootstrap sample $\mathbf{V}^{A*}_n$ and calculate $T_n(\mathbf{V}^{A*}_n)$. Denote this by T^*_1.

iii) Independently repeat step ii) a number B times to obtain B bootstrap replications $T^*_1, T^*_2, \ldots, T^*_B$.

iv) $\hat{P}^A_{boot} = \frac{1}{B}\sum_{b=1}^{B} I(T^*_b \geq \hat{C}_n(\alpha; \mathbf{X}_n))$.

In the next section a number of different hypothesis tests will be considered, and techniques used to apply the transformation method in order to test these hypotheses will be discussed. We also describe how to estimate the power of these tests at a specific alternative.

3.2.2 Transformation method applied to various statistical tests

(a) The mean in the univariate case

Let $\mathbf{X}_n = (X_1, X_2, \ldots, X_n)$ denote a random sample from an unknown univariate distribution F with finite mean μ.

Suppose we wish to test the hypothesis:

$$H_0 : \mu = \mu_0 \quad \text{vs.} \quad H_A : \mu > \mu_0.$$

Application of the bootstrap for testing this hypothesis has been considered by a number of authors, including Young (1988); Noreen (1989); Hall and Wilson (1991) and Efron and Tibshirani (1993). One usually applies the (asymptotic) pivotal test statistic

$$T_n(\mathbf{X}_n) = \frac{\sqrt{n}(\overline{X}_n - \mu_0)}{S_n(\mathbf{X}_n)}, \tag{3.5}$$

where

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \text{ and } S^2_n(\mathbf{X}_n) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2.$$

The test rejects H_0 if and only if

$$T_n(\mathbf{X}_n) \geq C_n(\alpha), \text{ where}$$

$$P_{H_0}(T_n(\mathbf{X}_n) \geq C_n(\alpha)) \cong \alpha.$$

Consider the following transformation of the data $\{X_i,\ i = 1, 2, \ldots, n\}$:

$$V_i^0 = X_i - \overline{X}_n + \mu_0,\ i = 1, 2, \ldots, n.$$

The sample mean of $\{V_i^0,\ i = 1, 2, \ldots, n\}$ is now μ_0. Choose $C_{n,P}^R(\alpha; \mathbf{X}_n)$, the bootstrap estimator of $C_n(\alpha)$, such that

$$P_{H_0^*}^*\left(\frac{\sqrt{n}(\overline{V}_n^{0*} - \mu_0)}{S_n(\mathbf{V}_n^{0*})} \geq C_{n,P}^R(\alpha; \mathbf{X}_n)\right) \cong \alpha,$$

where $\overline{V}_n^{0*} = \frac{1}{n}\sum_{i=1}^{n} V_i^{0*}$ and $S_n^2(\mathbf{V}_n^{0*}) = \frac{1}{n}\sum_{i=1}^{n}(V_i^{0*} - \overline{V}_n^{0*})^2$.

Because of the transformation, $C_{n,P}^R(\alpha; \mathbf{X}_n)$ is therefore given by

$$P_{H_0^*}^*\left(\frac{\sqrt{n}(\overline{X}_n^* - \overline{X}_n)}{S_n(\mathbf{X}_n^*)} \geq C_{n,P}^R(\alpha; \mathbf{X}_n)\right) \cong \alpha, \tag{3.6}$$

where $\overline{X}_n^* = \frac{1}{n}\sum_{i=1}^{n} X_i^*$ and $S_n^2(\mathbf{X}_n^*) = \frac{1}{n}\sum_{i=1}^{n}(X_i^* - \overline{X}_n^*)^2$.

Remarks:

(a) If the data are not transformed to have mean μ_0, then (3.6) becomes

$$P_{H_0^*}^*\left(\frac{\sqrt{n}(\overline{X}_n^* - \mu_0)}{S_n(\mathbf{X}_n^*)} \geq C_{n,P}^W(\alpha; \mathbf{X}_n)\right) \cong \alpha.$$

(b) The symbol R in the notation of $C_{n,P}^R(\alpha; \mathbf{X}_n)$ is used to denote the critical value obtained when the bootstrap is applied "correctly", whereas the W in the notation of $C_{n,P}^W(\alpha; \mathbf{X}_n)$ is used to indicate the critical value derived when the bootstrap is applied "wrongly". The subscript P in both these critical values refers to the fact that it is a pivotal (or asymptotically pivotal) statistic being used. We will elaborate on these concepts later on in the text.

(c) In Chapter 4 some properties of these two estimators ($C_{n,P}^R(\alpha; \mathbf{X}_n)$ and $C_{n,P}^W(\alpha; \mathbf{X}_n)$) as well as their non-pivotal counterparts $C_{n,N-P}^R(\alpha; \mathbf{X}_n)$ and $C_{n,N-P}^W(\alpha; \mathbf{X}_n)$ will be discussed.

In order to estimate the power of the test, at a specific alternative μ_A, transform the data as follows:

$$V_i^A = X_i - \overline{X}_n + \mu_A,\ i = 1, 2, \ldots, n.$$

By construction, the sample mean of $\{V_i^A,\ i = 1, 2, \ldots, n\}$ is μ_A.

The estimated power of this test, at the specific alternative, is then given by

$$P^A_{boot} = P^*_{H^*_A}\left(\frac{\sqrt{n}(\overline{V}^{A*}_n - \mu_0)}{S_n(\mathbf{V}^{A*}_n)} \geq C^R_{n,P}(\alpha;\mathbf{X}_n)\right),$$

where $\overline{V}^{A*}_n = \frac{1}{n}\sum_{i=1}^{n} V^{A*}_i$ and $S^2_n(\mathbf{V}^{A*}_n) = \frac{1}{n}\sum_{i=1}^{n}\left(V^{A*}_i - \overline{V}^{A*}_n\right)^2$.

Using the transformation, the estimated power is therefore given by

$$P^A_{boot} = P^*_{H^*_A}\left(\frac{\sqrt{n}(\overline{X}^*_n - \overline{X}_n - \mu_0 + \mu_A)}{S_n(\mathbf{X}^*_n)} \geq C^R_{n,P}(\alpha;\mathbf{X}_n)\right). \tag{3.7}$$

Chapter 8 presents the results of a Monte-Carlo study where we compare the bootstrap estimate of power with the Monte-Carlo estimate of the true power.

(b) Comparing two means

Let $\mathbf{X}_n = (X_1, X_2, \ldots, X_n)$ and $\mathbf{Y}_m = (Y_1, Y_2, \ldots, Y_m)$ be independent random samples from unknown distributions F and G respectively, with sample means $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $\overline{Y}_m = \frac{1}{m}\sum_{j=1}^{m} Y_j$ and sample variances $S^2_n(\mathbf{X}_n) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$ and $S^2_m(\mathbf{Y}_m) = \frac{1}{m}\sum_{j=1}^{m}(Y_j - \overline{Y}_m)^2$.
Consider testing the hypothesis

$$H_0 : \mu_x = \mu_y \quad \text{vs.} \quad H_A : \mu_x > \mu_y,$$

where μ_x is the mean of F and μ_y is the mean of G.
Application of the bootstrap for testing this hypothesis was considered in some detail by Efron and Tibshirani (1993); Westfall and Young (1993); Davison and Hinkley (1997) and more recently by Martin (2007).
If equal variances are assumed, one can use the following test statistic:

$$T^{(1)}_{n,m}(\mathbf{X}_n, \mathbf{Y}_m) = \frac{\overline{X}_n - \overline{Y}_m}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}}, \tag{3.8}$$

where

$$S_p = \sqrt{\frac{nS^2_n(\mathbf{X}_n) + mS^2_m(\mathbf{Y}_m)}{n+m-2}}.$$

Further, if it is assumed that F and G are normally distributed, then $T^{(1)}_{n,m}(\mathbf{X}_n, \mathbf{Y}_m)$ has a t_{n+m-2} distribution under the null hypothesis.

If we do not assume that the variances in the two populations are equal, we could base the test on the test statistic

$$T^{(2)}_{n,m}(\mathbf{X}_n, \mathbf{Y}_m) = \frac{\overline{X}_n - \overline{Y}_m}{\sqrt{\frac{S^2_n(\mathbf{X}_n)}{n} + \frac{S^2_m(\mathbf{Y}_m)}{m}}}.$$

However, even if F and G are assumed to be normally distributed, $T^{(2)}_{n,m}(\mathbf{X}_n, \mathbf{Y}_m)$ no longer has a t-distribution. In the literature this is known as the Behrens-Fisher problem.

For the purpose of this discussion the statistic $T^{(2)}_{n,m}(\mathbf{X}_n, \mathbf{Y}_m)$ will be deemed the appropriate test statistic.

The test rejects H_0 if and only if

$$T^{(2)}_{n,m}(\mathbf{X}_n, \mathbf{Y}_m) \geq C_n(\alpha), \text{ where}$$

$$P_{H_0}(T^{(2)}_{n,m}(\mathbf{X}_n, \mathbf{Y}_m) \geq C_n(\alpha)) \cong \alpha.$$

Consider the following transformations of $\{X_i, i = 1, 2, \ldots, n\}$ and $\{Y_j, j = 1, 2, \ldots, m\}$:

$$\begin{aligned} V_i^{x0} &= X_i - \overline{X}_n \ , i = 1, 2, \ldots, n \\ V_j^{y0} &= Y_j - \overline{Y}_m \ , j = 1, 2, \ldots, m. \end{aligned}$$

By construction both $\{V_i^{x0},\ i = 1, 2, \ldots, n\}$ and $\{V_j^{y0},\ j = 1, 2, \ldots, m\}$ have sample means equal to 0.

Remarks:

(a) There are numerous other ways to transform the X_i's and Y_j's to "mimic" H_0.

Efron and Tibshirani (1993) proposed the following transformations:

$$\begin{aligned} V_i^{x0} &= X_i - \overline{X}_n + \overline{Z} \ , i = 1, 2, \ldots, n \\ V_j^{y0} &= Y_j - \overline{Y}_m + \overline{Z} \ , j = 1, 2, \ldots, m, \end{aligned}$$

where $\overline{Z}$ is the mean of the combined sample, i.e., $\overline{Z} = \dfrac{\sum_{i=1}^{n} X_i + \sum_{j=1}^{m} Y_j}{m+n}$.

The sample means of $\{V_i^{x0},\ i = 1, 2, \ldots, n\}$ and $\{V_j^{y0},\ j = 1, 2, \ldots, m\}$ are now equal to $\overline{Z}$.

(b) Another possible transformation proposed by Martin (2007) is:

$$\begin{aligned} V_i^{x0} &= X_i, \ i = 1, 2, \ldots, n \\ V_j^{y0} &= Y_j - \overline{Y}_m + \overline{X}_n, \ j = 1, 2, \ldots, m. \end{aligned}$$

The sample means of $\{V_i^{x0},\ i = 1, 2, \ldots, n\}$ and $\{V_j^{y0},\ j = 1, 2, \ldots, m\}$ are now equal to $\overline{X}_n$.

(c) The choice of transformation is not important, as long as both $\{V_i^{x0},\ i = 1, 2, \ldots, n\}$ and $\{V_j^{y0},\ j = 1, 2, \ldots, m\}$ have *equal* sample means, as required by the null hypothesis.

(d) In the case of the transformation proposed by Efron and Tibshirani (1993), the combined sample mean $\overline{Z}$ will disappear in the test statistic. The same will happen with the quantity $\overline{X}_n$ in the transformation proposed by Martin (2007).

Define $C_n(\alpha; \mathbf{X}_n, \mathbf{Y}_m)$, the bootstrap estimator of $C_n(\alpha)$, by

$$P^*_{H_0^*}\left(\frac{\overline{V}_n^{x0*} - \overline{V}_m^{y0*}}{\sqrt{\frac{S_n^2(\mathbf{V}_n^{x0*})}{n} + \frac{S_m^2(\mathbf{V}_m^{y0*})}{m}}} \geq C_n(\alpha; \mathbf{X}_n, \mathbf{Y}_m)\right) \cong \alpha,$$

where

$\mathbf{V}_n^{x0*} = (V_1^{x0*}, V_2^{x0*}, \ldots, V_n^{x0*})$ and the components of $\mathbf{V}_n^{x0*}$ are i.i.d. drawn from the e.d.f. of $\mathbf{V}_n^{x0} = (V_1^{x0}, V_2^{x0}, \ldots, V_n^{x0})$; $\mathbf{V}_m^{y0*} = (V_1^{y0*}, V_2^{y0*}, \ldots, V_m^{y0*})$ and the components of $\mathbf{V}_m^{y0*}$ are i.i.d. drawn from the e.d.f. of $\mathbf{V}_m^{y0} = (V_1^{y0}, V_2^{y0}, \ldots, V_m^{y0})$; $\overline{V}_n^{x0*} = \frac{1}{n}\sum_{i=1}^{n} V_i^{x0*}$; $\overline{V}_m^{y0*} = \frac{1}{m}\sum_{j=1}^{m} V_j^{y0*}$;

$$S_n^2(\mathbf{V}_n^{x0*}) = \frac{1}{n}\sum_{i=1}^{n}(V_i^{x0*} - \overline{V}_n^{x0*})^2 \quad \text{and} \quad S_m^2(\mathbf{V}_m^{y0*}) = \frac{1}{m}\sum_{j=1}^{m}(V_j^{y0*} - \overline{V}_m^{y0*})^2.$$

Using the transformations above, $C_n(\alpha; \mathbf{X}_n, \mathbf{Y}_m)$ is therefore chosen such that

$$P^*_{H_0^*}\left(\frac{\overline{X}_n^* - \overline{Y}_m^* - (\overline{X}_n - \overline{Y}_m)}{\sqrt{\frac{S_n^2(\mathbf{X}_n^*)}{n} + \frac{S_m^2(\mathbf{Y}_m^*)}{m}}} \geq C_n(\alpha; \mathbf{X}_n, \mathbf{Y}_m)\right) \cong \alpha,$$

where $\mathbf{X}_n^* = (X_1^*, X_2^*, \ldots, X_n^*)$ and the components of $\mathbf{X}_n^*$ are i.i.d. obtained from the e.d.f. of $\mathbf{X}_n$; $\mathbf{Y}_m^* = (Y_1^*, X_2^*, \ldots, Y_m^*)$ and the components of $\mathbf{Y}_m^*$ are i.i.d. obtained from the e.d.f. of $\mathbf{Y}_m$; $\overline{X}_n^* = \frac{1}{n}\sum_{i=1}^{n} X_i^*$; $\overline{Y}_m^* = \frac{1}{m}\sum_{j=1}^{m} Y_j^*$; $S_n^2(\mathbf{X}_n^*) = \frac{1}{n}\sum_{i=1}^{n}(X_i^* - \overline{X}_n^*)^2$ and $S_m^2(\mathbf{Y}_m^*) = \frac{1}{m}\sum_{j=1}^{m}(Y_j^* - \overline{Y}_m^*)^2$.

Next, consider estimating the power of this test, at a specific alternative, say $\mu_x - \mu_y = \delta_A$. Martin (2007) described one possible approach that can be used to do this. Consider the following transformed data:

$$\begin{aligned} V_i^{xA} &= X_i - \overline{X}_n + \delta_A,\ i = 1, 2, \ldots, n \\ V_j^{yA} &= Y_j - \overline{Y}_m,\ j = 1, 2, \ldots, m. \end{aligned}$$

The sample mean of $\{V_i^{xA},\ i = 1,2,\dots,n\}$ is equal to δ_A, while the sample mean of $\{V_j^{yA},\ j = 1,2,\dots,m\}$ is equal to 0.

The estimated power of the test, at the specific alternative $\mu_x - \mu_y = \delta_A$, is given by

$$P_{boot}^A = P_{H_A^*}^*\left(\frac{\overline{V}_n^{xA*} - \overline{V}_m^{yA*}}{\sqrt{\frac{S_n^2(\mathbf{V}_n^{xA*})}{n} + \frac{S_m^2(\mathbf{V}^{yA*})}{m}}} \geq C_n(\alpha; \mathbf{X}_n, \mathbf{Y}_m)\right),$$

where

$\mathbf{V}_n^{xA*} = (V_1^{xA*}, V_2^{xA*}, \dots, V_n^{xA*})$ and the components of $\mathbf{V}_n^{xA*}$ are i.i.d. drawn from the e.d.f. of $\mathbf{V}_n^{xA} = (V_1^{xA}, V_2^{xA}, \dots, V_n^{xA})$; $\mathbf{V}_m^{yA*} = (V_1^{yA*}, V_2^{yA*}, \dots, V_m^{yA*})$ and the components of $\mathbf{V}_m^{yA*}$ are i.i.d. drawn from the e.d.f. of $\mathbf{V}_m^{yA} = (V_1^{yA}, V_2^{yA}, \dots, V_m^{yA})$; $\overline{V}_n^{xA*} = \frac{1}{n}\sum_{i=1}^{n} V_i^{xA*}$; $\overline{V}_m^{yA*} = \frac{1}{m}\sum_{j=1}^{m} V_j^{yA*}$; $S_n^2(\mathbf{V}_n^{xA*}) = \frac{1}{n}\sum_{i=1}^{n}(V_i^{xA*} - \overline{V}_n^{xA*})^2$ and $S_m^2(\mathbf{V}_m^{yA*}) = \frac{1}{m}\sum_{j=1}^{m}(V_j^{yA*} - \overline{V}_m^{yA*})^2$.

Using the transformations, the estimated power of the test is therefore given by

$$P_{boot}^A = P_{H_A^*}^*\left(\frac{\overline{X}_n^* - \overline{Y}_m^* - (\overline{X}_n - \overline{Y}_m) + \delta_A}{\sqrt{\frac{S_n^2(\mathbf{X}_n^*)}{n} + \frac{S_m^2(\mathbf{Y}_m^*)}{m}}} \geq C_n(\alpha; \mathbf{X}_n, \mathbf{Y}_m)\right).$$

(c) One-way ANOVA models

Assume that a random sample $\{Y_{ij}, i = 1,2,\dots,r; j = 1,2,\dots,n_i\}$ is available, where r is the number of treatments and n_i is the number of responses within the ith treatment group. Thus, Y_{ij} is the value of the response variable in the jth trial for the ith factor level or treatment.

The one-way ANOVA model can be stated as follows:

$$Y_{ij} = \mu_i + \varepsilon_{ij},\ i = 1,2,\dots,r; j = 1,2,\dots,n_i,$$

where μ_i is the mean response for the ith treatment and ε_{ij} are i.i.d. with unknown distribution $F_i,\ i = 1,2,\dots r$, with zero mean.

An alternative, but entirely equivalent, formulation of the single-factor ANOVA model is:

$$Y_{ij} = \mu_. + \tau_i + \varepsilon_{ij},\ i = 1,2,\dots,r; j = 1,2,\dots,n_i,$$

where $\mu_. = \sum_{i=1}^{r} \mu_i/r$, $\tau_i = \mu_i - \mu_.$ is the ith factor level effect and $\varepsilon_{ij},\ j = 1,2,\dots,n_i$ are i.i.d. with unknown distribution $F_i, i = 1,2,\dots,r$, with zero mean.

The first model is called the *cell means* model, where as the second model is known as the *factor effects* model.

Normally, one would begin a single-factor study by determining whether or not the factor level means μ_i are equal or, equivalently, whether all the factor effects are zero.
Thus, for the cell means model interest centers around testing the hypothesis

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_r \quad \text{vs.} \quad H_A : \text{not all } \mu_i \text{ are equal,}$$

or, for the factor effects model,

$$H_0 : \tau_1 = \tau_2 = \cdots = \tau_r = 0 \quad \text{vs.} \quad H_A : \text{not all } \tau_i \text{ equal zero.}$$

Bootstrap-based testing in ANOVA models have been considered in some detail by Fisher and Hall (1990); Westfall and Young (1993) and recently by Martin (2007).
Fisher and Hall (1990) considered two possible test statistics, namely:

$$T_1(\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_r) = \frac{(n_T - r)\sum_{i=1}^{r} n_i(\overline{Y}_{i.} - \overline{Y}_{..})^2}{(r-1)\sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{i.})^2} \quad \text{and}$$

$$T_2(\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_r) = \sum_{i=1}^{r}\left[\frac{n_i(n_i-1)(\overline{Y}_{i.} - \overline{Y}_{..})^2}{\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{i.})^2}\right],$$

where

$$n_T = \sum_{i=1}^{r} n_i, \quad \overline{Y}_{i.} = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij},$$

$$\overline{Y}_{..} = \frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{n_i} Y_{ij} \quad \text{and} \quad \mathbf{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{in_i}).$$

The statistic T_1 is the classic F-ratio introduced by Fisher, while T_2 is a statistic proposed by James (1951).

Remarks:

(a) When H_0 holds and it is assumed that each F_i, $i = 1, 2, \ldots, r$, are normally distributed with a constant variance σ^2, then T_1 has a $F_{r-1, n_T - r}$ distribution. However, if $F_i \sim N(0, \sigma_i^2)$, T_1 has a distribution depending on the σ_i's in a complex manner. In this situation the distribution of T_2, however, does not depend on the unknown σ_i's (see Fisher and Hall, 1990).

(b) If the assumption that F_i is normally distributed is dropped, but one still allows for heteroscedasticity, the distribution of T_2 converges to a χ^2_{r-2} distribution, while the limiting distribution of T_1 does depend on the σ_i's (see James, 1951; Fisher and Hall, 1990). Thus, T_2 is an asymptotically pivotal statistic in the heteroscedasticity scenario, but T_1 is not.

(c) In the homoscedastic case it is recommended that T_1 should be used as the test statistic, since the use of T_2 in homoscedastic cases leads to a less powerful test (James, 1951; Beran, 1988).

We will first discuss a bootstrap-based test for the heteroscedastic scenario and then discuss a modification of the test for the homoscedastic scenario.

Heteroscedastic scenario

The test rejects H_0 if and only if

$$T_2(\mathbf{Y}_1, \mathbf{Y}_2, \dots \mathbf{Y}_r) \geq C_n(\alpha), \text{ where}$$

$$P_{H_0}(T_2(\mathbf{Y}_1, \mathbf{Y}_2, \dots \mathbf{Y}_r) \geq C_n(\alpha)) \cong \alpha.$$

Consider the following transformation of $\{Y_{ij}, i = 1, 2, \dots, r; j = 1, 2, \dots, n_i\}$ proposed by Fisher and Hall (1990):

$$V^0_{ij} = Y_{ij} - \overline{Y}_{i.}\ , \ i = 1, 2, \dots, r; j = 1, 2, \dots n_i.$$

The sample mean of each $\{V^0_{ij}, j = 1, 2, \dots, n_i\}$, $i = 1, 2, \dots r$, is now equal to 0.

Remark:

Martin (2007) proposed the transformation $V^0_{ij} = Y_{ij} - \overline{Y}_{i.} + \overline{Y}_{..}$, $i = 1, 2, \dots r; j = 1, 2, \dots, n_i$. With this transformation, the sample mean of each $\{V^0_{ij},\ j = 1, 2, \dots, n_i\}$, $i = 1, 2, \dots r$, is equal to $\overline{Y}_{..}$. This value will be cancelled out in the test statistic.

Choose $C^{(2)}_n(\alpha; \mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_r)$, the bootstrap estimator of $C_n(\alpha)$, such that

$$P^*_{H^*_0}\left(\sum_{i=1}^{r} \frac{n_i(n_i-1)(\overline{V}^{0*}_{i.} - \overline{V}^{0*}_{..})^2}{\sum\limits_{j=1}^{n_i}(V^{0*}_{ij} - \overline{V}^{0*}_{i.})^2} \geq C^{(2)}_n(\alpha; \mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_r)\right) \cong \alpha,$$

where

$V_{i1}^{0*}, V_{i2}^{0*}, \ldots, V_{in_i}^{0*}$ are i.i.d obtained from the e.d.f. of $V_{i1}^{0}, V_{i2}^{0}, \ldots, V_{in_i}^{0}$, $i = 1, 2, \ldots, r$;
$\overline{V}_{i.}^{0*} = \frac{1}{n_i}\sum_{j=1}^{n_i} V_{ij}^{0*}$ and $\overline{V}_{..}^{0*} = \frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{n_i} V_{ij}^{0*}$.
The homoscedastic scenario will now be considered.

Homoscedastic scenario

In this case the test rejects H_0 if and only if

$$T_1(\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_r) \geq C_n(\alpha), \text{ where}$$

$$P_{H_0}(T_1(\mathbf{Y}_1, \mathbf{Y}_2, \ldots \mathbf{Y}_r) \geq C_n(\alpha)) \cong \alpha.$$

Consider the following transformation of $\{Y_{ij},\ i = 1, 2, \ldots, r; j = 1, 2, \ldots, n_i\}$:

$$V_{ij}^0 = \frac{Y_{ij} - \overline{Y}_{i.}}{S_i}\ , i = 1, 2, \ldots r; j = 1, 2, \ldots, n_i,$$

where

$$S_i^2 = \frac{1}{n_i}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{i.})^2.$$

By construction each $\{V_{ij}^0, j = 1, 2, \ldots, n_i\}$, $i = 1, 2, \ldots, r$ has a sample mean equal to 0 and a sample variance equal to 1.

Remark:

Martin (2007) proposed a transformation where each $\{V_{ij}^0, j = 1, 2, \ldots, n_i\}$, $i = 1, 2, \ldots, r$ has a sample mean equal to $\overline{Y}_{..}$ and a sample variance equal to $S_p^2 = \sum_{i=1}^{r} n_i S_i^2/(n_T - r)$. Again, this quantity will disappear in the test statistic.

We define $C_n^{(1)}(\alpha; \mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_r)$, the bootstrap estimator of $C_n(\alpha)$, by

$$P_{H_0^*}^*\left(\frac{(n_T - r)\sum_{i=1}^{r} n_i(\overline{V}_{i.}^{0*} - \overline{V}_{..}^{0*})^2}{(r-1)\sum_{i=1}^{r}\sum_{j=1}^{n_i}(V_{ij}^{0*} - \overline{V}_{i.}^{0*})^2} \geq C_n^{(1)}(\alpha; \mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_r)\right) \cong \alpha.$$

(d) Testing the equality of two distributions

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with unknown distribution function F and let $Y_1, Y_2, \ldots, Y_m$ be i.i.d. random variables with unknown distribution function G. Assume F and G are *continuous and consider testing the following hypothesis:*

$$H_0 : F = G \quad \text{vs.} \quad H_A : F \neq G.$$

Efron and Tibshirani (1993) proposed testing this hypothesis using (3.8) as the test statistic. Under H_0 there is no difference between the probability structures that generated $\mathbf{X}_n = (X_1, X_2, \ldots, X_n)$ and $\mathbf{Y}_m = (Y_1, Y_2, \ldots, Y_m)$ and thus, must have come from the same probability structure. The transformation suggested by Efron and Tibshirani (1993) to transform the data to "mimic" H_0 is simple: Pool all the data together to form a new data set, i.e.,

$$\mathbf{V}_{n+m}^0 = (X_1, X_2, \ldots, X_n, Y_1, Y_2, \ldots, Y_m).$$

Let the distribution function H denote the distribution function of the data under H_0 (i.e., $H = F = G$). Thus, generating n elements from F and then m elements from G is the same as generating $n+m$ elements from H, if H_0 is true.

It is now possible to create an empirical distribution function, H_n, from the null distribution by placing probability $\dfrac{1}{n+m}$ on each of the elements in the pooled sample $\mathbf{V}_{n+m}^0$.

The bootstrap sample is then given by: $V_1^{0*}, V_2^{0*}, \ldots, V_{n+m}^{0*}$ i.i.d. from H_n.

Now, denote the first n observations by $\mathbf{V}_n^{xo*}$ and the remaining m observations by $\mathbf{V}_m^{yo*}$.

Choose $C_n(\alpha; \mathbf{X}_n, \mathbf{Y}_m)$, the bootstrap estimator of the critical value, such that

$$P_{H_0^*}^*\left(\frac{\overline{V}_n^{x0*} - \overline{V}_m^{y0*}}{S_p(\mathbf{V}_n^{x0*}, \mathbf{V}_m^{y0*})\sqrt{\frac{1}{n} + \frac{1}{m}}} \geq C_n(\alpha; \mathbf{X}_n, \mathbf{Y}_m)\right) \cong \alpha,$$

where

$$\begin{aligned}
\overline{V}_n^{x0*} &= \frac{1}{n}\sum_{i=1}^{n} V_i^{x0*}; \\
\overline{V}_m^{y0*} &= \frac{1}{m}\sum_{j=1}^{m} V_j^{y0*} \text{ and} \\
S_p^2(\mathbf{V}_n^{x0*}, \mathbf{V}_m^{y0*}) &= \frac{nS_n^2(\mathbf{V}_n^{x0*}) + mS_m^2(\mathbf{V}_m^{y0*})}{n+m-2}.
\end{aligned}$$

Remark:

Various other tests have been developed to test $H_0 : F = G$. Permutation tests are often applied in practice. The interested reader is referred to, among others, Anderson et al. (1994); Jing (1995); Allen (1997) and Cao and Van Keilegom (2006).

(e) The variance in the univariate case

In this section we consider a bootstrap test for the hypothesis

$$H_0 : \sigma^2 = \sigma_0^2 \quad \text{vs.} \quad H_A : \sigma^2 > \sigma_0^2$$

from a random sample $\mathbf{X}_n = (X_1, X_2, \ldots, X_n)$ taken from an unknown distribution F with variance σ^2.

The test statistic most commonly used for testing this hypothesis is

$$T_n(\mathbf{X}_n) = \frac{nS_n^2(\mathbf{X}_n)}{\sigma_0^2}.$$

Martin (2007) proposed the following transformation of $\{X_i,\ i = 1, 2, \ldots, n\}$:

$$V_i^0 = \frac{X_i}{S_n(\mathbf{X}_n)}.\sigma_0\ ,\ i = 1, 2, \ldots, n. \tag{3.9}$$

By construction, the sample variance of $\{V_i^0,\ i = 1, 2, \ldots, n\}$ is σ_0^2.

$C_n^R(\alpha; \mathbf{X}_n)$, the bootstrap estimator of the critical value, is now defined by

$$P^*_{H_0^*}\left(\frac{nS_n^2(\mathbf{V}_n^{0*})}{\sigma_0^2} \geq C_n^R(\alpha; \mathbf{X}_n)\right) \cong \alpha.$$

Because of the transformation, $C_n^R(\alpha; \mathbf{X}_n)$ can also be computed from

$$P^*_{H_0^*}\left(\frac{nS_n^2(\mathbf{X}_n^*)}{S_n^2(\mathbf{X}_n)} \geq C_n^R(\alpha; \mathbf{X}_n)\right) \cong \alpha. \tag{3.10}$$

If, however, the data are not transformed to have variance σ_0^2, then (3.10) becomes

$$P^*_{H_0^*}\left(\frac{nS_n^2(\mathbf{X}_n^*)}{\sigma_0^2} \geq C_n^W(\alpha; \mathbf{X}_n)\right) \cong \alpha. \tag{3.11}$$

The following theorem states that the size of the test based on $C_n^W(\alpha; \mathbf{X}_n)$ converges to zero as $n \to \infty$.

Theorem 3.2.1

$$P_{H_0}\left(\frac{nS_n^2(\mathbf{X}_n)}{\sigma_0^2} \geq C_n^W(\alpha; \mathbf{X}_n)\right) \to 0 \quad as \quad n \to \infty.$$

Proof.

$$\begin{aligned} \alpha &\cong P^*_{H_0^*}\left(\frac{nS_n^2(\mathbf{X}_n^*)}{\sigma_0^2} \geq C_n^W(\alpha; \mathbf{X}_n)\right) \\ &= P^*_{H_0^*}\left(\frac{nS_n^2(\mathbf{X}_n^*)}{S_n^2(\mathbf{X}_n)} \geq C_n^W(\alpha; \mathbf{X}_n)\cdot \frac{\sigma_0^2}{S_n^2(\mathbf{X}_n)}\right). \end{aligned} \tag{3.12}$$

Recall that $C_n^R(\alpha; \mathbf{X}_n)$ is defined by (see (3.10))

$$P^*_{H_0^*}\left(\frac{nS_n^2(\mathbf{X}_n^*)}{S_n^2(\mathbf{X}_n)} \geq C_n^R(\alpha; \mathbf{X}_n)\right) \cong \alpha. \tag{3.13}$$

It now follows from (3.12) and (3.13) that

$$C_n^W(\alpha; \mathbf{X}_n) = \frac{C_n^R(\alpha; \mathbf{X}_n)S_n^2(\mathbf{X}_n)}{\sigma_0^2}. \tag{3.14}$$

Hence, from (3.14) it follows that

$$\begin{aligned} P_{H_0}\left(\frac{nS_n^2(\mathbf{X}_n)}{\sigma_0^2} \geq C_n^W(\alpha;\mathbf{X}_n)\right) &= P_{H_0}\left(\frac{nS_n^2(\mathbf{X}_n)}{\sigma_0^2} \geq \frac{C_n^R(\alpha;\mathbf{X}_n)S_n^2(\mathbf{X}_n)}{\sigma_0^2}\right) \\ &= P_{H_0}(n \geq C_n^R(\alpha;\mathbf{X}_n)) \\ &\to 0 \quad \text{as} \quad n \to \infty, \end{aligned}$$

since it follows from Section 2.4 of Hall (1992) that

$$\frac{C_n^R(\alpha;\mathbf{X}_n) - n}{\sqrt{2n}} \to \Phi^{-1}(1-\alpha) \quad \text{a.s. as} \quad n \to \infty.$$

Q.E.D.■

In Chapter 8 we will present the results of a Monte-Carlo study where we compare the performance of the test based on $C_n^R(\alpha;\mathbf{X}_n)$ with the test based on $C_n^W(\alpha;\mathbf{X}_n)$.

In order to estimate the power of the test, at a specific alternative σ_A^2, transform $\{X_i, i = 1,2,\ldots,n\}$ as follows:

$$V_i^A = \frac{X_i}{S_n(\mathbf{X}_n)}.\sigma_A \ , i = 1,2,\ldots,n.$$

The transformed data $\{V_i^0,\ i = 1,2,\ldots,n\}$ now have a sample variance equal to σ_A^2.
The estimated power of the test, at the specific alternative, is given by

$$P_{boot}^A = P_{H_A^*}^*\left(\frac{nS_n^2(\mathbf{V}_n^{A*})}{\sigma_0^2} \geq C_n^R(\alpha;\mathbf{X}_n)\right).$$

From the transformation, the estimated power of the test, at the specific alternative, is therefore given by

$$P_{boot}^A = P_{H_A^*}^*\left(\frac{nS_n^2(\mathbf{X}_n^*)}{S_n^2(\mathbf{X}_n)} \cdot \frac{\sigma_A^2}{\sigma_0^2} \geq C_n^R(\alpha;\mathbf{X}_n)\right).$$

(f) Testing the equality of variances

Let $\mathbf{X}_n = (X_1, X_2, \ldots, X_n)$ and $\mathbf{Y}_m = (Y_1, Y_2, \ldots, Y_m)$ be independent random samples from unknown distributions F and G, respectively. Consider testing the hypothesis:

$$H_0 : \sigma_x^2 = \sigma_y^2 \quad \text{vs.} \quad H_A : \sigma_x^2 > \sigma_y^2,$$

where σ_x^2 is the variance of F and σ_y^2 is the variance of G.
The test statistic most commonly used for this test is

$$T_{n,m}(\mathbf{X}_n, \mathbf{Y}_m) = \frac{S_n^2(\mathbf{X}_n)}{S_m^2(\mathbf{Y}_m)},$$

where $\frac{n(m-1)}{m(n-1)} T_{n,m}(\mathbf{X}_n, \mathbf{Y}_m)$ has a $F_{n-1,m-1}$-distribution under H_0, if F and G are normally distributed.

Boos and Brownie (1989) and Boos et al. (1989) discussed two possible procedures that can be used to test this hypothesis.

The first procedure that they describe involves transforming $\{X_i, i = 1, 2, \ldots, n\}$ and $\{Y_j, j = 1, 2, \ldots, m\}$ as follows:

$$\begin{aligned} V_i^{x0} &= \frac{X_i}{S_n(\mathbf{X}_n)} \ , i = 1, 2, \ldots, n \\ V_j^{y0} &= \frac{Y_j}{S_m(\mathbf{Y}_m)} \ , j = 1, 2, \ldots, m. \end{aligned}$$

By construction, both $\{V_i^{x0},\ i = 1, 2, \ldots, n\}$ and $\{V_j^{y0},\ j = 1, 2, \ldots, m\}$ have a sample variance of 1, thus satisfying H_0.

Remark:

Another possible transformation would be to transform $\mathbf{X}_n$ and $\mathbf{Y}_n$ to have a sample variance equal to S_p^2, the pooled variance. This, however, will yield exactly the same results as the transformation described above, because S_p^2 will disappear in the test statistic.

Choose $C_n(\alpha; \mathbf{X}_n, \mathbf{Y}_m)$, the bootstrap estimator of the critical value, such that

$$P^*_{H_0^*}\left(\frac{S_n^2(\mathbf{V}_n^{x0*})}{S_m^2(\mathbf{V}_m^{y0*})} \geq C_n(\alpha; \mathbf{X}_n, \mathbf{Y}_m) \right) \cong \alpha,$$

which is equivalent to choosing $C_n(\alpha; \mathbf{X}_n \mathbf{Y}_m)$ such that

$$P^*_{H_0^*}\left(\frac{S_n^2(\mathbf{X}_n^*)}{S_n^2(\mathbf{X}_n)} \Big/ \frac{S_m^2(\mathbf{Y}_m^*)}{S_m^2(\mathbf{Y}_m)} \geq C_n(\alpha; \mathbf{X}_n, \mathbf{Y}_m) \right) \cong \alpha.$$

In order to estimate the power of the test, at a specific alternative $H_A : \frac{\sigma_x^2}{\sigma_y^2} = k > 1$, the data need to be transformed as follows:

$$\begin{aligned} V_i^{xA} &= \frac{X_i}{S_n(\mathbf{X}_n)} \cdot \sqrt{k} \ , \ i = 1, 2, \ldots, n \\ V_j^{yA} &= \frac{Y_j}{S_m(\mathbf{Y}_m)}, \ j = 1, 2, \ldots, m. \end{aligned}$$

The sample variance of $\{V_i^{xA},\ i = 1, 2, \ldots, n\}$ is equal to k, while the sample variance of $\{V_j^{yA},\ j = 1, 2, \ldots, m\}$ is equal to 1.

The estimated power of the test, at the specific alternative $\dfrac{\sigma_x^2}{\sigma_y^2} = k$, is given by

$$P_{boot}^{A} = P_{H_A^*}^{*}\left(\frac{S_n^2(\mathbf{V}_n^{xA*})}{S_m^2(\mathbf{V}_m^{yA*})} \geq C_n(\alpha;\mathbf{X}_n,\mathbf{Y}_m)\right),$$

or equivalently by

$$P_{boot}^{A} = P_{H_A^*}^{*}\left(\frac{kS_n^2(\mathbf{X}_n^*)}{S_n^2(\mathbf{X}_n)} / \frac{S_m^2(\mathbf{Y}_m^*)}{S_m^2(\mathbf{Y}_m)} \geq C_n(\alpha;\mathbf{X}_n,\mathbf{Y}_m)\right).$$

The second procedure proposed by Boos and Brownie (1989) is very similar to the one used for testing the equality of two distribution functions.
First they make a location-scale assumption. That is, they assume that $F(x) = H\left(\dfrac{x-\mu_x}{\sigma_x}\right)$, $G(y) = H\left(\dfrac{y-\mu_y}{\sigma_y}\right)$, with H unknown and μ_x and μ_y unknown location parameters.
The data $\{X_i, i = 1,2,\ldots,n\}$ and $\{Y_j, j = 1,2,\ldots,m\}$ are then transformed in the following manner:

$$\begin{aligned}\tilde{X}_i &= X_i - \overline{X}_n\ ,\ i = 1,2,\ldots,n\\ \tilde{Y}_j &= Y_j - \overline{Y}_m\ ,\ j = 1,2,\ldots,m.\end{aligned}$$

The next step is to pool the newly transformed data, $\{\tilde{X}_i, i = 1,2,\ldots,n\}$ and $\{\tilde{Y}_j, j = 1,2,\ldots m\}$, to form a new data set

$$\mathbf{V}_{n+m}^0 = (\tilde{X}_1,\tilde{X}_2,\ldots,\tilde{X}_n,\tilde{Y}_1,\tilde{Y}_2,\ldots,\tilde{Y}_m).$$

Resampling is now done, in exactly the same manner as described in Subsection (d) - **Testing the equality of two distributions**.

Boos et al. (1989) showed that both procedures will have the correct significance level asymptotically. They also indicated that the pooled procedure has an advantage over the other procedure in small to moderate sample sizes.

Remark:

The same procedures, as described in this section, can be used to test the equality of variances of more than two populations. The only difference is that we will base the test on a different test statistic, for example Bartlett's k-sample test statistic given by:

$$T = \frac{\sum_{i=1}^{k}(n_i - 1)\log\frac{S_p^2}{S_i^2}}{1 + \frac{1}{3(k-1)}\left[\sum_{i=1}^{k}\left(\frac{1}{n_i - 1}\right) - \frac{1}{n_T - k}\right]},$$

where k is the number of samples, n_i is the number of elements in the ith sample, S_i^2 is the sample variance of the ith sample, S_p^2 is the pooled sample variance and $n_T = \sum_{i=1}^{k} n_i$.

(g) Pearson's correlation coefficient

Let $\mathbf{Z}_n = \big((X_1, Y_1), \ldots, (X_n, Y_n)\big)$ be a bivariate random sample from a population (X, Y) with unknown joint distribution function $H(x, y)$ and variance-covariance matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_x^2 & \rho_p\sigma_x\sigma_y \\ \rho_p\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix},$$

where ρ_p denotes Pearson's correlation coefficient.

Consider testing the hypothesis

$$H_0 : \rho_p = 0 \quad \text{vs.} \quad H_A : \rho_p > 0.$$

A test statistic often used in this scenario is

$$T_n(\mathbf{Z}_n) = \frac{\hat{\rho}_p(\mathbf{Z}_n)\sqrt{n-2}}{\sqrt{1 - \hat{\rho}_p(\mathbf{Z}_n)^2}},$$

where

$$\hat{\rho}_p(\mathbf{Z}_n) = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)(Y_i - \overline{Y}_n)}{S_n(\mathbf{X}_n).S_n(\mathbf{Y}_n)}$$

is an estimator for ρ_p; $\mathbf{X}_n = (X_1, X_2, \ldots, X_n)$ and $\mathbf{Y}_n = (Y_1, Y_2, \ldots, Y_n)$. If $H(x, y)$ is a bivariate normal distribution, then $T(\mathbf{Z}_n)$ has a t_{n-2}-distribution (see e.g., Neter et al., 1996).

Westfall and Young (1993) proposed the following bootstrap approach to enable us to test this hypothesis:

(a) Consider $\mathbf{X}_n$ and $\mathbf{Y}_n$ separately.

(b) Independently obtain resample data $\mathbf{V}_n^{x0*} = (V_1^{x0*}, V_2^{x0*}, \ldots, V_n^{x0*})$ and $\mathbf{V}_n^{y0*} = (V_1^{y0*}, V_2^{y0*}, \ldots, V_n^{y0*})$, where the components of $\mathbf{V}_n^{x0*}$ are i.i.d. drawn from the e.d.f. of $\mathbf{X}_n$, and the components of $\mathbf{V}_n^{y0*}$ are i.i.d. drawn from the e.d.f. of $\mathbf{Y}_n$.

(c) Now, construct the vector $\mathbf{Z}_n^{0*} = \big((V_1^{x0*}, V_1^{y0*}), (V_2^{x0*}, V_2^{y0*}), \ldots, (V_n^{x0*}, V_n^{y0*})\big)$.

- Westfall and Young (1993) thus obtained resamples according to the independence hypothesis and they choose $C_n(\alpha; \mathbf{Z}_n)$, the bootstrap estimator of the critical value, such that

$$P^*_{H_0^*}(\hat{\rho}_p(\mathbf{Z}_n^{0*}) \geq C_n(\alpha; \mathbf{Z}_n)) \cong \alpha,$$

where

$$\hat{\rho}_p(\mathbf{Z}_n^{0*}) = \frac{\frac{1}{n}\sum_{i=1}^{n}(V_i^{x0*} - \overline{V}_n^{x0*})(V_i^{y0*} - \overline{V}_n^{y0*})}{S_n(\mathbf{V}_n^{x0*}).S_n(\mathbf{V}_n^{y0*})}$$

Note that the correlation coefficient itself is used as the test statistic and not a pivotal test statistic. This choice of test statistic is due to the difficulty of forming a reliable variance estimator for the sample correlation coefficient (Hall et al., 1988).

Next, consider testing the more general hypothesis:

$$H_0 : \rho_p = \rho_0 \quad \text{vs.} \quad H_A : \rho_p > \rho_0.$$

In order to test this hypothesis, Martin (2007) proposed the following transformation of $\{(X_i, Y_i), i = 1, 2, \ldots, n\}$:

$$\begin{aligned} V_i^{x0} &= 0.5(\frac{X_i - \overline{X}}{S_n(\mathbf{X}_n)})\left[\sqrt{\frac{1+\rho_0}{1+\hat{\rho}_p(\mathbf{Z}_n)}} + \sqrt{\frac{1-\rho_0}{1-\hat{\rho}_p(\mathbf{Z}_n)}}\right] + \\ &\quad 0.5(\frac{X_i - \overline{X}}{S_n(\mathbf{X}_n)})\left[\sqrt{\frac{1+\rho_0}{1+\hat{\rho}_p(\mathbf{Z}_n)}} - \sqrt{\frac{1-\rho_0}{1-\hat{\rho}_p(\mathbf{Z}_n)}}\right], \ i = 1, 2, \ldots, n, \end{aligned} \tag{3.15}$$

$$\begin{aligned} V_i^{y0} &= 0.5(\frac{Y_i - \overline{Y}}{S_n(\mathbf{Y}_n)})\left[\sqrt{\frac{1+\rho_0}{1+\hat{\rho}_p(\mathbf{Z}_n)}} - \sqrt{\frac{1-\rho_0}{1-\hat{\rho}_p(\mathbf{Z}_n)}}\right] + \\ &\quad 0.5(\frac{Y_i - \overline{Y}}{S_n(\mathbf{Y}_n)})\left[\sqrt{\frac{1+\rho_0}{1+\hat{\rho}_p(\mathbf{Z}_n)}} + \sqrt{\frac{1-\rho_0}{1-\hat{\rho}_p(\mathbf{Z}_n)}}\right], \ i = 1, 2, \ldots, n. \end{aligned} \tag{3.16}$$

Some calculations show that Pearson's sample correlation coefficient between $\{V_i^{x0},\ i = 1, 2, \ldots, n\}$ and $\{V_i^{y0},\ i = 1, 2, \ldots, n\}$ is now equal to ρ_0. Resampling is done from the pairs (V_1^{x0}, V_1^{y0}), (V_2^{x0}, V_2^{y0}), $\ldots$, (V_n^{x0}, V_n^{y0}) to obtain the bootstrap sample $\mathbf{Z}_n^{0*} = ((V_1^{x0*}, V_1^{y0*}), (V_2^{x0*}, V_2^{y0*}), \ldots, (V_n^{x0*}, V_n^{y0*}))$. *The appropriate bootstrap estimator of the critical value, $C_n(\alpha; \mathbf{Z}_n)$, is then defined by*

$$P^*_{H_0^*}\left(\hat{\rho}_p(\mathbf{Z}_n^{0*}) \geq C_n(\alpha; \mathbf{Z}_n)\right) \cong \alpha.$$

To estimate the power, at a specific alternative $\rho_p = \rho_A$, transform $\{(X_i, Y_i), i = 1, 2, \ldots, n\}$ as follows:

$$\begin{aligned}
V_i^{xA} &= 0.5(\frac{X_i - \overline{X}}{S_n(\mathbf{X}_n)})\left[\sqrt{\frac{1+\rho_A}{1+\hat{\rho}_p(\mathbf{Z}_n)}} + \sqrt{\frac{1-\rho_A}{1-\hat{\rho}_p(\mathbf{Z}_n)}}\right] + \\
& \quad 0.5(\frac{X_i - \overline{X}}{S_n(\mathbf{X}_n)})\left[\sqrt{\frac{1+\rho_A}{1+\hat{\rho}_p(\mathbf{Z}_n)}} - \sqrt{\frac{1-\rho_A}{1-\hat{\rho}_p(\mathbf{Z}_n)}}\right], \ i = 1, 2, \ldots, n,
\end{aligned}$$

$$\begin{aligned}
V_i^{yA} &= 0.5(\frac{Y_i - \overline{Y}}{S_n(\mathbf{Y}_n)})\left[\sqrt{\frac{1+\rho_A}{1+\hat{\rho}_p(\mathbf{Z}_n)}} - \sqrt{\frac{1-\rho_A}{1-\hat{\rho}_p(\mathbf{Z}_n)}}\right] + \\
& \quad 0.5(\frac{Y_i - \overline{Y}}{S_n(\mathbf{Y}_n)})\left[\sqrt{\frac{1+\rho_A}{1+\hat{\rho}_p(\mathbf{Z}_n)}} + \sqrt{\frac{1-\rho_A}{1-\hat{\rho}_p(\mathbf{Z}_n)}}\right], \ i = 1, 2, \ldots, n.
\end{aligned}$$

By construction, Pearson's sample correlation coefficient between $\{V_i^{xA},\ i = 1, 2, \ldots, n\}$ and $\{V_i^{yA},\ i = 1, 2, \ldots, n\}$ is equal to ρ_A. Resampling is now done from the pairs (V_1^{xA}, V_1^{yA}), (V_2^{xA}, V_2^{yA}), $\ldots, (V_n^{xA}, V_n^{yA})$ to obtain the bootstrap sample $\mathbf{Z}_n^{A*} = ((V_1^{xA*}, V_1^{yA*}), (V_2^{xA*}, V_2^{yA*}), \ldots, (V_n^{xA*}, V_n^{yA*}))$. The estimated power of the test, at the specific alternative $\rho_p = \rho_A$, is then given by:

$$P_{boot}^{A} = P_{H_A^*}^{*}\left(\hat{\rho}_p(\mathbf{Z}_n^{A*}) \geq C_n(\alpha; \mathbf{Z}_n)\right).$$

3.3 Exponentially tilted version of the e.d.f.

In Section 3.2 we discussed methods to transform the data to "mimic" H_0 or H_A. An alternative approach is to leave the data values fixed and instead change the probability (p_i) on each data value X_i to "mimic" H_0 or H_A (Efron, 1981).

To be more specific, consider the hypothesis in (3.1).

Let $\mathbf{p} = (p_1, p_2, \ldots, p_n)$ and let F_{Tilt} be the distribution which places probability p_i on each $X_i,\ i = 1, 2, \ldots, n$. It is now possible to choose $\mathbf{p}$ in such a way that the distance between F_{Tilt} and F_n is minimised, subject to the constraints:

$$\theta(F_{Tilt}) = \theta_0 \quad \text{and} \quad \sum_{i=1}^{n} p_i = 1.$$

Denote this solution by $\mathbf{p}^0 = (p_1^0, p_2^0, \ldots, p_n^0)$

Remark: In the literature F_{Tilt}^0 (i.e., F_{Tilt} based on $\mathbf{p}^0$) is often referred to as the exponentially tilted version of F_n.

A useful distance measure for this constrained minimisation problem is the reverse information distance or *Kullback-Leibler* distance

$$\sum_{i=1}^{n} p_i \log(np_i).$$

Alternative measures that can be used include maximising $\prod_{i=1}^{n} p_i$ (or minimise $\sum_{i=1}^{n} \frac{1}{n} \log\left(\frac{1}{np_i}\right)$) or to minimise the expression: $\max_i \left|p_i - \frac{1}{n}\right|$ (Young, 1988; Davison and Hinkley, 1997).

After obtaining $\mathbf{p}^0$, the solution to the constrained minimisation problem, the bootstrap random sample $(V_1^{0*}, V_2^{0*}, \ldots, V_n^{0*})$ drawn from F_{Tilt}^0 can be obtained. The bootstrap estimator of the critical value, $C_n(\alpha; \mathbf{X}_n)$, is then defined by

$$P_{H_0^*}^*(T_n(V_1^{0*}, V_2^{0*}, \ldots, V_n^{0*}) \geq C_n(\alpha; \mathbf{X}_n)) \cong \alpha.$$

In order to estimate the power of the test, at the specific alternative θ_A, the vector $\mathbf{p} = (p_1, p_2, \ldots, p_n)$ is chosen in such a way that the distance between F_{Tilt} and F_n is minimised, subject to the constraints:

$$\theta(F_{Tilt}) = \theta_A \quad \text{and} \quad \sum_{i=1}^{n} p_i = 1.$$

Denote this solution by $\mathbf{p}^A = (p_1^A, p_2^A, \ldots, p_n^A)$.
The bootstrap random sample is then given by $(V_1^{A*}, V_2^{A*}, \ldots, V_n^{A*})$ obtained from F_{Tilt}^A (i.e., F_{Tilt} based on $\mathbf{p}^A$). The bootstrap estimated power, at a specific alternative, is given by

$$P_{boot}^A = P_{H_A^*}^*(T_n(V_1^{A*}, V_2^{A*}, \ldots, V_n^{A*}) \geq C_n(\alpha; \mathbf{X}_n)).$$

Two common testing scenarios will be used to illustrate this method, namely testing the mean (see Efron, 1981) and testing the variance in the univariate case. Chapter 8 contains the results of a Monte-Carlo study conducted to compare the transformation method to this method in terms of estimated sizes for testing the variance in the univariate case. For a Monte-Carlo comparison between the two methods for the mean in the univariate case, the interested reader is referred to Young (1988).

3.3.1 The mean in the univariate case

Refer to Section 3.2.2(a), where the notation for testing the mean was introduced.

The probability vector $\mathbf{p}^0 = (p_1^0, p_2^0, \ldots, p_n^0)$ is obtained by minimising $\sum_{i=1}^{n} p_i \log(np_i)$, subject to the constraint that the mean of F_{Tilt}($\overline{X}_p = \sum_{i=1}^{n} p_i X_i$) is equal to μ_0 and $\sum_{i=1}^{n} p_i = 1$.
The solution of this constrained minimisation problem is given by:

$$p_i^0 = \frac{e^{\lambda X_i}}{\sum_{j=1}^{n} e^{\lambda X_j}},$$

where λ is chosen such that $\sum_{i=1}^{n} p_i^0 X_i = \mu_0$ (Efron, 1981).

3.3.2 The variance in the univariate case

Refer to Section 3.3.2(e) for the notation used for testing the variance.

In order to obtain $\mathbf{p}^0$ we minimise the Lagrange expression

$$L(\mathbf{p}, \lambda) = \sum_{i=1}^{n} p_i \log(np_i) - \lambda \left(\sum_{i=1}^{n} p_i (X_i - \overline{X}_p)^2 - \sigma_0^2 \right),$$

where $\sum_{i=1}^{n} p_i (X_i - \overline{X}_p)^2$ is the variance of F_{Tilt}, with $\overline{X}_p = \sum_{i=1}^{n} p_i X_i$.
Setting the derivative, with respect to p_i, equal to zero produces the following equality:

$$\log(np_i) + 1 - \lambda(X_i - \overline{X}_p)^2 = 0,$$

which together with the constraint $\sum_{i=1}^{n} p_i = 1$, gives the implicit solution

$$p_i^0 = \frac{e^{\lambda(X_i - \overline{X}_p^0)^2}}{\sum_{j=1}^{n} e^{\lambda(X_j - \overline{X}_p^0)^2}}. \tag{3.17}$$

If we assume that $\overline{X}_p^0 = \overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$, (3.17) becomes:

$$p_i^0 = \frac{e^{\lambda(X_i - \overline{X})^2}}{\sum_{j=1}^{n} e^{\lambda(X_j - \overline{X})^2}},$$

where λ is chosen such that $\sum_{i=1}^{n} p_i^0 (X_i - \overline{X})^2 = \sigma_0^2$. That is, λ is chosen such that

$$\frac{\sum_{i=1}^{n} (X_i - \overline{X})^2 e^{\lambda(X_i - \overline{X})^2}}{\sum_{j=1}^{n} e^{\lambda(X_j - \overline{X})^2}} = \sigma_0^2.$$

Two drawbacks of the tilted bootstrap are that a solution for λ can not always be obtained (especially in small samples) and that the method is very computationally intensive.

In the next chapter a new method to evaluate the performance of bootstrap-based tests will be discussed.

Chapter 4

A new method of evaluating the performance of bootstrap-based tests

In this chapter we discuss two methods to evaluate the performance of bootstrap-based tests: the first method is one that is traditionally used in the literature, while the second is an alternative, more robust, method that we propose. The chapter concludes by applying these two methods to an example.

4.1 Introduction

When a bootstrap-based test is proposed, one would like to evaluate the performance of the test in order to assess how "good" the proposed test is. This evaluation can be done theoretically and/or by means of a Monte-Carlo simulation.

In Section 4.3 the evaluation method that is currently in use in the literature is discussed. We will refer to this evaluation method as *Method I.* In Section 4.4 we propose a new method of evaluating the performance of a bootstrap-based test. We will refer to this new evaluation method as *Method II.* Section 4.2 introduces the basic notation that will be used in discussing these two methods.

4.2 Notation

Assume that observations $X_1, X_2, \ldots, X_n$ are available from some model with joint d.f. $F_{\boldsymbol{\theta},\boldsymbol{\nu}}(x_1, \ldots, x_n)$, depending on some unknown parameters $\boldsymbol{\theta}$ and $\boldsymbol{\nu}$. Again, let $\mathbf{X}_n = (X_1, X_2, \ldots, X_n)$ and denote by $\mathbf{x}_n = (x_1, x_2, \ldots, x_n)$ an observed realization of $\mathbf{X}_n$.

Consider the hypothesis

$$H_0 : \boldsymbol{\theta} \in \Theta_0 \qquad \text{vs.} \qquad H_A : \boldsymbol{\theta} \in \Theta_A,$$

where Θ_0 and Θ_A are two disjoint subsets of some parameter space $\Theta = \Theta_0 \cup \Theta_A$.
Assume, without loss of generality, that the test procedure is of the form:
Reject H_0 if and only if

$$T_n(\mathbf{X}_n) \geq C_n(\alpha; \mathbf{X}_n),$$

where $T_n(\mathbf{X}_n)$ is an appropriate test statistic, $C_n(\alpha; \mathbf{X}_n)$ is a bootstrap critical value and α is the significance level of the test.
For testing H_0 we consider $\boldsymbol{\nu}$ as a nuisance parameter, which will be replaced by an estimator that is strongly consistent under both H_0 and H_A. These estimated nuisance parameters are therefore not indicated in the notation of $T_n(\mathbf{X}_n)$ and $C_n(\alpha; \mathbf{X}_n)$.
Furthermore, let $\boldsymbol{\theta}_T$ denote the true value of the parameter. If $\boldsymbol{\theta}_T \in \Theta_0$ it is denoted by $\boldsymbol{\theta}^0$, otherwise we write it as $\boldsymbol{\theta}^A$.

4.3 Method I

In order to assess the accuracy of the bootstrap critical value $C_n(\alpha; \mathbf{X}_n)$, the following measure is currently in use in the literature:

$$P_{\boldsymbol{\theta}^0}(T_n(\mathbf{X}_n) \geq C_n(\alpha; \mathbf{X}_n)).$$

(See, e.g., De Beer and Swanepoel, 1989; Boos et al., 1989; Boos and Brownie, 1989; Fisher and Hall, 1990; Rayner, 1990; Nankervis and Savin, 1996; Cao and Van Keilegom, 2006; Davidson and Mackinnon, 2007; Martin, 2007.)
The power of the bootstrap-based test is evaluated similarly:

$$P_{\boldsymbol{\theta}^A}(T_n(\mathbf{X}_n) \geq C_n(\alpha; \mathbf{X}_n)).$$

It is important to note that $C_n(\alpha; \mathbf{X}_n)$ is a random variable which depends on the sample. Sakov (1998) remarked that *"this is exactly the problem in using the bootstrap to estimate the critical value. One property that a 'good' estimate of the critical value should satisfy is that the critical value estimate should be the same whether the null hypothesis is correct or not"*.

This remark by Sakov (1998) led us to develop an alternative, more robust, evaluation method.

4.4 Method II

Suppose that $\mathbf{V}_n^0 = (V_1^0, V_2^0, \ldots, V_n^0)$ is a "pseudo random 'test' sample" with joint d.f. $F_{\boldsymbol{\theta}^0,\boldsymbol{\tau}}(\cdot)$ and assume that $\mathbf{V}_n^0$ is independent of the "training" sample $\mathbf{X}_n$. Here, $\boldsymbol{\tau}$ is a nuisance parameter which may differ from $\boldsymbol{\nu}$, the nuisance parameter defined in Section 4.2. It will also be replaced by some strongly consistent estimator. In order to assess the accuracy of the bootstrap critical value $C_n(\alpha; \mathbf{X}_n)$, we propose the following measure:

$$P(T_n(\mathbf{V}_n^0) \geq C_n(\alpha; \mathbf{X}_n)).$$

Remark:

Let $\varphi^0(\mathbf{x}_n) = P(T_n(\mathbf{V}_n^0) \geq C_n(\alpha; \mathbf{x}_n))$, then

$$\begin{aligned} P(T_n(\mathbf{V}_n^0) \geq C_n(\alpha; \mathbf{X}_n)) &= E_{\boldsymbol{\theta}}\left(P(T_n(\mathbf{V}_n^0) \geq C_n(\alpha; \mathbf{X}_n) | \mathbf{X}_n)\right) \\ &= E_{\boldsymbol{\theta}}\left(\varphi^0(\mathbf{X}_n)\right), \end{aligned}$$

since

$$\begin{aligned} P(T_n(\mathbf{V}_n^0) \geq C_n(\alpha; \mathbf{X}_n) | \mathbf{X}_n = \mathbf{x}_n) &= P(T_n(\mathbf{V}_n^0) \geq C_n(\alpha; \mathbf{x}_n) | \mathbf{X}_n = \mathbf{x}_n) \\ &= P(T_n(\mathbf{V}_n^0) \geq C_n(\alpha; \mathbf{x}_n)) \\ &= \varphi^0(\mathbf{x}_n). \end{aligned}$$

The power of the bootstrap-based test can be evaluated similarly:

$$P(T_n(\mathbf{V}_n^A) \geq C_n(\alpha; \mathbf{X}_n)) = E_{\boldsymbol{\theta}}(\varphi^A(\mathbf{X}_n)),$$

where

$$\varphi^A(\mathbf{x}_n) = P(T_n(\mathbf{V}_n^A) \geq C_n(\alpha; \mathbf{x}_n)).$$

Now, $\mathbf{V}_n^A = (V_1^A, V_2^A, \ldots, V_n^A)$, has joint d.f. $F_{\boldsymbol{\theta}^A,\boldsymbol{\tau}}(\cdot)$ and is independent of the training sample $\mathbf{X}_n$.

Next, these two evaluation methods will be applied in a simple example.

4.5 Example: The mean in the univariate case

Let $\mathbf{X}_n = (X_1, X_2, \ldots, X_n)$ denote a random sample from an unknown univariate distribution F with finite mean μ and finite variance σ^2. In this section we will derive some theoretical properties of the bootstrap estimate of the critical value for the test

$$H_0 : \mu = \mu_0 \quad \text{vs.} \quad H_A : \mu > \mu_0,$$

based on both evaluation methods. This will be done using the "right" and "wrong" bootstrap critical value for the pivotal and non-pivotal test statistic. The term "right" refers to the case where resampling is done from data that are transformed in order to "mimic" H_0, whereas the term "wrong" refers to resampling being done from the original sample (see Section 3.2.2(a)). Results of a Monte-Carlo study, where both methods are used, will also be presented.

4.5.1 Method I

Suppose the procedure is to reject H_0 if and only if

$$T_{n,N-P}(\mathbf{X}_n) = \sqrt{n}(\overline{X}_n - \mu_0) \geq C^R_{n,N-P}(\alpha; \mathbf{X}_n),$$

or if and only if

$$T_{n,P}(\mathbf{X}_n) = \frac{\sqrt{n}(\overline{X}_n - \mu_0)}{S_n(\mathbf{X}_n)} \geq C^R_{n,P}(\alpha; \mathbf{X}_n),$$

where $C^R_{n,N-P}(\alpha; \mathbf{X}_n)$ is defined by

$$P^*_{H_0^*}\left(\sqrt{n}(\overline{X}^*_n - \overline{X}_n) \geq C^R_{n,N-P}(\alpha; \mathbf{X}_n)\right) \cong \alpha,$$

where $C^R_{n,P}(\alpha; \mathbf{X}_n)$ is defined by

$$P^*_{H_0^*}\left(\frac{\sqrt{n}(\overline{X}^*_n - \overline{X}_n)}{S_n(\mathbf{X}^*_n)} \geq C^R_{n,P}(\alpha; \mathbf{X}_n)\right) \cong \alpha.$$

Henceforth, we shall denote the above test procedures by (R,N-P) and (R,P), respectively.

Beran (1988) and Sakov (1998) showed that both

$$P_{H_0}\left(\sqrt{n}(\overline{X}_n - \mu_0) \geq C^R_{n,N-P}(\alpha; \mathbf{X}_n)\right)$$

and

$$P_{H_0}\left(\frac{\sqrt{n}(\overline{X}_n - \mu_0)}{S_n(\mathbf{X}_n)} \geq C^R_{n,P}(\alpha; \mathbf{X}_n)\right)$$

converges to α.

Next, suppose the procedure is to reject H_0 if and only if

$$T_{n,N-P}(X_n) = \sqrt{n}(\overline{X}_n - \mu_0) \geq C^W_{n,N-P}(\alpha; \mathbf{X}_n),$$

where $C^W_{n,N-P}(\alpha;\mathbf{X}_n)$ is defined by

$$P^*_{H_0^*}\left(\sqrt{n}(\overline{X}^*_n - \mu_0) \geq C^W_{n,N-P}(\alpha;\mathbf{X}_n)\right) \cong \alpha.$$

Refer to this test procedure as (W,N-P).

Theorem 4.5.1 *If $\alpha < 0.5$, then*

$$P_{H_0}(\sqrt{n}(\overline{X}_n - \mu_0) \geq C^W_{n,N-P}(\alpha;\mathbf{X}_n)) \to 0 \quad as \quad n \to \infty.$$

Proof.

$$\begin{aligned}
\alpha &\cong P^*_{H_0^*}(\sqrt{n}(\overline{X}^*_n - \mu_0) \geq C^W_{n,N-P}(\alpha;\mathbf{X}_n)) \\
&= P^*_{H_0^*}(\sqrt{n}(\overline{X}^*_n - \overline{X}_n) \geq C^W_{n,N-P}(\alpha;\mathbf{X}_n) + \sqrt{n}(\mu_0 - \overline{X}_n)). \qquad (4.1)
\end{aligned}$$

Recall that $C^R_{n,N-P}(\alpha;\mathbf{X}_n)$ is defined by

$$P^*_{H_0^*}(\sqrt{n}(\overline{X}^*_n - \overline{X}_n) \geq C^R_{n,N-P}(\alpha;\mathbf{X}_n)) \cong \alpha. \qquad (4.2)$$

It now follows from (4.1) and (4.2) that

$$C^W_{n,N-P}(\alpha;\mathbf{X}_n) = C^R_{n,N-P}(\alpha;\mathbf{X}_n) + \sqrt{n}(\overline{X}_n - \mu_0). \qquad (4.3)$$

Hence, from (4.3) and (4.12) below, it follows that

$$\begin{aligned}
& P_{H_0}(\sqrt{n}(\overline{X}_n - \mu_0) \geq C^W_{n,N-P}(\alpha;\mathbf{X}_n)) \\
= \; & P_{H_0}(\sqrt{n}(\overline{X}_n - \mu_0) \geq C^R_{n,N-P}(\alpha;\mathbf{X}_n) + \sqrt{n}(\overline{X}_n - \mu_0)) \\
= \; & P_{H_0}(0 \geq C^R_{n,N-P}(\alpha;\mathbf{X}_n)) \\
\to \; & I(0 \geq \Phi^{-1}(1-\alpha)) \\
= \; & 0.
\end{aligned}$$

Q.E.D.■

Suppose now the procedure is to reject H_0 if and only if

$$T_{n,P}(\mathbf{X}_n) = \frac{\sqrt{n}(\overline{X}_n - \mu_0)}{S_n(\mathbf{X}_n)} \geq C^W_{n,P}(\alpha;\mathbf{X}_n),$$

where $C^W_{n,P}(\alpha;\mathbf{X}_n)$ is defined by

$$P^*_{H_0^*}\left(\frac{\sqrt{n}(\overline{X}^*_n - \mu_0)}{S_n(\mathbf{X}^*_n)} \geq C^W_{n,P}(\alpha;\mathbf{X}_n)\right) \cong \alpha.$$

We will refer to this test procedure as (W,P).

Theorem 4.5.2 *Suppose that $E(X_1^4) < \infty$. Then*

$$P_{H_0}\left(\frac{\sqrt{n}(\overline{X}_n - \mu_0)}{S_n(\mathbf{X}_n)} \geq C^W_{n,P}(\alpha;\mathbf{X}_n)\right) \to 0 \quad as \quad n \to \infty.$$

Proof. Clearly,

$$\begin{aligned} & P^*_{H_0^*}\left(\frac{\sqrt{n}(\overline{X}^*_n - \mu_0)}{S_n(\mathbf{X}^*_n)} \geq C^W_{n,P}(\alpha;\mathbf{X}_n)\right) \cong \alpha \\ \Leftrightarrow \quad & P^*_{H_0^*}\left(\frac{\sqrt{n}(\overline{X}^*_n - \overline{X}_n)}{S_n(\mathbf{X}^*_n)} + \frac{\Delta^*_n}{\sqrt{n}} \leq C^W_{n,P}(\alpha;\mathbf{X}_n) - \frac{\sqrt{n}(\overline{X}_n - \mu_0)}{S_n(\mathbf{X}_n)}\right) \cong 1-\alpha, \end{aligned} \tag{4.4}$$

where

$$\Delta^*_n = \frac{\sqrt{n}(\overline{X}_n - \mu_0)}{S_n(\mathbf{X}_n)} \cdot \frac{\sqrt{n}(S_n(\mathbf{X}_n) - S_n(\mathbf{X}^*_n))}{S_n(\mathbf{X}^*_n)} = \Delta_{n1}\cdot \Delta^*_{n2}.$$

The CLT implies that, under H_0, $\Delta_{n1} \xrightarrow{d} N(0,1)$ as $n \to \infty$, so that $\Delta_{n1} = O_P(1)$. Also, applying classical bootstrap theory (see, e.g., Shao and Tu, 1995) it readily follows, using the assumption $E(X_1^4) < \infty$, that

$$\Delta^*_{n2} \xrightarrow{d^*} N(0,(\mu_4 - \sigma^4)/4\sigma^4) \quad \text{a.s. as} \quad n \to \infty, \tag{4.5}$$

where d^* denotes convergence in conditional, given $\mathbf{X}_n$, distribution (i.e. weak convergence in the bootstrap world) and $\mu_4 = E((X_1 - \mu)^4)$.

Let $a_n = n^{-1/2}\log n$. Then, using the fact that $\Delta_{n1} = O_P(1)$ and (4.5) we have that

$$\begin{aligned} P^*(n^{-1/2}|\Delta^*_n| > a_n) &= P^*(|\Delta_{n1}|\cdot|\Delta^*_{n2}| > \log n) \\ &= o_{a.s.}(1) \quad \text{as} \quad n \to \infty. \end{aligned} \tag{4.6}$$

From (4.4), (4.6) and the Lemma given on pp. 228 – 229 by Serfling (1980) we therefore have that

$$\Phi\left(C^W_{n,P}(\alpha;\mathbf{X}_n) - \frac{\sqrt{n}(\overline{X}_n - \mu_0)}{S_n(\mathbf{X}_n)}\right) = 1 - \alpha + o_{a.s.}(1),$$

implying that

$$C^W_{n,P}(\alpha;\mathbf{X}_n) = \frac{\sqrt{n}(\overline{X}_n - \mu_0)}{S_n(\mathbf{X}_n)} + \Phi^{-1}(1-\alpha) + o_{a.s.}(1). \tag{4.7}$$

Hence, from (4.7) it follows that

$$\begin{aligned} & P_{H_0}\left(\frac{\sqrt{n}(\overline{X}_n - \mu_0)}{S_n(\mathbf{X}_n)} \geq C^W_{n,P}(\alpha;\mathbf{X}_n)\right) \\ = \quad & P_{H_0}\left(\frac{\sqrt{n}(\overline{X}_n - \mu_0)}{S_n(\mathbf{X}_n)} \geq \frac{\sqrt{n}(\overline{X}_n - \mu_0)}{S_n(\mathbf{X}_n)} + \Phi^{-1}(1-\alpha) + o_{a.s.}(1)\right) \\ \to \quad & 0 \quad \text{as} \quad n \to \infty. \end{aligned}$$

Q.E.D.■

Remark: Hall (1992, p. 149) provides a heuristic proof of this theorem.

4.5.2 Monte-Carlo study: Method I

In this section we present the results of a Monte-Carlo study, based on Method I, in order to compare the sizes of the tests based on (R,P), (W,P), (R,N-P) and (W,N-P) for sample sizes $n = 20, 30, 50, 100$ and 200, respectively. A nominal significance level of $\alpha = 0.05$ was used throughout and data were generated from five different distributions:

- Normal distribution with mean μ_0 and variance 1;
- Uniform distribution with mean μ_0 and variance 1/12;
- Double exponential distribution with mean μ_0 and variance 2;
- Exponential distribution with mean μ_0 and variance 1;
- Chi-squared distribution with mean μ_0 and variance 6.

Size estimates were calculated as the proportion of 20 000 Monte-Carlo samples that resulted in rejection of H_0 using the bootstrap critical values $\hat{C}^R_{n,N-P}(\alpha; \mathbf{X}_n)$, $\hat{C}^R_{n,P}(\alpha; \mathbf{X}_n)$, $\hat{C}^W_{n,N-P}(\alpha; \mathbf{X}_n)$ and $\hat{C}^W_{n,P}(\alpha; \mathbf{X}_n)$, respectively. The standard errors of the estimated sizes are less than or equal to $\sqrt{0.25/20\ 000} = 0.0035$. Bootstrap critical values were based on $B = 1000$ independent replications.

Table 4.1: Estimated size of bootstrap hypothesis tests for the mean using evaluation Method I

Distribution	μ_0	n	(R,P)	(W,P)	(R,N-P)	(W,N-P)
Normal	0.0	20	0.048	0.000	0.060	0.000
		30	0.049	0.000	0.059	0.000
		50	0.049	0.000	0.055	0.000
		100	0.051	0.000	0.054	0.000
		200	0.050	0.000	0.050	0.000
Uniform	0.5	20	0.037	0.000	0.065	0.000
		30	0.043	0.000	0.059	0.000
		50	0.048	0.000	0.058	0.000
		100	0.049	0.000	0.054	0.000
		200	0.050	0.000	0.052	0.000
Double exponential	0.0	20	0.062	0.000	0.060	0.000
		30	0.060	0.000	0.056	0.000
		50	0.057	0.000	0.053	0.000
		100	0.054	0.000	0.051	0.000
		200	0.050	0.000	0.050	0.000
Exponential	1.0	20	0.038	0.000	0.024	0.000
		30	0.044	0.000	0.025	0.000
		50	0.047	0.000	0.028	0.000
		100	0.050	0.000	0.031	0.000
		200	0.050	0.000	0.036	0.000
Chi-squared	3.0	20	0.043	0.000	0.030	0.000
		30	0.045	0.000	0.030	0.000
		50	0.048	0.000	0.032	0.000
		100	0.050	0.000	0.033	0.000
		200	0.050	0.000	0.039	0.000

From Table 4.1 we can conclude the following:

(1) The estimated sizes of (W,N-P) and (W,P) are 0, even for large sample sizes. This agrees with the results of Theorem 4.5.1 and Theorem 4.5.2.

(2) For the tests (R,N-P) and (R,P), the estimated sizes converge to the nominal significance level as $n \to \infty$.

(3) In most cases the test (R,P) outperforms the test (R,N-P), especially for small sample sizes.

4.5.3 Method II

Again, consider the hypothesis

$$H_0 : \mu = \mu_0 \quad \text{vs.} \quad H_A : \mu > \mu_0.$$

Suppose the procedure is to reject H_0 if and only if

$$T_{n,N-P}(\mathbf{X}_n) = \sqrt{n}(\overline{X}_n - \mu_0) \geq C^W_{n,N-P}(\alpha; \mathbf{X}_n),$$

where $C^W_{n,N-P}(\alpha; \mathbf{X}_n)$ is defined by

$$P^*_{H_0^*}\left(\sqrt{n}(\overline{X}^*_n - \mu_0) \geq C^W_{n,N-P}(\alpha; \mathbf{X}_n)\right) \cong \alpha.$$

Theorem 4.5.3 *Suppose that $E(|X_1|^3) < \infty$, $E(X_1) = \mu_0$ and $\sigma^2 = \sigma^2(V_1^0)$, the variance of V_1^0. Then, as $n \to \infty$,*

$$P\left(\sqrt{n}(\overline{V}^0_n - \mu_0) \geq C^W_{n,N-P}(\alpha; \mathbf{X}_n)\right) \quad \to \quad 1 - \Phi\left(\frac{z_{1-\alpha}}{\sqrt{2}}\right),$$

where $z_{1-\alpha} = \Phi^{-1}(1-\alpha)$.

Proof. By applying the Berry-Esséen theorem in the bootstrap world (w.r.t. P^*) we deduce that

$$\begin{aligned}
\alpha &\cong P^*_{H_0^*}\left(\sqrt{n}(\overline{X}^*_n - \mu_0) \geq C^W_{n,N-P}(\alpha; \mathbf{X}_n)\right) \\
&= 1 - P^*_{H_0^*}\left(\frac{\sqrt{n}(\overline{X}^*_n - \overline{X}_n)}{S_n(\mathbf{X}_n)} \leq \frac{C^W_{n,N-P}(\alpha; \mathbf{X}_n)}{S_n(\mathbf{X}_n)} - \frac{\sqrt{n}(\overline{X}_n - \mu_0)}{S_n(\mathbf{X}_n)}\right) \\
&= 1 - \Phi\left(\frac{C^W_{n,N-P}(\alpha; \mathbf{X}_n)}{S_n(\mathbf{X}_n)} - \frac{\sqrt{n}(\overline{X}_n - \mu_0)}{S_n(\mathbf{X}_n)}\right) + O_{a.s.}(n^{-1/2}).
\end{aligned}$$

Hence, as $n \to \infty$, we have that

$$C^W_{n,N-P}(\alpha; \mathbf{X}_n) = \sqrt{n}(\overline{X}_n - \mu_0) + S_n(\mathbf{X}_n) z_{1-\alpha} + S_n(\mathbf{X}_n) o_{a.s.}(1). \tag{4.8}$$

Using the Berry-Esséen theorem in the real world (w.r.t. P) we find that

$$\begin{aligned}
\varphi^0(\mathbf{x}_n) &= P\left(\sqrt{n}(\overline{V}^0_n - \mu_0) \geq C^W_{n,N-P}(\alpha; \mathbf{x}_n)\right) \\
&= 1 - \Phi\left(\frac{C^W_{n,N-P}(\alpha; \mathbf{x}_n)}{\sigma}\right) + O(n^{-1/2}).
\end{aligned}$$

From (4.8), the CLT and Kolmogorov's Strong Law of Large Numbers, it follows that

$$\varphi^0(\mathbf{X}_n) \xrightarrow{d} 1 - \Phi(Z + z_{1-\alpha}) \quad \text{as} \quad n \to \infty, \tag{4.9}$$

where Z is a $N(0,1)$ random variable.

Consequently, we conclude from (4.9) and Theorem 5.4 of Billingsley (1968) that

$$\begin{aligned} P\left(\sqrt{n}(\overline{V}_n^0 - \mu_0) \geq C_{n,N-P}^W(\alpha; \mathbf{X}_n)\right) &= E_\mu(\varphi^0(\mathbf{X}_n)) \\ &\to 1 - E(\Phi(Z + z_{1-\alpha})) \\ &= 1 - \Phi\left(\frac{z_{1-\alpha}}{\sqrt{2}}\right). \end{aligned}$$

The last step follows from the fact that

$$E(\Phi(Z + z_{1-\alpha})) = \Phi * \Phi(z_{1-\alpha}).$$

Q.E.D.■

Remark: If we assume in Theorem 4.5.3 that $E(X_1) = \mu_A$, then $P\left(\sqrt{n}(\overline{V}_n^0 - \mu_0) \geq C_{n,P}^W(\alpha; \mathbf{X}_n)\right) \to 0$ or 1, if $\mu_A > \mu_0$ or $\mu_A < \mu_0$, respectively.

Next, suppose the procedure is to reject H_0 if and only if

$$T_{n,P}(\mathbf{X}_n) = \frac{\sqrt{n}(\overline{X}_n - \mu_0)}{S_n(\mathbf{X}_n)} \geq C_{n,P}^W(\alpha; \mathbf{X}_n),$$

where $C_{n,P}^W(\alpha; \mathbf{X}_n)$ is defined by

$$P_{H_0^*}^*\left(\frac{\sqrt{n}(\overline{X}_n^* - \mu_0)}{S_n(\mathbf{X}_n^*)} \geq C_{n,P}^W(\alpha; \mathbf{X}_n)\right) \cong \alpha.$$

Theorem 4.5.4 *Suppose that $E(|X_1|^3) < \infty$ and $E(X_1) = \mu_0$. Then, as $n \to \infty$,*

$$P\left(\frac{\sqrt{n}(\overline{V}_n^0 - \mu_0)}{S_n(\mathbf{V}_n^0)} \geq C_{n,P}^W(\alpha; \mathbf{X}_n)\right) \to 1 - \Phi\left(\frac{z_{1-\alpha}}{\sqrt{2}}\right),$$

where $S_n(\mathbf{V}_n^0)$ is the sample standard deviation based on $\mathbf{V}_n^0 = (V_1^0, V_2^0, \ldots, V_n^0)$.

Proof. From (4.7) we have that

$$C_{n,P}^W(\alpha; \mathbf{X}_n) = \frac{\sqrt{n}(\overline{X}_n - \mu_0)}{S_n(\mathbf{X}_n)} + z_{1-\alpha} + o_{a.s.}(1). \tag{4.10}$$

Using the Berry-Esséen theorem in the real world (w.r.t P) we find that

$$\begin{aligned}\varphi^0(\mathbf{x}_n) &= P\left(\frac{\sqrt{n}(\overline{V}_n^0-\mu_0)}{S_n(\mathbf{V}_n^0)} \geq C_{n,P}^W(\alpha;\mathbf{x}_n)\right) \\ &= 1-\Phi\left(C_{n,P}^W(\alpha;\mathbf{x}_n)\right)+O(n^{-1/2}).\end{aligned}$$

From (4.10) and the CLT it follows that

$$\varphi^0(\mathbf{X}_n) \xrightarrow{d} 1-\Phi(Z+z_{1-\alpha}) \quad \text{as} \quad n\to\infty, \tag{4.11}$$

where Z is a $N(0,1)$ random variable.

Hence, from (4.11) and Theorem 5.4 of Billingsley (1968) it follows that

$$\begin{aligned}P\left(\frac{\sqrt{n}(\overline{V}_n^0-\mu_0)}{S_n(\mathbf{V}_n^0)} \geq C_{n,P}^W(\alpha;\mathbf{X}_n)\right) &= E_\mu(\varphi^0(\mathbf{X}_n)) \\ &\to 1-E(\Phi(Z+z_{1-\alpha})) \\ &= 1-\Phi\left(\frac{z_{1-\alpha}}{\sqrt{2}}\right).\end{aligned}$$

Q.E.D.■

Remark: If we assume in Theorem 4.5.4 that $E(X_1)=\mu_A$, then $P\left(\frac{\sqrt{n}(\overline{V}_n^0-\mu_0)}{S_n(\mathbf{V}_n^0)} \geq C_{n,P}^W(\alpha;\mathbf{X}_n)\right) \to 0$ or 1, if $\mu_A>\mu_0$ or $\mu_A<\mu_0$, respectively.

Lemma 4.1 Suppose that $E(|X_1|^5)<\infty$. Then, as $n\to\infty$,
a) $E(S_n(\mathbf{X}_n)-\sigma) = -\dfrac{(\mu_4+3\sigma^4)}{8n\sigma^3}+O(n^{-2})$,
b) $E((S_n(\mathbf{X}_n)-\sigma)^2) = \dfrac{\mu_4-\sigma^4}{4n\sigma^2}+O(n^{-2})$.

Proof. Using a Taylor series expansion we deduce that

$$E(S_n(\mathbf{X}_n)) = (E(S_n^2(\mathbf{X}_n)))^{1/2}-\frac{1}{8}\cdot\frac{\text{Var}(S_n^2(\mathbf{X}_n))}{(E(S_n^2(\mathbf{X}_n)))^{3/2}}+O(n^{-2}).$$

Applying the result in (27.4.1) and (27.4.2) of Cramér (1945) we conclude that

$$\begin{aligned}E(S_n(\mathbf{X}_n)) &= \sigma\left(1-\frac{1}{n}\right)^{1/2}-\frac{1}{8\sigma^3}\left(1-\frac{1}{n}\right)^{-3/2}\frac{(\mu_4-\sigma^4)}{n}+O(n^{-2}) \\ &= \sigma-\frac{\sigma}{2n}-\frac{1}{8\sigma^3}\frac{(\mu^4-\sigma^4)}{n}+O(n^{-2}) \\ &= \sigma-\frac{(\mu_4+3\sigma^4)}{8n\sigma^3}+O(n^{-2}),\end{aligned}$$

which proves a).

Furthermore, from a) and (27.7.2) of Cramér (1945) it follows that

$$\begin{aligned} E((S_n(\mathbf{X}_n)-\sigma)^2) &= \mathrm{Var}(S_n(\mathbf{X}_n))+\{E(S_n(\mathbf{X}_n))-\sigma\}^2 \\ &= \frac{\mu_4-\sigma^4}{4n\sigma^2}+O(n^{-2}), \end{aligned}$$

which proves b). **Q.E.D.■**

Next, suppose the procedure is to reject H_0 if and only if

$$T_{n,N-P}(\mathbf{X}_n)=\sqrt{n}(\overline{X}_n-\mu_0)\geq C^R_{n,N-P}(\alpha;\mathbf{X}_n),$$

where $C^R_{n,N-P}(\alpha;\mathbf{X}_n)$ is defined by

$$\begin{aligned} & P^*_{H_0^*}\left(\sqrt{n}(\overline{X}^*_n-\overline{X}_n)\geq C^R_{n,N-P}(\alpha;\mathbf{X}_n)\right)\cong\alpha \\ \Leftrightarrow\quad & P^*_{H_0^*}\left(\frac{\sqrt{n}(\overline{X}^*_n-\overline{X}_n)}{S_n(\mathbf{X}_n)}\geq\frac{C^R_{n,N-P}(\alpha;\mathbf{X}_n)}{S_n(\mathbf{X}_n)}\right)\cong\alpha. \end{aligned}$$

From Hall (1992) it follows that a Cornish-Fisher expansion for $C^R_{n,N-P}(\alpha;\mathbf{X}_n)$ is given by

$$C^R_{n,N-P}(\alpha;\mathbf{X}_n)=S_n(\mathbf{X}_n)\{z_{1-\alpha}+n^{-1/2}\hat{P}^{cf}_1(z_{1-\alpha})+n^{-1}\hat{P}^{cf}_2(z_{1-\alpha})+O_P(n^{-3/2})\} \tag{4.12}$$

where $z_{1-\alpha}=\Phi^{-1}(1-\alpha)$, and

$$\hat{P}^{cf}_1(z_{1-\alpha})=-\hat{P}_1(z_{1-\alpha}); \tag{4.13}$$

$$\hat{P}^{cf}_2(z_{1-\alpha})=\hat{P}_1(z_{1-\alpha})\hat{P}'_1(z_{1-\alpha})-\frac{1}{2}z_{1-\alpha}(\hat{P}_1(z_{1-\alpha}))^2-\hat{P}_2(z_{1-\alpha}); \tag{4.14}$$

$$\hat{P}_1(z_{1-\alpha})=\hat{\kappa}_3(1-z^2_{1-\alpha})/6;\quad \hat{P}'_1(z_{1-\alpha})=-\hat{\kappa}_3 z_{1-\alpha}/3; \tag{4.15}$$

$$\hat{P}_2(z_{1-\alpha})=-z_{1-\alpha}\left\{\frac{\hat{\kappa}_4(z^2_{1-\alpha}-3)}{24}+\frac{\hat{\kappa}^2_3(z^4_{1-\alpha}-10z^2_{1-\alpha}+15)}{72}\right\}; \tag{4.16}$$

$$\hat{\kappa}_3=S_n^{-3}(\mathbf{X}_n)n^{-1}\sum_{i=1}^n(X_i-\overline{X}_n)^3;\quad \hat{\kappa}_4=S_n^{-4}(\mathbf{X}_n)n^{-1}\sum_{i=1}^n(X_i-\overline{X}_n)^4-3;$$

$$S^2_n(\mathbf{X}_n)=n^{-1}\sum_{i=1}^n(X_i-\overline{X}_n)^2.$$

Throughout the discussion below, $\hat{P}$ denotes the Lebesgue-Stieltjes probability measure determined by the distribution function F of the X_i's.

Theorem 4.5.5 *Suppose* $E(|X_1|^5) < \infty$, $\sigma^2 = \sigma^2(V_1^0)$ *and that Cramér's continuity condition holds, i.e.,*

$$\limsup_{|t|\to\infty} |\chi(t)| < 1,$$

where $\chi(t)$ *is the characteristic function of* X_1. *Then, as* $n \to \infty$,

$$P\left(\sqrt{n}(\overline{V}_n^0 - \mu_0) \geq C_{n,N-P}^R(\alpha; \mathbf{X}_n)\right) = \alpha + C_n + O(n^{-3/2}),$$

where

$$C_n = \frac{z_{1-\alpha}\phi(z_{1-\alpha})\{z_{1-\alpha}^2(\mu_4 - \sigma^4) + (\mu_4 + 3\sigma^4)\}}{8\sigma^4 n}.$$

Proof. Applying an Edgeworth expansion (see, e.g., Hall, 1992) we deduce that

$$\begin{aligned}
\varphi^0(\mathbf{x}_n) &= P\left(\sqrt{n}(\overline{V}_n^0 - \mu_0) \geq C_{n,N-P}^R(\alpha; \mathbf{x}_n)\right) \\
&= 1 - P\left(\frac{\sqrt{n}(\overline{V}_n^0 - \mu_0)}{\sigma} \leq \frac{C_{n,N-P}^R(\alpha; \mathbf{x}_n)}{\sigma}\right) \\
&= 1 - \left\{\Phi\left(\frac{C_{n,N-P}^R(\alpha; \mathbf{x}_n)}{\sigma}\right) + n^{-1/2} R_1\left(\frac{C_{n,N-P}^R(\alpha; \mathbf{x}_n)}{\sigma}\right)\right. \\
&\quad \left. + n^{-1} R_2\left(\frac{C_{n,N-P}^R(\alpha; \mathbf{x}_n)}{\sigma}\right) + O_{\hat{P}}(n^{-3/2})\right\}, \qquad (4.17)
\end{aligned}$$

where

$$R_j(x) = P_j(x)\phi(x); \qquad (4.18)$$

$$P_1(x) = \kappa_3(1 - x^2)/6; \quad \kappa_3 = E((X_1 - \mu)^3)/\sigma^3; \qquad (4.19)$$

$$P_2(x) = -x\left\{\frac{\kappa_4(x^2 - 3)}{24} + \frac{\kappa_3^2(x^4 - 10x^2 + 15)}{72}\right\}; \quad \kappa_4 = \frac{E((X_1 - \mu)^4)}{\sigma^4} - 3. \qquad (4.20)$$

Let

$$c_{n,\alpha}(\mathbf{x}_n) = z_{1-\alpha} + n^{-1/2}\hat{P}_1^{cf}(z_{1-\alpha}) + n^{-1}\hat{P}_2^{cf}(z_{1-\alpha}) + O_{\hat{P}}(n^{-3/2}). \qquad (4.21)$$

Using (4.12), (4.22) and a Taylor series expansion, we find that

$$\begin{aligned}
\Phi\left(\frac{C_{n,N-P}^R(\alpha; \mathbf{x}_n)}{\sigma}\right) &= \Phi\left(\frac{S_n(\mathbf{x}_n)}{\sigma} \cdot c_{n,\alpha}(\mathbf{x}_n)\right) \\
&= \Phi(c_{n,\alpha}(\mathbf{x}_n)) + (c_{n,\alpha}(\mathbf{x}_n)/\sigma)\phi(c_{n,\alpha}(\mathbf{x}_n))(S_n(\mathbf{x}_n) - \sigma) \\
&\quad - \sigma(c_{n,\alpha}(\mathbf{x}_n)/\sigma)^3\phi(c_{n,\alpha}(\mathbf{x}_n))(S_n(\mathbf{x}_n) - \sigma)^2/2 + O_{\hat{P}}(n^{-3/2}) \\
&= 1 - \alpha + \phi(z_{1-\alpha})\left\{\hat{P}_1^{cf}(z_{1-\alpha})n^{-1/2} + \hat{P}_2^{cf}(z_{1-\alpha})n^{-1}\right\} \\
&\quad - z_{1-\alpha}\phi(z_{1-\alpha})\{\hat{P}_1^{cf}(z_{1-\alpha})\}^2 n^{-1}/2 \\
&\quad + (z_{1-\alpha}/\sigma)\phi(z_{1-\alpha})(S_n(\mathbf{x}_n) - \sigma) \\
&\quad - \sigma(z_{1-\alpha}/\sigma)^3\phi(z_{1-\alpha})(S_n(\mathbf{x}_n) - \sigma)^2/2 + O_{\hat{P}}(n^{-3/2}). \qquad (4.22)
\end{aligned}$$

From (4.19), (4.22) and Taylor's theorem it easily follows that

$$\begin{aligned}
R_1\left(\frac{C^R_{n,N-P}(\alpha;\mathbf{x}_n)}{\sigma}\right) &= R_1\left(\frac{S_n(\mathbf{x}_n)}{\sigma}\cdot c_{n,\alpha}(\mathbf{x}_n)\right) \\
&= R_1(c_{n,\alpha}(\mathbf{x}_n)) + R_1'(c_{n,\alpha}(\mathbf{x}_n))c_{n,\alpha}(\mathbf{x}_n)(S_n(\mathbf{x}_n)-\sigma)/\sigma + O_{\hat{P}}(n^{-1}) \\
&= P_1(z_{1-\alpha})\phi(z_{1-\alpha}) + n^{-1/2}\hat{P}_1^{cf}(z_{1-\alpha})\phi(z_{1-\alpha})\times \\
&\quad \{P_1'(z_{1-\alpha}) - z_{1-\alpha}P_1(z_{1-\alpha})\} + O_{\hat{P}}(n^{-1}). \qquad (4.23)
\end{aligned}$$

A first-order Taylor series expansion also yields

$$\begin{aligned}
R_2\left(\frac{C^R_{n,N-P}(\alpha;\mathbf{x}_n)}{\sigma}\right) &= R_2\left(\frac{S_n(\mathbf{x}_n)}{\sigma}.c_{n,\alpha}(\mathbf{x}_n)\right) \\
&= P_2(z_{1-\alpha})\phi(z_{1-\alpha}) + O_{\hat{P}}(n^{-1/2}). \qquad (4.24)
\end{aligned}$$

From (4.13) – (4.21) and (4.23) – (4.25) we find, after tedious calculations, that

$$\begin{aligned}
\varphi^0(\mathbf{X}_n) &= \alpha - (z_{1-\alpha}/\sigma)\phi(z_{1-\alpha})(S_n(\mathbf{X}_n)-\sigma) + \sigma(z_{1-\alpha}/\sigma)^3\phi(z_{1-\alpha})(S_n(\mathbf{X}_n)-\sigma)^2/2 \\
&\quad - n^{-1/2}(\hat{\kappa}_3-\kappa_3)(z_{1-\alpha}^2-1)\phi(z_{1-\alpha})/6 \\
&\quad - n^{-1}(\hat{\kappa}_3-\kappa_3)\hat{P}_1(z_{1-\alpha})\phi(z_{1-\alpha})z_{1-\alpha}(z_{1-\alpha}^2-3)/6 \\
&\quad + n^{-1}\{\hat{P}_2(z_{1-\alpha}) - P_2(z_{1-\alpha})\}\phi(z_{1-\alpha}) + O_P(n^{-3/2}). \qquad (4.25)
\end{aligned}$$

From the results in Chapter 7 of Cramér (1945) it follows that

$$E(\hat{\kappa}_3) = \kappa_3 + O(n^{-1}), \quad E(\hat{\kappa}_4) = \kappa_4 + O(n^{-1}). \qquad (4.26)$$

Hence, from Lemma 4.1, (4.26) and (4.27) we conclude that

$$\begin{aligned}
P\left(\sqrt{n}(\overline{V}^0_n - \mu_0) \geq C^R_{n,N-P}(\alpha;\mathbf{X}_n)\right) &= E_\mu(\varphi^0(\mathbf{X}_n)) \\
&= \alpha + (z_{1-\alpha}/\sigma)\phi(z_{1-\alpha})(\mu_4+3\sigma^4)/8n\sigma^3 \\
&\quad + \sigma(z_{1-\alpha}/\sigma)^3\phi(z_{1-\alpha})(\mu_4-\sigma^4)/8n\sigma^2 \\
&\quad + O(n^{-3/2}) \\
&= \alpha + \frac{z_{1-\alpha}\phi(z_{1-\alpha})}{8n\sigma^4}\{z_{1-\alpha}^2(\mu_4-\sigma^4) + (\mu_4+3\sigma^4)\} \\
&\quad + O(n^{-3/2}),
\end{aligned}$$

which completes the proof of the theorem. **Q.E.D.■**

Remark: It is important to note that $C_n \geq 0$ in Theorem 4.5.5.

Lemma 4.2 Suppose that $E(|X_1|^6) < \infty$. Then we have that as $n \to \infty$,

$$E(\hat{\kappa}_3) = \kappa_3 + \frac{\mu_3(21\sigma^4 + 15\mu_4) - 12\sigma^2\mu_5}{8\sigma^7 n} + O(n^{-2}).$$

Proof. Using Taylor's theorem we deduce that

$$\begin{aligned} E(\hat{\kappa}_3) &= \frac{E(m_3)}{(E(m_2))^{3/2}} + \frac{15}{8} \cdot \frac{E(m_3)}{(E(m_2))^{7/2}} \cdot \mathrm{Var}(m_2) \\ &- \frac{3}{2} \cdot \frac{1}{(E(m_2))^{5/2}} \cdot \mathrm{Cov}(m_2, m_3) + O(n^{-2}), \end{aligned} \tag{4.27}$$

where $m_3 = n^{-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)^3$ and $m_2 = S_n^2(\mathbf{X}_n)$.

Note that from (27.4.5) and (27.5.1) of Cramér (1945) and Taylor's theorem we find that

$$\begin{aligned} \frac{E(m_3)}{(E(m_2))^{3/2}} &= \frac{1}{\sigma^3}\left(1 - \frac{1}{n}\right)^{-3/2}\left(1 - \frac{3}{n}\right)\mu_3 + O(n^{-2}) \\ &= \left(\frac{\mu_3}{\sigma^3}\right)\left(1 - \frac{3}{2n}\right) + O(n^{-2}). \end{aligned} \tag{4.28}$$

Equations (27.4.2) and (27.5.6) of Cramér (1945) also reveal that

$$\mathrm{Var}(m_2) = \frac{\mu_4 - \sigma^4}{n} + O(n^{-2}), \tag{4.29}$$

and

$$\mathrm{Cov}(m_2, m_3) = \frac{\mu_5 - 4\mu_3\sigma^2}{n} + O(n^{-2}). \tag{4.30}$$

The proof of the Lemma now follows directly by substituting (4.29), (4.30) and (4.31) into (4.28). **Q.E.D.■**

Lastly, suppose the procedure is to reject H_0 if and only if

$$T_{n,P}(\mathbf{X}_n) = \frac{\sqrt{n}(\overline{X}_n - \mu_0)}{S_n(\mathbf{X}_n)} \geq C^R_{n,P}(\alpha; \mathbf{X}_n),$$

where $C^R_{n,P}(\alpha; \mathbf{X}_n)$ is defined by

$$P^*_{H_0^*}\left(\frac{\sqrt{n}(\overline{X}^*_n - \overline{X}_n)}{S_n(\mathbf{X}^*_n)} \geq C^R_{n,P}(\alpha; \mathbf{X}_n)\right) \cong \alpha.$$

From Hall (1992) it follows that a Cornish-Fisher expansion for $C^R_{n,P}(\alpha; \mathbf{X}_n)$ is given by

$$C^R_{n,P}(\alpha; X_n) = z_{1-\alpha} + n^{-1/2}\hat{Q}^{cf}_1(z_{1-\alpha}) + n^{-1}\hat{Q}^{cf}_2(z_{1-\alpha}) + \dots \tag{4.31}$$

where

$$\hat{Q}^{cf}_1(z_{1-\alpha}) = -\hat{Q}_1(z_{1-\alpha}); \tag{4.32}$$

$$\hat{Q}_2^{cf}(z_{1-\alpha}) = \hat{Q}_1(z_{1-\alpha})\hat{Q}_1'(z_{1-\alpha}) - \frac{1}{2}z_{1-\alpha}(\hat{Q}_1(z_{1-\alpha}))^2 - \hat{Q}_2(z_{1-\alpha}); \tag{4.33}$$

$$\hat{Q}_1(z_{1-\alpha}) = \frac{1}{6}\hat{\kappa}_3(1+2z_{1-\alpha}^2); \quad \hat{Q}_1'(z_{1-\alpha}) = \frac{2}{3}\hat{\kappa}_3 z_{1-\alpha}; \tag{4.34}$$

$$\hat{Q}_2(z_{1-\alpha}) = z_{1-\alpha}\left\{\frac{1}{12}\hat{\kappa}_4(z_{1-\alpha}^2 - 3) - \frac{1}{18}\hat{\kappa}_3^2(z_{1-\alpha}^4 + 2z_{1-\alpha}^2 - 3) - \frac{1}{4}(z_{1-\alpha}^2 + 3)\right\}. \tag{4.35}$$

Theorem 4.5.6 *Suppose that $E(X_1^6) < \infty$ and that Cramér's continuity condition holds. Then, as $n \to \infty$, we have that*

$$P\left(\frac{\sqrt{n}(\overline{V}_n^0 - \mu_0)}{S_n(\mathbf{V}_n^0)} \geq C_{n,P}^R(\alpha;\mathbf{X}_n)\right) = \alpha + D_n + O(n^{-2}),$$

where

$$D_n = \frac{(1+2z_{1-\alpha}^2)\phi(z_{1-\alpha})\{\mu_3(21\sigma^4 + 15\mu_4) - 12\sigma^2\mu_5\}}{48\sigma^7 n^{3/2}}.$$

Proof. Applying an Edgeworth expansion (see, e.g., Hall, 1992) we derive that

$$\begin{aligned}
\varphi^0(\mathbf{x}_n) &= P\left(\frac{\sqrt{n}(\overline{V}_n^0 - \mu_0)}{S_n(\mathbf{V}_n^0)} \geq C_{n,P}^R(\alpha;\mathbf{x}_n)\right) \\
&= 1 - P\left(\frac{\sqrt{n}(\overline{V}_n^0 - \mu_0)}{S_n(\mathbf{V}_n^0)} \leq C_{n,P}^R(\alpha;\mathbf{x}_n)\right) \\
&= 1 - \{\Phi(C_{n,P}^R(\alpha;\mathbf{x}_n)) + n^{-1/2}Q_1(C_{n,P}^R(\alpha;\mathbf{x}_n))\phi(C_{n,P}^R(\alpha;\mathbf{x}_n)) \\
&\quad + n^{-1}Q_2(C_{n,P}^R(\alpha;\mathbf{x}_n))\phi(C_{n,P}^R(\alpha;\mathbf{x}_n)) + \ldots\},
\end{aligned} \tag{4.36}$$

where

$$Q_1(x) = \kappa_3(1+2x^2)/6, \tag{4.37}$$

and

$$Q_2(x) = x\left\{\frac{1}{12}\kappa_4(x^2-3) - \frac{1}{18}\kappa_3^2(x^4+2x^2-3) - \frac{1}{4}(x^2+3)\right\}. \tag{4.38}$$

Using (4.32) and a Taylor series expansion, we find that

$$\begin{aligned}
\Phi(C_{n,P}^R(\alpha;\mathbf{x}_n)) &= 1 - \alpha + \phi(z_{1-\alpha})\{n^{-1/2}\hat{Q}_1^{cf}(z_{1-\alpha}) + n^{-1}\hat{Q}_2^{cf}(z_{1-\alpha}) + \ldots\} \\
&\quad - z_{1-\alpha}\phi(z_{1-\alpha})(\hat{Q}_1^{cf})^2 n^{-1}/2.
\end{aligned} \tag{4.39}$$

From (4.32) and Taylor's theorem it follows that

$$\begin{aligned}
&Q_1(C_{n,P}^R(\alpha;\mathbf{x}_n))\phi(C_{n,P}^R(\alpha;\mathbf{x}_n)) \\
= \;& Q_1(z_{1-\alpha})\phi(z_{1-\alpha}) + \{Q_1'(z_{1-\alpha}) - z_{1-\alpha}Q_1(z_{1-\alpha})\}\phi(z_{1-\alpha})n^{-1/2}\hat{Q}_1^{cf}(z_{1-\alpha}) + \ldots
\end{aligned} \tag{4.40}$$

A first-order Taylor series expansion also yields

$$Q_2(C^R_{n,P}(\alpha;\mathbf{x}_n))\phi(C^R_{n,P}(\alpha;\mathbf{x}_n)) = Q_2(z_{1-\alpha})\phi(z_{1-\alpha}) + \dots \tag{4.41}$$

From (4.33) – (4.42) we find, after laborious calculations, that

$$\begin{aligned}\varphi^0(\mathbf{X}_n) &= \alpha + \frac{(1+2z^2_{1-\alpha})\phi(z_{1-\alpha})(\hat\kappa_3-\kappa_3)}{6n^{1/2}}\\ &\quad - n^{-1}(\hat\kappa_3-\kappa_3)\hat Q_1(z_{1-\alpha})\phi(z_{1-\alpha})z_{1-\alpha}(3-2z^2_{1-\alpha})/6\\ &\quad + n^{-1}\{\hat Q_2(z_{1-\alpha}) - Q_2(z_{1-\alpha})\}\phi(z_{1-\alpha}) + \dots \end{aligned}\tag{4.42}$$

Hence, from Lemma 4.2, (4.27) and (4.43) we conclude that

$$\begin{aligned}P\left(\frac{\sqrt{n}(\overline{V}^0_n-\mu_0)}{S_n(\mathbf{V}^0_n)} \geq C^R_{n,P}(\alpha;\mathbf{X}_n)\right) &= E_\mu(\varphi^0(\mathbf{X}_n))\\ &= \alpha + \frac{(1+2z^2_{1-\alpha})\phi(z_{1-\alpha})\{\mu_3(21\sigma^4+15\mu_4)-12\sigma^2\mu_5\}}{48\sigma^7 n^{3/2}}\\ &\quad + O(n^{-2}),\end{aligned}\tag{4.43}$$

which completes the proof of the theorem. **Q.E.D.■**

4.5.4 Monte-Carlo study: Method II

In this section we present the results of a Monte-Carlo study, based on Method II, in order to compare the sizes of the tests based on (R,P), (W,P), (R,N-P) and (W,N-P) for sample sizes $n = 20, 30, 50, 100$ and 200, respectively. A nominal significance level of $\alpha = 0.05$ was used throughout and data were again generated from five different distributions:

- Normal distribution with mean μ and variance 1;
- Uniform distribution with mean μ and variance 1/12;
- Double exponential distribution with mean μ and variance 2;
- Exponential distribution with mean μ and variance 1;
- Chi-squared distribution with mean μ and variance 6.

The column label "Data from" in the tables that follow refers to the training sample $\mathbf{X}_n$ generated from a distribution with the specified parameter. The μ_0 column refers to the pseudo test data $\mathbf{V}^0_n$ generated from a distribution with parameter μ_0.

Size estimates ($\overline{p}$) were calculated according to

$$\overline{p} = \frac{1}{MC_1} \sum_{j=1}^{MC_1} p_j,$$

where

$$p_j = \frac{1}{MC_2} \sum_{i=1}^{MC_2} I\left(T_n(\mathbf{V}_{n,i}^0) \geq \hat{C}_n(\alpha; \mathbf{X}_n^j)\right), \ j = 1, 2, \ldots, MC_1.$$

The statistic T_n was replaced by $T_{n,P}$ or $T_{n,N-P}$, while $\hat{C}_n(\alpha; \mathbf{X}_n^j)$ was replaced by $\hat{C}_{n,P}^R(\alpha, \mathbf{X}_n^j)$, $\hat{C}_{n,P}^W(\alpha, \mathbf{X}_n^j)$, $\hat{C}_{n,N-P}^R(\alpha, \mathbf{X}_n^j)$ or $\hat{C}_{n,N-P}^W(\alpha, \mathbf{X}_n^j)$, for independent training samples $\mathbf{X}_n^j$, $j = 1, 2, \ldots, MC_1$.

The standard errors of the estimated sizes were calculated according to

$$\sqrt{\frac{\frac{1}{MC_1} \sum_{j=1}^{MC_1} (p_j - \overline{p})^2}{MC_1}}.$$

In our Monte-Carlo study we took $MC_1 = 15\ 000$ and $MC_2 = 10\ 000$. All standard errors were found to be negligibly small and are not reported in the tables. Bootstrap critical values were again based on $B = 1000$ independent replications.

Table 4.2: Estimated size of bootstrap hypothesis tests for the mean using Method II (**normal distribution**).

		Data from: $\mu = 0.0$				Data from: $\mu = 0.2$			
μ_0	n	(R,P)	(W,P)	(R,N-P)	(W,N-P)	(R,P)	(W,P)	(R,N-P)	(W,N-P)
	20	0.051	0.130	0.063	0.136	0.051	0.040	0.064	0.044
	30	0.051	0.127	0.059	0.131	0.051	0.029	0.059	0.031
0.0	50	0.051	0.125	0.056	0.128	0.051	0.017	0.056	0.018
	100	0.050	0.124	0.053	0.126	0.051	0.005	0.053	0.005
	200	0.050	0.122	0.052	0.123	0.050	0.001	0.052	0.001

Table 4.3: Estimated size of bootstrap hypothesis tests for the mean using Method II (**double exponential distribution**).

		Data from: $\mu = 0.0$				Data from: $\mu = 0.2$			
μ_0	n	(R,P)	(W,P)	(R,N-P)	(W,N-P)	(R,P)	(W,P)	(R,N-P)	(W,N-P)
	20	0.050	0.131	0.072	0.144	0.050	0.052	0.072	0.066
	30	0.051	0.130	0.066	0.139	0.050	0.042	0.066	0.052
0.0	50	0.051	0.127	0.060	0.132	0.051	0.029	0.060	0.034
	100	0.050	0.126	0.055	0.128	0.050	0.014	0.055	0.016
	200	0.050	0.126	0.053	0.128	0.050	0.004	0.053	0.005

Table 4.4: Estimated size of bootstrap hypothesis tests for the mean using Method II (**uniform distribution**).

		Data from: $\mu = 0.5$				Data from: $\mu = 0.6$			
μ_0	n	(R,P)	(W,P)	(R,N-P)	(W,N-P)	(R,P)	(W,P)	(R,N-P)	(W,N-P)
	20	0.051	0.132	0.060	0.134	0.051	0.017	0.060	0.016
	30	0.051	0.129	0.057	0.130	0.051	0.007	0.056	0.007
0.5	50	0.051	0.127	0.054	0.127	0.051	0.002	0.054	0.002
	100	0.051	0.121	0.052	0.121	0.051	0.000	0.053	0.000
	200	0.050	0.122	0.052	0.122	0.050	0.000	0.052	0.000

Table 4.5: Estimated size of bootstrap hypothesis tests for the mean using Method II (**exponential distribution**).

		Data from: $\mu = 1.0$				Data from: $\mu = 1.2$			
μ_0	n	(R,P)	(W,P)	(R,N-P)	(W,N-P)	(R,P)	(W,P)	(R,N-P)	(W,N-P)
1.0	20	0.043	0.159	0.080	0.172	0.043	0.047	0.080	0.071
	30	0.046	0.155	0.072	0.165	0.046	0.029	0.071	0.047
	50	0.048	0.146	0.064	0.151	0.047	0.014	0.064	0.025
	100	0.049	0.140	0.058	0.143	0.049	0.004	0.058	0.009
	200	0.050	0.133	0.054	0.134	0.050	0.000	0.054	0.000

Table 4.6: Estimated size of bootstrap hypothesis tests for the mean using Method II (**chi-squared distribution**).

		Data from: $\mu = 3.0$				Data from: $\mu = 3.2$			
μ_0	n	(R,P)	(W,P)	(R,N-P)	(W,N-P)	(R,P)	(W,P)	(R,N-P)	(W,N-P)
3.0	20	0.045	0.154	0.075	0.165	0.044	0.099	0.076	0.116
	30	0.046	0.149	0.068	0.157	0.046	0.083	0.067	0.010
	50	0.048	0.143	0.061	0.147	0.048	0.066	0.062	0.076
	100	0.050	0.137	0.056	0.139	0.050	0.044	0.056	0.051
	200	0.050	0.130	0.054	0.131	0.050	0.024	0.054	0.029

Table 4.7 presents the values of the constant C_n (defined in Theorem 4.5.5) and the constant D_n (defined in Theorem 4.5.6) for the five distributions under consideration.

Table 4.7: Values of the constants C_n and D_n.

Distribution	n	C_n	D_n
Normal Mean: 0.0 Variance: 1.0	20	0.012	0.000
	30	0.008	0.000
	50	0.005	0.000
	100	0.002	0.000
	200	0.001	0.000
Uniform Mean: 0.5 Variance: 1/12	20	0.007	0.000
	30	0.005	0.000
	50	0.003	0.000
	100	0.001	0.000
	200	0.001	0.000
Double exponential Mean: 0.0 Variance: 2.0	20	0.035	0.000
	30	0.024	0.000
	50	0.014	0.000
	100	0.007	0.000
	200	0.003	0.000
Exponential Mean: 1.0 Variance: 1.0	20	0.036	-0.033
	30	0.024	-0.018
	50	0.014	-0.008
	100	0.007	-0.003
	200	0.004	-0.001
Chi-squared Mean: 3.0 Variance: 6.0	20	0.027	-0.021
	30	0.018	-0.012
	50	0.011	-0.005
	100	0.005	-0.002
	200	0.003	-0.001

Various conclusions can be drawn from Tables 4.2 – 4.7:

(1) The estimated sizes of (W,N-P) and (W,P) converge to $1 - \Phi(\frac{1.645}{\sqrt{2}}) \approx 0.123$ as n increases, when the data $\mathbf{X}_n$ are generated from a distribution with the parameter specified by the *null hypothesis*. This agrees with the results of Theorem 4.5.3 and Theorem 4.5.4.

(2) The estimated sizes of (W,N-P) and (W,P) converge to 0 as n increases, when the data $\mathbf{X}_n$ are generated from a distribution with the parameter specified by the *alternative hypothesis*.

(3) It is clear that the tests (R,P) and (R,N-P) produce the same results when the data are generated from a distribution with the parameter specified by the null and alternative hypotheses. These findings are not surprising, because both $C^R_{n,P}(\alpha; \mathbf{X}_n)$ and $C^R_{n,N-P}(\alpha; \mathbf{X}_n)$ are location-invariant.

(4) The estimated sizes of (R,N-P) decreases monotonically to the nominal significance level as n increases. This is in accordance with the result of Theorem 4.5.5, where it was shown that $C_n \geq 0$. This test, therefore, tends to be "liberal". It is clear from Table 4.1, that if we use evaluation Method I, the estimated sizes of (R,N-P) do not necessarily decrease to the nominal significance level as $n \to \infty$ (see, e.g., the results from the exponential distribution).

(5) For the *symmetrical distributions* (Tables 4.2–4.4) the constant D_n (defined in Theorem 4.5.6) is equal to zero. The estimated sizes of (R,P) for these symmetrical distributions attain the nominal significance level, even for *small* values of n.

(6) For the *asymmetric distributions* (Table 4.5 and 4.6) we find that, while the estimated sizes of (R,P) are slightly more conservative than their symmetric counterparts, they are still close to the nominal significance level, even for small sample sizes.

(7) From Tables 4.2 – 4.6 we can see that the estimated sizes of (R,P) appear to converge much quicker to the nominal significance level than the test (R,N-P). This agrees with the results of Theorem 4.5.5 and Theorem 4.5.6, where it was shown that

$$P\left(\sqrt{n}(\overline{V}^0_n - \mu_0) \geq C^R_{n,N-P}(\alpha; \mathbf{X}_n)\right) = \alpha + O(n^{-1}),$$

and

$$P\left(\frac{\sqrt{n}(\overline{V}^0_n - \mu_0)}{S_n(\mathbf{V}^0_n)} \geq C^R_{n,P}(\alpha; \mathbf{X}_n)\right) = \alpha + O(n^{-3/2}).$$

Next, we present the results of a Monte-Carlo study, again based on Method II, in order to evaluate the robustness of the tests (R,P) and (R,N-P).

Data were generated in the following three ways:

- $\mathbf{V}_n^0$ was generated from a $N(0,1)$ distribution, while $\mathbf{X}_n$ was generated from a $N(0,4)$ distribution. We tested the hypothesis

$$H_0 : \mu = 0 \quad \text{vs.} \quad H_A : \mu > 0.$$

- $\mathbf{V}_n^0$ was generated from an exponential distribution with mean and variance 1, while $\mathbf{X}_n$ was generated from an exponential distribution with mean 1.2 and variance 1.2^2. We tested the hypothesis

$$H_0 : \mu = 1 \quad \text{vs.} \quad H_A : \mu > 1.$$

- $\mathbf{V}_n^0$ was generated from a $N(0,1)$ distribution, while $\mathbf{X}_n$ was generated from a contaminated normal distribution defined by $0.6\Phi(x) + 0.4\Phi(\frac{x}{\sqrt{2}})$. We tested the hypothesis

$$H_0 : \mu = 0 \quad \text{vs.} \quad H_A : \mu > 0.$$

Table 4.8: Estimated sizes.

	Normal		Exponential		Con-Normal	
n	(R,P)	(R,N-P)	(R,P)	(R,N-P)	(R,P)	(R,N-P)
20	0.051	0.003	0.044	0.051	0.049	0.038
30	0.051	0.002	0.046	0.046	0.051	0.036
50	0.050	0.001	0.047	0.038	0.051	0.031
100	0.050	0.001	0.050	0.031	0.050	0.028
200	0.050	0.000	0.050	0.028	0.050	0.028

From the results in Table 4.8 it is clear that (R,N-P) performs poorly with regard to estimated sizes. This can, among others, be ascribed to the fact that (R,N-P) is not scale-invariant. This defect of (R,N-P) cannot be observed if one uses evaluation Method I. In contrast with this, (R,P) behaves satisfactorily. This is to be expected in the normal and exponential cases, since (R,P) is location and scale-invariant. However, the good performance of (R,P) in the contaminated normal case shows that it is robust with respect to contamination of the $\mathbf{X}_n$ data.

Concluding remarks:

- It is clear from the discussions in this section, that it is preferable to make use of a pivotal test statistic and the "right" bootstrap critical value. These findings are in agreement with the two guidelines proposed by Hall and Wilson (1991).

Chapter 5

Resampling residuals to apply the bootstrap to hypothesis testing

In this chapter we present a general formulation for resampling residuals when applying the bootstrap methodology to hypothesis testing. The chapter concludes with the application of this technique to three common testing scenarios.

5.1 Introduction

When working with complicated data sets (like regression models), transforming the data in order to "mimic" H_0 (or H_A) can be very difficult. An alternative way is to resample residuals, and then make use of the resampled residuals to build a model that conforms to H_0 (or H_A). Examples of where this technique is applied to perform bootstrap-based testing are especially common in the econometric literature. Some of these applications include:

- bootstrapping p-values and power in the first-order autoregression model (Rayner, 1990);
- the level and power of the bootstrap $t-$test in the AR(1)-model with trend (Nankervis and Savin, 1996);
- bootstrap testing in nonlinear models (Davidson and MacKinnon, 1999);
- implementing the bootstrap in static or stable dynamic regression models (van Giesbergen and Kiviet, 2002);
- bootstrap unit root tests (Park, 2003); and

- testing for autoregressive conditional heteroscedasticity (Davidson and MacKinnon, 2007).

In Section 5.2 a general formulation, based on two models (Model I and Model II), will be provided to show how this resampling scheme can be implemented *correctly*. Some of the notation used when discussing Model I in this section is similar to that found in Paparoditis and Politis (2005).

5.2 Model based method

We will first provide a general formulation based on Model I, followed by a formulation based on Model II.

5.2.1 General formulation (Model I)

Consider data $\big((\mathbf{X}_1, Y_1), (\mathbf{X}_2, Y_2), \ldots, (\mathbf{X}_n, Y_n)\big)$. Suppose we are interested in testing hypotheses concerning parameters from the following model:

$$Y_t = G(\boldsymbol{X}_t, \boldsymbol{\beta}, \boldsymbol{Y}_t^L, \boldsymbol{\phi}) + \varepsilon_t, \quad t = p+1, p+2, \ldots, n,$$

where

- $\mathbf{Y}_n = (Y_1, Y_2, \ldots, Y_n)^T$ are the response variables;
- $\{\varepsilon_t\}$ are i.i.d. random errors from an unknown distribution F, such that $\mathrm{E}(\varepsilon_t) = 0$ and $\mathrm{E}(\varepsilon_t^2) = \sigma_\varepsilon^2$;
- $\boldsymbol{X}_t = (X_{t,1}, X_{t,2}, \ldots, X_{t,k+m})$ are regressors (which can be fixed or random);
- $\boldsymbol{\beta} = (\boldsymbol{\theta}, \boldsymbol{\nu}) \in \Pi = \Theta \times N \subset \mathbb{R}^k \times \mathbb{R}^m$ is a set of parameters associated with $\mathbf{X}_t$;
- $\boldsymbol{Y}_t^L = (Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p})$ are the lagged values of the response variables;
- $\boldsymbol{\phi} = (\boldsymbol{\rho}, \boldsymbol{\gamma}) \in \Omega = P \times \Gamma \subset \mathbb{R}^q \times \mathbb{R}^s$ is a set of parameters associated with $\mathbf{Y}_t^L$, $q + s = p$; and
- G is some known function.

Let $\boldsymbol{\theta}$ be the parameter of interest and suppose we wish to test the hypothesis:

$$H_0 : \boldsymbol{r}'\boldsymbol{\theta} = \boldsymbol{r}'\boldsymbol{\theta}_0 \qquad \text{vs.} \qquad H_A : \boldsymbol{r}'\boldsymbol{\theta} > \boldsymbol{r}'\boldsymbol{\theta}_0,$$

where $\boldsymbol{r}$ is a known k-dimensional column vector and $\boldsymbol{\nu}$,$\boldsymbol{\rho}$ and $\boldsymbol{\gamma}$ are a set of m, q and s nuisance parameters, respectively.

The test rejects H_0 if and only if $T_n(\boldsymbol{Y}_n, \chi_n) \geq C_n(\alpha)$, where

$$P_{H_0}(T_n(\mathbf{Y}_n, \chi_n) \geq C_n(\alpha)) \cong \alpha$$

and $\chi_n = (\boldsymbol{X}_{p+1}, \boldsymbol{X}_{p+2}, \ldots, \boldsymbol{X}_n)$.

The critical value $C_n(\alpha)$ is unknown, since F is unknown. However, a bootstrap estimator $C_n(\alpha; \mathbf{Y}_n, \chi_n)$ can be derived for $C_n(\alpha)$.

Bootstrap algorithm

The following residual-based bootstrap algorithm can be used to find the bootstrap estimator $C_n(\alpha; \boldsymbol{Y}_n, \chi_n)$:

i) Let $\hat{\boldsymbol{\beta}}_n = (\hat{\boldsymbol{\theta}}_n, \hat{\boldsymbol{\nu}}_n) \in \Pi$ and $\hat{\boldsymbol{\phi}}_n = (\hat{\boldsymbol{\rho}}_n, \hat{\gamma}_n) \in \Omega$ be estimators for $\boldsymbol{\beta}$ and $\boldsymbol{\phi}$, respectively.

ii) Define centered residuals by:

$$\hat{\varepsilon}_t = Y_t - G(\boldsymbol{X}_t, \hat{\boldsymbol{\beta}}_n, \boldsymbol{Y}_t^L, \hat{\boldsymbol{\phi}}_n) - \frac{1}{n-p} \sum_{j=p+1}^{n} (Y_j - G(\boldsymbol{X}_j, \hat{\boldsymbol{\beta}}_n, \boldsymbol{Y}_j^L, \hat{\boldsymbol{\phi}}_n)), \quad t = p+1, p+2, \ldots, n. \tag{5.1}$$

iii) Under H_0, Y_t can be represented as:

$$V_t^0 = G(\boldsymbol{X}_t, \tilde{\boldsymbol{\beta}}_0, \boldsymbol{V}_t^{L0}, \boldsymbol{\phi}) + \varepsilon_t^0, \quad t = p+1, p+2, \ldots, n,$$

where

$$\tilde{\boldsymbol{\beta}}_0 = (\boldsymbol{\theta}_0, \boldsymbol{\nu}); \qquad \boldsymbol{V}_t^{L0} = (V_{t-1}^0, V_{t-2}^0, \ldots, V_{t-p}^0)$$

and $\{\varepsilon_t^0\}$ are unknown error terms under H_0.

iv) Under H_0^*, V_t^0 can be represented as:

$$V_t^{0*} = G(\boldsymbol{X}_t, \hat{\tilde{\boldsymbol{\beta}}}_{0n}, \boldsymbol{V}_t^{L0*}, \hat{\boldsymbol{\phi}}_n) + \varepsilon_t^*, \quad t = p+1, p+2, \ldots, n,$$

where $\hat{\tilde{\boldsymbol{\beta}}}_{0n} = (\boldsymbol{\theta}_0, \hat{\boldsymbol{\nu}}_n)$ and $\{\varepsilon_t^*\}$ are i.i.d from F_n, the e.d.f. of $\{\hat{\varepsilon}_t\}$.

Remarks:

(a) There are various ways of obtaining the initial values $\{V_t^{0*},\ t = 1, 2, \ldots, p\}$. The simplest and most commonly used method is to set $\{V_t^{0*}, t = 1, 2, \ldots, p\}$ equal to the original values $\{Y_t,\ t = 1, 2, \ldots, p\}$, i.e., $V_t^{0*} = Y_t,\ t = 1, 2, \ldots, p$.

(b) Another method, which is probably more suitable for hypothesis testing as it allows one to "mimic" H_0 more accurately, is to use the so called "burn-in" period (see, e.g., Davison and Hinkley, 1997). This procedure entails generating $n + k + 1$ bootstrap residuals to obtain $\varepsilon^*_{-k}, \varepsilon^*_{-k+1}, \ldots, \varepsilon^*_0, \ldots, \varepsilon^*_n$ (the number k is usually chosen to be a fairly large number). Set

$$\begin{aligned}
V_{-k}^{0*} &= \varepsilon^*_{-k} \\
V_{-k+1}^{0*} &= G(\mathbf{X}_t, \hat{\hat{\boldsymbol{\beta}}}_{0n}, V_{-k}^{0*}, \hat{\boldsymbol{\phi}}_n) + \varepsilon^*_{-k+1} \\
&\vdots \\
V_0^{0*} &= G(\mathbf{X}_t, \hat{\hat{\boldsymbol{\beta}}}_{0n}, V_{-1}^{0*}, \hat{\boldsymbol{\phi}}_n) + \varepsilon^*_0 \\
&\vdots \\
V_n^{0*} &= G(\mathbf{X}_t, \hat{\hat{\boldsymbol{\beta}}}_{0n}, V_{n-1}^{0*}, \hat{\boldsymbol{\phi}}_n) + \varepsilon^*_n.
\end{aligned}$$

In the end, the terms $V_{-k}^{0*}, \ldots, V_0^{0*}$ are discarded.

v) Define $C_n(\alpha; \boldsymbol{Y}_n, \boldsymbol{\chi}_n)$ by

$$P^*_{H_0^*}(T_n(\boldsymbol{V}_n^{0*}, \boldsymbol{\chi}_n) \geq C_n(\alpha; \boldsymbol{Y}_n, \boldsymbol{\chi}_n)) \cong \alpha,$$

where

$$\boldsymbol{V}_n^{0*} = (V_1^{0*}, V_2^{0*}, \ldots, V_n^{0*})^T.$$

Next, an algorithm which can be used to estimate the power of this test at the specific alternative $\boldsymbol{r}'\boldsymbol{\theta} = \boldsymbol{r}'\boldsymbol{\theta}_A$ will be provided:

a) Obtain $C_n(\alpha; \boldsymbol{Y}_n, \boldsymbol{\chi}_n)$ by following the residual-based bootstrap algorithm as discussed above.

b) Under H_A, Y_t can be represented as:

$$V_t^A = G(\boldsymbol{X}_t, \tilde{\boldsymbol{\beta}}_A, \boldsymbol{V}_t^{LA}, \boldsymbol{\phi}) + \varepsilon_t^A, \quad t = p+1, p+2, \ldots, n,$$

where

$$\tilde{\boldsymbol{\beta}}_A = (\boldsymbol{\theta}_A, \boldsymbol{\nu}); \qquad \boldsymbol{V}_t^{LA} = (V_{t-1}^A, V_{t-2}^A, \ldots, V_{t-p}^A)$$

and $\{\varepsilon_t^A\}$ are unknown error terms under H_A.

c) Under H_A^*, V_t^A can be represented as:

$$V_t^{A*} = G(\boldsymbol{X}_t, \hat{\hat{\boldsymbol{\beta}}}_{An}, \boldsymbol{V}_t^{LA*}, \hat{\boldsymbol{\phi}}_n) + \varepsilon_t^*, \quad t = p+1, p+2, \ldots, n,$$

where $\hat{\hat{\boldsymbol{\beta}}}_{An} = (\boldsymbol{\theta}_A, \hat{\boldsymbol{\nu}}_n)$ and $\{\varepsilon_t^*\}$ are i.i.d from F_n, the e.d.f. of $\{\hat{\varepsilon}_t\}$ (obtained in (5.1)).

d) The estimated power of the test, at the specific alternative, is then given by:

$$P_{boot}^A = P_{H_A^*}^*(T_n(\boldsymbol{V}_n^{A*}, \boldsymbol{\chi}_n) \geq C_n(\alpha; \boldsymbol{Y}_n, \boldsymbol{\chi}_n)),$$

where

$$\boldsymbol{V}_n^{A*} = (V_1^{A*}, V_2^{A*}, \ldots, V_n^{A*})^T.$$

5.2.2 General formulation (Model II)

Consider data $\big((\mathbf{X}_1, Y_1), (\mathbf{X}_2, Y_2), \ldots, (\mathbf{X}_n, Y_n)\big)$. Interest centers around testing the parameters from the model:

$$Y_t = H(\mathbf{X}_t, \boldsymbol{\lambda}) + u_t, \quad t = 1, 2, \ldots, n,$$

where

- $\mathbf{Y}_n = (Y_1, Y_2, \ldots, Y_n)^T$ are the response variables;
- $\mathbf{X}_t = (X_{t,1}, X_{t,2}, \ldots, X_{t,k+m})$ are regressors (which can be fixed or random);
- $\boldsymbol{\lambda} = (\boldsymbol{\delta}, \boldsymbol{\psi}) \in \Upsilon = \Delta \times \Psi \subset \mathbb{R}^k \times \mathbb{R}^m$ is a set of parameters associated with $\mathbf{X}_t$;
- H is some known function; and
- $\{u_t\}$ is a stationary process given by:

  $$u_t = J(\mathbf{u}_t^L, \boldsymbol{\phi}) + \varepsilon_t, \quad t = p+1, p+2, \ldots, n,$$

  where

  * $\mathbf{u}_t^L = (u_{t-1}, u_{t-2}, \ldots, u_{t-p})$;

* $\boldsymbol{\phi} = (\boldsymbol{\rho}, \boldsymbol{\gamma}) \in \Omega = P \times \Gamma \subset \mathbb{R}^q \times \mathbb{R}^s$ is a set of parameters, $q + s = p$;

* J is a known function; and

* $\{\varepsilon_t\}$ are i.i.d. random errors from an unknown distribution F, such that $E(\varepsilon_t) = 0$ and $E(\varepsilon_t^2) = \sigma_\varepsilon^2$.

Consider the following two scenarios:

(1) Let $\boldsymbol{\rho}$ be the parameter of interest and suppose we wish to test the hypothesis:

$$H_0 : \mathbf{d}'\boldsymbol{\rho} = \mathbf{d}'\boldsymbol{\rho}_0 \qquad \text{vs.} \qquad H_A : \mathbf{d}'\boldsymbol{\rho} > \mathbf{d}'\boldsymbol{\rho}_0,$$

where $\mathbf{d}$ is a known q-dimensional column vector and $\boldsymbol{\delta}, \boldsymbol{\psi}$ and $\boldsymbol{\gamma}$ are a set of k, m and s nuisance parameters, respectively.

(2) Let $\boldsymbol{\delta}$ be the parameter of interest, and suppose we wish to test the hypothesis:

$$H_0 : \mathbf{r}'\boldsymbol{\delta} = \mathbf{r}'\boldsymbol{\delta}_0 \qquad \text{vs.} \qquad H_A : \mathbf{r}'\boldsymbol{\delta} > \mathbf{r}'\boldsymbol{\delta}_0,$$

where $\mathbf{r}$ is a known k-dimensional column vector and $\boldsymbol{\psi}, \boldsymbol{\rho}$ and $\boldsymbol{\gamma}$ are a set of m, q and s nuisance parameters, respectively.

We will present a general formulation only for scenario (1), since the formulation for scenario (2) is analogous.

Suppose the hypothesis in (1) is rejected if and only if $T_n(\mathbf{Y}_n, \boldsymbol{\chi}_n) \geq C_n(\alpha)$, where

$$P_{H_0}(T_n(\mathbf{Y}_n, \boldsymbol{\chi}_n) \geq C_n(\alpha)) \cong \alpha \quad \text{and} \quad \boldsymbol{\chi}_n = (\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n).$$

The critical value $C_n(\alpha)$ is unknown, since F is unknown. However, a bootstrap estimator $C_n(\alpha; \mathbf{Y}_n, \boldsymbol{\chi}_n)$ can be derived for $C_n(\alpha)$.

Bootstrap algorithm

The following residual-based bootstrap algorithm can be used to find the bootstrap estimator $C_n(\alpha; \mathbf{Y}_n, \boldsymbol{\chi}_n)$:

i) Let $\hat{\boldsymbol{\lambda}}_n = (\hat{\boldsymbol{\delta}}_n, \hat{\boldsymbol{\psi}}_n) \in \Upsilon$ be estimators for $\boldsymbol{\lambda}$ in the regression equation:

$$Y_t = H(\mathbf{X}_t, \boldsymbol{\lambda}) + u_t, \quad t = 1, 2, \ldots, n.$$

ii) Define residuals for this model by:

$$\hat{u}_t = Y_t - H(\mathbf{X}_t, \hat{\boldsymbol{\lambda}}_n), \quad t = 1, 2, \ldots, n.$$

iii) Estimate $\boldsymbol{\phi}$ by $\hat{\boldsymbol{\phi}}_n = (\hat{\boldsymbol{\rho}}_n, \hat{\gamma}_n) \in \Omega$ in the regression equation:

$$\hat{u}_t = J(\hat{\mathbf{u}}_t^L, \boldsymbol{\phi}) + \varepsilon_t,$$

where

$$\hat{\mathbf{u}}_t^L = (\hat{u}_{t-1}, \hat{u}_{t-2}, \ldots, \hat{u}_{t-p}), \quad t = p+1, p+2, \ldots, n.$$

iv) Define centered residuals by:

$$\hat{\varepsilon}_t = \hat{u}_t - J(\hat{\mathbf{u}}_t^L, \hat{\boldsymbol{\phi}}_n) - \frac{1}{n-p} \sum_{j=p+1}^{n} (\hat{u}_j - J(\hat{\mathbf{u}}_j^L, \hat{\boldsymbol{\phi}}_n)), \quad t = p+1, p+2, \ldots, n. \tag{5.2}$$

v) Under H_0, Y_t and $\hat{u}_t$ can be represented as:

$$\begin{aligned} V_t^0 &= H(\mathbf{X}_t, \boldsymbol{\lambda}) + \hat{u}_t^0, \\ \hat{u}_t^0 &= J(\hat{\mathbf{u}}_t^{L0}, \tilde{\boldsymbol{\phi}}_0) + \varepsilon_t^0, \quad t = p+1, p+2, \ldots, n, \end{aligned}$$

where $\tilde{\boldsymbol{\phi}}_0 = (\boldsymbol{\rho}_0, \gamma)$; $\hat{\mathbf{u}}_t^{L0} = (\hat{u}_{t-1}^0, \hat{u}_{t-2}^0, \ldots, \hat{u}_{t-p}^0)$ and $\{\varepsilon_t^0\}$ are the unknown error terms under H_0.

vi) Under H_0^*, V_t and $\hat{u}_t^0$ can be represented as:

$$\begin{aligned} V_t^{0*} &= H(\mathbf{X}_t, \hat{\boldsymbol{\lambda}}_n) + \hat{u}_t^{0*}, \\ \hat{u}_t^{0*} &= J(\hat{\mathbf{u}}_t^{L0*}, \hat{\tilde{\boldsymbol{\phi}}}_{0n}) + \varepsilon_t^*, \quad t = p+1, p+2, \ldots, n, \end{aligned}$$

where $\hat{\tilde{\boldsymbol{\phi}}}_{0n} = (\boldsymbol{\rho}_0, \hat{\gamma}_n)$; $\hat{\mathbf{u}}_t^{L0*} = (\hat{u}_{t-1}^{0*}, \hat{u}_{t-2}^{0*}, \ldots, \hat{u}_{t-p}^{0*})$ and $\{\varepsilon_t^*\}$ are i.i.d. from F_n, the e.d.f. of $\{\hat{\varepsilon}_t\}$ (obtained in (5.2)).

vii) Choose $C_n(\alpha; \mathbf{Y}_n, \boldsymbol{\chi}_n)$ such that

$$P_{H_0^*}^*(T_n(\mathbf{V}_n^{0*}, \boldsymbol{\chi}_n) \geq C_n(\alpha; \mathbf{Y}_n, \boldsymbol{\chi}_n)) \cong \alpha,$$

where

$$\mathbf{V}_n^{0*} = (V_1^{0*}, V_2^{0*}, \ldots, V_n^{0*})^T.$$

The following procedure can be used to estimate the power of this test, at the specific alternative $\mathbf{r}'\boldsymbol{\rho} = \mathbf{r}'\boldsymbol{\rho}_A$:

a) Obtain $C_n(\alpha; \mathbf{Y}_n, \boldsymbol{\chi}_n)$ by following the residual-based bootstrap algorithm as discussed above.

b) Under H_A, Y_t and $\hat{u}_t$ can be represented as:

$$\begin{aligned} V_t^A &= H(\mathbf{X}_t, \boldsymbol{\lambda}) + \hat{u}_t^A, \\ \hat{u}_t^A &= J(\hat{\mathbf{u}}_t^{LA}, \tilde{\boldsymbol{\phi}}_A) + \varepsilon_t^A, \quad t = p+1, p+2, \ldots, n, \end{aligned}$$

where $\tilde{\boldsymbol{\phi}}_A = (\boldsymbol{\rho}_A, \boldsymbol{\gamma})$; $\hat{\mathbf{u}}_t^{LA} = (\hat{u}_{t-1}^A, \hat{u}_{t-2}^A, \ldots, \hat{u}_{t-p}^A)$ and $\{\varepsilon_t^A\}$ are the unknown error terms under H_A.

c) Under H_A^*, V_t^A and $\hat{u}_t^A$ can be represented as:

$$\begin{aligned} V_t^{A*} &= H(\mathbf{X}_t, \hat{\boldsymbol{\lambda}}_n) + \hat{u}_t^{A*}, \\ \hat{u}_t^{A*} &= J(\hat{\mathbf{u}}_t^{LA*}, \hat{\tilde{\boldsymbol{\phi}}}_{An}) + \varepsilon_t^*, \quad t = p+1, p+2, \ldots, n, \end{aligned}$$

where $\hat{\tilde{\boldsymbol{\phi}}}_{An} = (\boldsymbol{\rho}_A, \hat{\gamma}_n)$; $\hat{\mathbf{u}}_t^{LA*} = (\hat{u}_{t-1}^{A*}, \hat{u}_{t-2}^{A*}, \ldots, \hat{u}_{t-p}^{A*})$ and $\{\varepsilon_t^*\}$ are i.i.d. from F_n, the e.d.f. of $\{\hat{\varepsilon}_t\}$ (obtained in (5.2)).

d) The estimated power of the test is then given by:

$$P_{boot}^A = P_{H_A^*}^*(T_n(\boldsymbol{V}_n^{A*}, \boldsymbol{\chi}_n) \geq C_n(\alpha; \boldsymbol{Y}_n, \boldsymbol{\chi}_n)),$$

where

$$\boldsymbol{V}_n^{A*} = (V_1^{A*}, V_2^{A*}, \ldots, V_n^{A*})^T.$$

Models with heteroscedasticity

In equations (5.1) and (5.2) we obtained the residuals under the assumption that $E(\varepsilon_t^2) = \sigma_\varepsilon^2$ for all t. If, however, we work with models that are heteroscedastic, i.e., $E(\varepsilon_t^2) = \sigma_t^2, \ t = 1, 2, \ldots, n$, bootstrap samples must be generated in such a way that the relationship between the variance of each error term and the corresponding regressors is retained. One possible way to do this is to use the *wild bootstrap* proposed by Wu (1986), Liu (1988) and further developed by Mammen (1993). The wild bootstrap is implemented by first obtaining the residuals in (5.1) or (5.2), without centering them. The bootstrap residuals $\{\varepsilon_t^*, t = 1, 2, \ldots, n\}$ are then obtained from a distribution with probability mass in two points. That is

$$\varepsilon_t^* = \begin{cases} a, & \text{with probability } p \\ b, & \text{with probability } 1-p, \end{cases}$$

so that $E_*(\varepsilon_t^*) = 0$, $E_*(\varepsilon_t^*)^2 = \hat{\varepsilon}_t^2$ and $E_*(\varepsilon_t^*)^3 = \hat{\varepsilon}_t^3$. One possible choice for a, b and p suggested by Mammen (1993) is $a = \hat{\varepsilon}_t(1-\sqrt{5})/2, \ b = \hat{\varepsilon}_t(1+\sqrt{5})/2$ and $p = (\sqrt{5}+1)/(2\sqrt{5})$.

Another possible choice for the two-point distribution is the Rademacher distribution. For this distribution $a = \hat{\varepsilon}_t, b = -\hat{\varepsilon}_t$ and $p = \frac{1}{2}$.

Remark: The reason why it is not necessary to center the residuals $\{\hat{\varepsilon}_t, t = 1, 2, \ldots, n\}$ is because the two-point distribution already has mean 0 (in the bootstrap world).

After obtaining the bootstrap residuals $\{\varepsilon_t^*, t = 1, 2, \ldots, n\}$ using the wild bootstrap, one can proceed with the residual-based bootstrap algorithm as previously described.

In the next section we discuss three models (the location-scale model, the multiple linear regression model and the simple linear regression model with correlated errors), where residuals are resampled in order to perform the bootstrap hypothesis tests.

5.2.3 Resampling residuals applied to common testing scenarios

(a) Location-scale model

Let $\mathbf{X}_n = (X_1, X_2, \ldots, X_n)$ denote a random sample from an unknown univariate distribution F with mean μ and variance σ^2.

Consider the model:

$$X_t = \mu + \sigma\varepsilon_t, \quad t = 1, 2, \ldots, n, \tag{5.3}$$

where $\{\varepsilon_t\}$ are i.i.d. random errors with zero mean and variance 1.

Suppose we wish to test the hypothesis:

$$H_0 : \mu = \mu_0 \qquad \text{vs.} \qquad H_A : \mu > \mu_0,$$

where, once again, H_0 is rejected if and only if

$$\frac{\sqrt{n}(\overline{X}_n - \mu_0)}{S_n(\mathbf{X}_n)} \geq C_n(\alpha).$$

We can estimate $C_n(\alpha)$ by the bootstrap estimator $C_n(\alpha; \mathbf{X}_n)$, by following the residual-based bootstrap algorithm discussed in the previous section (Model I):

i) Fit the model and obtain the residuals:

$$\hat{\varepsilon}_t = \frac{X_t - \overline{X}_n}{S_n(\mathbf{X}_n)}. \tag{5.4}$$

ii) Under H_0, X_t can be represented as:

$$V_t^0 = \mu_0 + \sigma\varepsilon_t^0.$$

iii) Under H_0^*, V_t^0 can be represented as:

$$V_t^{0*} = \mu_0 + S_n(\mathbf{X}_n)\varepsilon_t^*, \tag{5.5}$$

where $\{\varepsilon_t^*\}$ are i.i.d. from the e.d.f. of $\{\hat{\varepsilon}_t\}$ (obtained in (5.4)).

iv) Define $C_n(\alpha; \mathbf{X}_n)$ by

$$P_{H_0^*}^* \left(\frac{\sqrt{n}(\overline{V}_n^{0*} - \mu_0)}{S_n(\mathbf{V}_n^{0*})} \geq C_n(\alpha; \mathbf{X}_n) \right) \cong \alpha. \tag{5.6}$$

It is an elementary exercise to show that (5.6) is equivalent to

$$P_{H_0^*}^* \left(\frac{\sqrt{n}(\overline{X}_n^* - \overline{X}_n)}{S_n(\mathbf{X}_n^*)} \geq C_n(\alpha; \mathbf{X}_n) \right) \cong \alpha, \tag{5.7}$$

which is indeed identical to the expression given in (3.6).

If, however, we obtain our residuals in (5.4) by enforcing the null hypothesis, and we do not center the residuals, i.e.,

$$\hat{\varepsilon}_t = \frac{X_t - \mu_0}{S_n(\mathbf{X}_n)},$$

it is a simple matter to verify that (5.7) becomes

$$P_{H_0^*}^* \left(\frac{\sqrt{n}(\overline{X}_n^* - \mu_0)}{S_n(\mathbf{X}_n^*)} \geq C_n(\alpha; \mathbf{X}_n) \right) \cong \alpha. \tag{5.8}$$

It was shown in Chapter 4 that using a bootstrap test based on the critical value defined in (5.8), leads to a test that performs extremely poorly.

For the bootstrap estimate of power for the test

$$H_0 : \mu = \mu_0 \qquad \text{vs.} \qquad H_A : \mu > \mu_0,$$

the reader is referred back to Section 3.2.2 (a).

Next, consider testing

$$H_0 : \sigma^2 = \sigma_0^2 \qquad \text{vs.} \qquad H_A : \sigma^2 > \sigma_0^2,$$

where H_0 is rejected if and only if

$$\frac{nS_n^2(\mathbf{X}_n)}{\sigma_0^2} \geq C_n(\alpha).$$

Following the residual-based bootstrap algorithm, we can estimate $C_n(\alpha)$ by the bootstrap estimator $C_n(\alpha; \mathbf{X}_n)$ as follows:

i) Fit the model and obtain the residuals:

$$\hat{\varepsilon}_t = \frac{X_t - \overline{X}_n}{S_n(\mathbf{X}_n)}. \tag{5.9}$$

ii) Under H_0, X_t can be represented as:

$$V_t^0 = \mu + \sigma_0 \varepsilon_t^0.$$

iii) Under H_0^*, V_t^0 can be represented as:

$$V_t^{0*} = \overline{X}_n + \sigma_0 \varepsilon_t^*, \tag{5.10}$$

where $\{\varepsilon_t^*\}$ are i.i.d. from the e.d.f. of $\{\hat{\varepsilon}_t\}$ (obtained in (5.9)).

iv) Define $C_n(\alpha; \mathbf{X}_n)$ by

$$P^*_{H_0^*}\left(\frac{nS_n^2(\mathbf{V}_n^{0*})}{\sigma_0^2} \geq C_n(\alpha; \mathbf{X}_n)\right) \cong \alpha. \tag{5.11}$$

It is easy to see that (5.11) is the same as

$$P^*_{H_0^*}\left(\frac{nS_n^2(\mathbf{X}_n^*)}{S_n^2(\mathbf{X}_n)} \geq C_n(\alpha; \mathbf{X}_n)\right) \cong \alpha, \tag{5.12}$$

which is indeed identical to the expression given in (3.10).

If, however, we obtain our residuals in (5.9) by enforcing the null hypothesis, i.e.,

$$\hat{\varepsilon}_t = \frac{X_t - \overline{X}_n}{\sigma_0},$$

it is a simple matter to verify that (5.11) becomes:

$$P^*_{H_0^*}\left(\frac{nS_n^2(\mathbf{X}_n^*)}{\sigma_0^2} \geq C_n(\alpha; \mathbf{X}_n)\right) \cong \alpha, \tag{5.13}$$

which is identical to the expression given in (3.11).

It was shown in Theorem 3.2.1 that using the critical value defined in (5.13) leads to a bootstrap test that performs very poorly. Further evidence of this poor performance will be illustrated in Chapter 8.

The reader is again referred back to Section 3.2.2 (e) for the bootstrap estimate of power for the test

$$H_0 : \sigma^2 = \sigma_0^2 \qquad \text{vs.} \qquad H_A : \sigma^2 > \sigma_0^2.$$

Remarks:

(a) It is clear from the location-scale model, that if we follow the residual-based bootstrap algorithm, we end up with the same bootstrap test as when we transform the data in order to "mimic" H_0 (see Section 3.2.2 (a) and 3.2.2 (e)).

(b) Imposing the null hypothesis when obtaining the residuals can lead to serious problems. For testing the mean, this problem can be overcome by centering the residuals, but centering the residuals will not help to solve this problem when testing for the variance.

(c) We therefore advocate that *unrestricted residuals* (i.e., where we do not impose the null hypothesis) must be used when the residuals are resampled in order to apply the bootstrap to hypothesis testing.

(b) Multiple linear regression

Consider the model:

$$Y_t = \beta_0 + \sum_{i=1}^{p} \beta_i X_{t,i} + \varepsilon_t, \quad t = 1, 2, \dots, n, \tag{5.14}$$

where there are n observations $(\mathbf{X}_1, Y_1), (\mathbf{X}_2, Y_2), \dots, (\mathbf{X}_n, Y_n)$ and $\{\beta_i, i = 0, 1, 2, \dots, p\}$ are parameters; $\mathbf{X}_t = (X_{t,1}, \dots, X_{t,p})^T$ is a vector of covariates; $\mathbf{Y}_n = (Y_1, Y_2, \dots, Y_n)^T$ are the response variables and $\{\varepsilon_t\}$ are the random errors with some unknown d.f. F.

Suppose we are interested in testing the following hypothesis:

$$H_0 : \beta_1 = b_1 \quad \text{vs.} \quad H_A : \beta_1 > b_1.$$

The test rejects H_0 if and only if

$$T_n(\mathbf{Y}_n, \boldsymbol{\chi}_n) = \frac{\hat{\beta}_1 - b_1}{S_n(\hat{\beta}_1)} \geq C_n(\alpha),$$

where $\hat{\beta}_1 = \hat{\beta}_1(\mathbf{Y}_n, \boldsymbol{\chi}_n)$; $\boldsymbol{\chi}_n = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n)$ and $S_n(\hat{\beta}_1)$ denotes the estimated standard error of $\hat{\beta}_1$.

Bootstrap-based testing in the multiple linear regression framework was discussed by, among others, Davison and Hinkley (1997); Davidson and MacKinnon (2004) and Paparoditis and Politis (2005).

In order to obtain the bootstrap critical value, $C_n(\alpha; \mathbf{Y}_n, \boldsymbol{\chi}_n)$, we again follow the residual-based bootstrap algorithm:

i) Fit the model in (5.14) to obtain estimates $\{\hat{\beta}_i, i = 0, 1, \ldots, p\}$.

ii) Obtain the centered residuals:

$$\hat{\varepsilon}_t = Y_t - \hat{\beta}_0 - \sum_{i=1}^{n} \hat{\beta}_i X_{t,i} - \frac{1}{n}\sum_{j=1}^{n}(Y_j - \hat{\beta}_0 - \sum_{i=1}^{n} \hat{\beta}_i X_{j,i}). \tag{5.15}$$

iii) Under H_0, Y_t can be represented as:

$$V_t^0 = \beta_0 + b_1 X_{t,1} + \sum_{i=2}^{n} \beta_i X_{t,i} + \varepsilon_t^0. \tag{5.16}$$

iv) Under H_0^*, V_t^0 can be represented as:

$$V_t^{0*} = \hat{\beta}_0 + b_1 X_{t,1} + \sum_{i=2}^{n} \hat{\beta}_i X_{t,i} + \varepsilon_t^*, \tag{5.17}$$

where $\{\varepsilon_t^*\}$ are i.i.d. from the e.d.f. of $\{\hat{\varepsilon}_t\}$ (obtained in (5.15)).

v) $C_n(\alpha; \mathbf{Y}_n, \chi_n)$ is defined by

$$P^*_{H_0^*}\left(\frac{\hat{\beta}_1^* - b_1}{S_n(\hat{\beta}_1^*)} \geq C_n(\alpha; \mathbf{Y}_n, \chi_n)\right) \cong \alpha,$$

where $\hat{\beta}_1^* = \hat{\beta}_1(\mathbf{V}_n^{0*}, \chi_n)$.

In order to estimate the power of this test, at the specific alternative $\beta_1 = b_A$, we replace b_1 in (5.16) and (5.17) by b_A to obtain

$$V_t^{A*} = \hat{\beta}_0 + b_A X_{t,1} + \sum_{i=2}^{n} \hat{\beta}_i X_{t,i} + \varepsilon_t^*,$$

where $\{\varepsilon_t^*\}$ are i.i.d. from the e.d.f. of $\{\varepsilon_t\}$ (also obtained in (5.15)).
The estimated power of the test, at the specific alternative $\beta_1 = b_A$, is then:

$$P^A_{boot} = P^*_{H_A^*}\left(\frac{\hat{\beta}_1^* - b_1}{S_n(\hat{\beta}_1^*)} \geq C_n(\alpha; \mathbf{Y}_n, \chi_n)\right),$$

where $\hat{\beta}_1^* = \hat{\beta}_1(\mathbf{V}_n^{A*}, \chi_n)$.

(c) Simple linear regression model with errors following an AR(1) process

Consider the simple linear regression model:

$$Y_t = \beta_0 + \beta_1 X_t + u_t, \tag{5.18}$$

with

$$u_t = \rho u_{t-1} + \epsilon_t, \quad t = 2, 3, \ldots, n \tag{5.19}$$

and $|\rho| < 1$.

Here, β_0 and β_1 are parameters; Y_t is the value of the response variable in the t^{th} trial; X_t is the value of the covariate in the t^{th} trial and $\{\epsilon_t\}$ are the random errors with some unknown d.f. F, $t = 1, 2, \ldots, n$. Let $\mathbf{X}_n = (X_1, X_2, \ldots, X_n)^T$ and $\mathbf{Y}_n = (Y_1, Y_2, \ldots, Y_n)^T$.

Consider testing the hypothesis:

$$H_0 : \rho = \rho_0 \qquad \text{vs.} \quad H_A : \rho > \rho_0.$$

The test rejects H_0 if and only if

$$T_n(\mathbf{Y}_n, \mathbf{X}_n) \geq C_n(\alpha).$$

We can obtain $C_n(\alpha; \mathbf{X}_n, \mathbf{Y}_n)$, the bootstrap estimator of $C_n(\alpha)$, by following the residual-based bootstrap algorithm for Model II:

i) Estimate β_0 and β_1 by $\hat{\beta}_0$ and $\hat{\beta}_1$ in the regression equation:

$$Y_t = \beta_0 + \beta_1 X_t + u_t, \quad t = 1, 2, \ldots, n.$$

ii) Obtain the residuals:

$$\hat{u}_t = Y_t - \hat{\beta}_0 - \hat{\beta}_1 X_t.$$

iii) Estimate ρ by $\hat{\rho}$ in the regression equation:

$$\hat{u}_t = \rho \hat{u}_{t-1} + \varepsilon_t, \quad t = 2, 3, \ldots, n.$$

iv) Obtain the centered residuals as:

$$\hat{\varepsilon}_t = \hat{u}_t - \hat{\rho}\hat{u}_{t-1} - \frac{1}{n-1}\sum_{j=2}^{n}(\hat{u}_j - \hat{\rho}\hat{u}_{j-1}), \quad t = 2, 3, \ldots, n. \tag{5.20}$$

v) Under H_0, Y_t and $\hat{u}_t$ can be represented as:

$$\begin{aligned} V_t^0 &= \beta_0 + \beta_1 X_t + \hat{u}_t^0, \\ \hat{u}_t^0 &= \rho_0 \hat{u}_{t-1}^0 + \varepsilon_t^0, \quad t = 2, 3, \ldots, n. \end{aligned}$$

vi) Under H_0^*, V_t and $\hat{u}_t^0$ can be represented as:

$$\begin{aligned} V_t^{0*} &= \hat{\beta}_0 + \hat{\beta}_1 X_t + \hat{u}_t^{0*}, \\ \hat{u}_t^{0*} &= \rho_0 \hat{u}_{t-1}^{0*} + \varepsilon_t^*, \quad t = 2, 3 \ldots, n, \end{aligned}$$

where $\{\varepsilon_t^*\}$ are i.i.d. from the e.d.f. of $\{\hat{\varepsilon}_t\}$ (obtained in (5.20)).

In order to obtain the bootstrap residual value at $t = 1$ we can choose $\hat{u}_1^{0*} = \hat{u}_1$ or alternatively we can again make use of a "burn-in" period.

vii) Define $C_n(\alpha; \mathbf{X}_n, \mathbf{Y}_n)$ by

$$P_{H_0^*}^* \left(T_n(\mathbf{V}_n^{0*}, \mathbf{X}_n) \geq C_n(\alpha; \mathbf{X}_n, \mathbf{Y}_n) \right) \cong \alpha.$$

Remarks:

(a) If we combine (5.19) with (5.18) by substituting $\rho u_{t-1} + \varepsilon_t$ for u_t and then replacing u_{t-1} by $Y_{t-1} - \beta_0 - \beta_1 X_{t-1}$, we obtain the nonlinear regression model

$$Y_t = \beta_0(1-\rho) + \beta_1 X_t - \rho\beta_1 X_{t-1} + \rho Y_{t-1} + \varepsilon_t, \quad t = 2, 3, \ldots, n. \tag{5.21}$$

In order to test the hypothesis

$$H_0 : \rho = \rho_0 \quad \text{vs.} \qquad H_A : \rho > \rho_0,$$

we can now obtain the bootstrap critical value by following the residual-based bootstrap algorithm for Model I. It is just important to note that (5.21) is linear in the regressors, but nonlinear in the parameters ρ and β_1, and it therefore needs to be estimated by nonlinear least squares or some other nonlinear estimation method.

(b) Davidson and Mackinnon (1999, 2004) used a variant of Newton's Method, namely Gauss-Newton regression (GNR) to test the specific hypothesis:

$$H_0 : \rho = 0 \quad \text{vs.} \qquad H_A : \rho > 0$$

in model (5.21).

The power of the test, at the specific alternative $\rho = \rho_A$, can be estimated as follows:

a) Obtain $C_n(\alpha; \mathbf{X}_n, \mathbf{Y}_n)$.

b) Under H_A, Y_t and $\hat{u}_t$ can be represented as:

$$\begin{aligned} V_t^A &= \beta_0 + \beta_1 X_t + \hat{u}_t^A, \\ \hat{u}_t^A &= \rho_A \hat{u}_{t-1}^A + \varepsilon_t^A, \quad t = 2, 3, \ldots, n. \end{aligned}$$

c) Under H_A^*, V_t^A and $\hat{u}_t^A$ can be represented as:

$$\begin{aligned} V_t^{A*} &= \hat{\beta}_0 + \hat{\beta}_1 X_t + \hat{u}_t^{A*}, \\ \hat{u}_t^{A*} &= \rho_A \hat{u}_{t-1}^{A*} + \varepsilon_t^*, \quad t = 2, 3, \ldots, n, \end{aligned}$$

where $\{\varepsilon_t^*\}$ are i.i.d. from the e.d.f. of $\{\hat{\varepsilon}_t\}$ (obtained in 5.20).

We can again choose $\hat{u}_1^{A*} = \hat{u}_1$ or make use of a "burn-in" period.

d) The estimated power of the test, at the specific alternative $H_A : \rho = \rho_A$, is then given by:

$$P_{boot}^A = P_{H_A^*}^* \left(T_n(\mathbf{V}_n^{A*}, \mathbf{X}_n) \geq C_n(\alpha; \mathbf{X}_n, \mathbf{Y}_n) \right).$$

5.2.4 Concluding remarks

In a Monte-Carlo study we found, that for the test

$$H_0 : \rho = 0 \quad \text{vs.} \qquad H_A : \rho > 0,$$

the performances of the following three tests were almost identical:

- the test described in Section 5.2.3 (c);
- the test described in Section 5.2.3 (c), with the exception that the null hypothesis was imposed in (5.20);
- the test based on GNR.

Moreover, MacKinnon (2002) also found that for the hypothesis he considered, resampling the residuals from either a restricted (where the null hypothesis is imposed when obtaining the residuals) or unrestricted (where the null hypothesis is not imposed in the residual estimation step) model, made little difference to the performance of their tests. In the literature there are also some examples of authors choosing to resample residuals from a restricted model (see, e.g., Nankervis and Savin, 1996; Park, 2003; Davidson and MacKinnon, 2007).

However, many authors, including Van Giersbergen and Kiviet (2002); Paparoditis and Politis (2005) and Martin (2007) have advocated that residuals must be resampled from an unrestricted model. In this chapter we also gave an example (the location-scale model), where failing to resample from an unrestricted model, leads to a test that performs dismally. We therefore strongly advise that residuals should always be resampled from an unrestricted model (i.e., the null or alternative hypothesis must not be imposed when obtaining the residuals).

Chapter 6

A brief survey of copulas and two measures of association

In this chapter a very brief overview of copulas is provided, focusing specifically on how to simulate from different families of copulas. The chapter concludes with a discussion on two measures of association, namely Spearman's rho and Kendall's tau and how these measures can be expressed in terms of copulas.

6.1 Introduction

Nelsen (2006) informally defines a copula as *"a function which joins or 'couples' a multivariate distribution function to its one-dimensional marginal distribution functions."* Fisher (1997) notes in the Encyclopedia of Statistical Sciences that *"Copulas [are] of interest to statisticians for two main reasons: First, as a way of studying scale-free measures of dependence; and secondly as a starting point for constructing families of bivariate distributions ..."*

In this dissertation we make use of copulas for precisely these two reasons: a way to simulate data from different bivariate distributions and secondly, because of the fact that Spearman's rho and Kendall's tau (and other measures of association) can be expressed in terms of copulas.

We do not attempt to give a comprehensive overview of copulas in this chapter, partly because copulas are not the primary topic of this dissertation, but rather a means to an end (which will become clear in Chapters 7 and 8).

Readers seeking to learn more about copulas (and related topics) are referred to, among others, Hutchinson and Lai (1990); Joe (1997) and the excellent monograph on copulas by Nelsen (2006).

In what follows we shall, for simplicity, restrict our discussion to bivariate observations only. However, most of the definitions and theorems would be just as relevant in higher dimensions.

6.2 Copulas: some basic concepts

Let X and Y denote two continuous random variables with distribution functions $F(x) = P(X \leq x)$ and $G(y) = P(Y \leq y)$, $x, y \in \mathbb{R}$ and joint distribution function $H(x, y) = P(X \leq x, Y \leq y)$ for $(x, y) \in \mathbb{R}^2$.

Definition 6.2.1 *A copula is a function $C : [0,1]^2 \to [0,1]$ which satisfies:*

(a) for every u, v in $[0,1]$, $C(u,0) = 0 = C(0,v)$ and $C(u,1) = u$ and $C(1,v) = v$; and

(b) for every u_1, u_2, v_1, v_2 in $[0,1]$, such that $u_1 \leq u_2$ and $v_1 \leq v_2$,

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0.$$

The importance of copulas in statistics is described in Sklar's Theorem (see Sklar, 1959):

Theorem 6.2.1 *Let X and Y be random variables with joint distribution function H and marginal distribution functions F and G respectively. Then there exists a copula C such that*

$$H(x, y) = C(F(x), G(y)) \tag{6.1}$$

for all x, y in $\mathbb{R}$. If F and G are continuous, then C is unique. Otherwise, the copula C is uniquely determined on Range$(F) \times$ Range(G). Conversely, if C is a copula and F and G are distribution functions, then the function H defined by (6.1) is a joint distribution function with margins F and G.

A proof of this theorem can be found in Schweizer and Sklar (1983) and in Nelsen (2006).
Thus, given a joint distribution function H with continuous marginals F and G as in Theorem 6.2.1, it is easy to construct the corresponding copula:

$$C(u, v) = H(F^{-1}(u), G^{-1}(v)).$$

6.3 Conditional distribution and random variate generation

Given a copula $C(u, v)$, how does one simulate random variables from this given copula? This is a very important question as Nelsen (2006) noted: *"one of the primary applications of copulas is in simulations and Monte-Carlo studies"*.

One possible way would be to use the so-called conditional distribution method. For this method, one needs the conditional distribution function for V given $U = u$. Denote this conditional distribution by $C_u(v)$, then

$$\begin{aligned} C_u(v) &= P(V \leq v | U = u) \\ &= \lim_{\delta \to 0} \frac{P(u \leq U \leq u+\delta, V \leq v)}{P(u \leq U \leq u+\delta)} \\ &= \lim_{\delta \to 0} \frac{C(u+\delta, v) - C(u,v)}{\delta} \\ &= \frac{\partial}{\partial u} C(u,v). \end{aligned}$$

The following algorithm, based on the conditional distribution function (Nelsen, 2006), can be used to simulate pairs (U, V) from a given copula $C(u, v)$:

1) Generate a continuous uniform variable U from $(0, 1)$.

2) Simulate V from $C_u(v)$ as follows:

   - generate a second continuous uniform variate U_1 from $(0, 1)$, independent of U;
   - set $V = C_U^{-1}(U_1)$.

3) The pair of variables is then (U, V).

Remark: There exist other methods to simulate data from a given copula (see, e.g., Devroye, 1986).

In the next section we give examples of some families of copulas as well as algorithms which can be used to simulate observations from these copulas.

6.4 Examples of copulas

Nelsen (2006) noted that *"if we have a collection of copulas, then as a consequence of Sklar's Theorem, we automatically have a collection of bivariate or multivariate distributions with whatever marginal distributions we desire."*

It is thus desirable, for the purpose of statistical modelling, to have a collection of copulas at one's disposal. Numerous examples of copulas can be found in the literature (see, e.g., Hutchinson and Lai, 1990; Joe, 1997; Nelsen, 2006).

In our discussion we will only focus on a few single parameter families of copulas, denoted by $C_\theta(u, v)$. Where possible, an explicit expression for the copula as well as an algorithm for simulating variates from the specific copula will be given.

An algorithm which allows one to simulate a single pair (U, V) which has uniform marginals and the desired copula will be given. By repeating the algorithm n times, one then obtains a sample of n i.i.d. bivariate random variables from the desired copula.

6.4.1 Normal copula

The normal copula is a single parameter family of copulas and is given by:

$$C_\theta(u, v) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{e^{\dfrac{-(s^2 - 2\theta st + t^2)}{2(1-\theta^2)}}}{2\pi\sqrt{1-\theta^2}} dsdt,$$

where $-1 < \theta < 1$.

The following algorithm can be used to simulate random variates (U, V) from the bivariate normal copula (see Fishman, 1996):

1) Let $\mathbf{R} = \begin{bmatrix} 1 & \rho_{12} \\ \rho_{21} & 1 \end{bmatrix}$ be an arbitrary correlation matrix, where $\rho_{12} = \rho_{21}$ is the correlation coefficient between two variables X and Y.

2) Obtain a Cholesky-decomposition (see, e.g., Burden and Faires, 1989) of the matrix $\mathbf{R}$, i.e. $\mathbf{R} = BB^T$, where B is a (2×2) lower triangular matrix. The Cholesky-decomposition can be computed by most mathematical and statistical software.

3) Generate two i.i.d. standard normal random variables $\mathbf{Y} = (Y_1, Y_2)^T$.

4) Compute the matrix product $\mathbf{Z} = \mathbf{BY}$. Then $\mathbf{Z} = (Z_1, Z_2)^T$ has a bivariate normal distribution.

5) Obtain $U = \Phi(Z_1)$ and $V = \Phi(Z_2)$.

6.4.2 Farlie-Gumbel-Morgenstern copula (FGM)

The Farlie-Gumbel-Morgenstern copula is given by (Farlie, 1960):

$$C_\theta(u, v) = uv + \theta uv(1-u)(1-v), \quad -1 \leq \theta \leq 1.$$

Johnson (1986) proposed the following algorithm to simulate random variables (U, V) from the FGM copula.

1) Generate two independent continuous uniform variables U and U_1 from $(0,1)$.

2) Set $A = 1 + \theta(1-2U)$; $B = \sqrt{A^2 - 4(A-1)U_1}$.

3) Set $V = \dfrac{2U_1}{A+B}$.

4) The pair of variables is then (U,V).

Remark: This algorithm is based on the conditional distribution method.

6.4.3 Plackett family of copulas

The Plackett family of copulas is given by (Plackett, 1965):

$$C_\theta(u,v) = \frac{[1+(\theta-1)(u+v)] - \sqrt{[1+(\theta-1)(u+v)]^2 - 4uv\theta(\theta-1)}}{2(\theta-1)}, \quad \theta > 0, \theta \neq 1.$$

For $\theta = 1$, $C_\theta(u,v) = uv$, the independence copula.
Johnson (1986) proposed the following algorithm, based on the conditional distribution method, to simulate from a Plackett copula:

1) Generate two independent continuous uniform variables U and U_1 from $(0,1)$.

2) Set $A = U_1(1-U_1)$; $B = \theta + A(\theta-1)^2$; $C = 2A(U\theta^2 + 1 - U) + \theta(1-2A)$ and $D = \sqrt{\theta(\theta + 4AU(1-U)(1-\theta)^2}$.

3) Set $V = [C - (1-2U_1)D]/(2B)$.

4) The pair of variables is then (U,V).

6.4.4 Cuadras-Augé copula

The Cuadras-Augé copula is given by (Cuadras and Augé, 1981):

$$C_\theta(u,v) \quad = \quad [\min(u,v)]^\theta [uv]^{1-\theta}, \quad 0 \leq \theta \leq 1.$$

The following algorithm based on the conditional distribution method can be used to generate random variables (U,V) from the Cuadras-Augé copula.

1) Generate two independent continuous uniform variates U and U_1 from $(0,1)$.

2) Set

$$V = \begin{cases} \dfrac{U_1 U^{\theta}}{1-\theta}, & \text{if} \quad 0 \le U_1 \le (1-\theta)U^{1-\theta} \\ U, & \text{if} \quad (1-\theta)U^{1-\theta} < U_1 \le U^{1-\theta} \\ U_1^{1/(1-\theta)}, & \text{if} \quad U^{1-\theta} < U_1 \le 1. \end{cases}$$

3) The desired pair is then (U, V).

6.4.5 Clayton copula

The Clayton copula is given by (Clayton, 1978):

$$C_{\theta}(u,v) = [\max(u^{-\theta} + v^{-\theta} - 1, 0)]^{-\frac{1}{\theta}}, \quad \theta \ge -1, \theta \ne 0.$$

Two algorithms to simulate variables (U, V) from the Clayton copula are presented.
The first one is proposed by Devroye (1986) and the second one is based on the conditional distribution method.

Algorithm 1

1) Generate two standard exponential variables X and Y and a gamma variable Z (independently), with parameters $\alpha = \theta, \beta = 1$.

2) Set $U = (1 + \frac{X}{Z})^{-\theta}$ and $V = (1 + \frac{Y}{Z})^{-\theta}$.

3) The pair is then (U, V).

Algorithm 2

1) Generate two independent continuous uniform variables U and U_1 from $(0, 1)$.

2) Set $V = [1 + U^{-\theta}(U_1^{\frac{-\theta}{1+\theta}} - 1)]^{-\frac{1}{\theta}}$.

3) The desired pair is then (U, V).

6.4.6 Ali-Mikhail-Haq copula

The Ali-Mikhail-Haq copula is given by (Ali et al., 1978):

$$C_{\theta}(u,v) = \frac{uv}{1 - \theta(1-u)(1-v)}, \quad -1 \le \theta \le 1.$$

In order to simulate random variates from this copula, we can again make use of the conditional distribution method. Johnson (1986) proposed the following algorithm:

1) Generate two independent continuous uniform variables U and U_1 from $(0,1)$.

2) Set $A = 1-U$; $B = -\theta(2AU_1+1)+2\theta^2A^2U_1+1$ and $C = \theta^2(4A^2U_1-4AU_1+1)-\theta(4AU_1-4U_1+2)+1$.

3) Set $V = 2U_1(A\theta-1)^2/(B+\sqrt{C})$.

4) The pair of variables is then (U,V).

Remarks:

(a) The two copulas discussed in Sections 6.4.5 and 6.4.6 are part of a class of copulas known as Archimedean copulas.

(b) For a detailed discussion on Archimedean copulas, the interested reader is referred to Nelsen (2006).

6.4.7 Raftery's bivariate exponential distribution

Although this distribution does not represent a copula, we discuss it in this section due to the fact that this bivariate exponential distribution is going to be used extensively in our Monte-Carlo studies (see Chapter 8).

The Raftery distribution is generated by considering

$$\begin{aligned} X &= (1-\theta)Z_1 + IZ_3 \quad \text{and} \\ Y &= (1-\theta)Z_2 + IZ_3, \end{aligned}$$

where Z_1, Z_2 and Z_3 are i.i.d. exponential random variables with parameter $\lambda > 0$ and I is a Bernoulli random variable with parameter θ, independent of Z_i, $i = 1,2,3$.

This bivariate exponential distribution is due to Raftery (1984).

6.5 Measures of association

In this section we discuss the two most commonly used nonparametric measures of association for two random variables, namely Kendall's tau (τ) and Spearman's rho (ρ_s). The relationship between these two measures is, e.g., discussed in Fredricks and Nelsen (2007).

6.5.1 Kendall's tau

The nonparametric measure of association known as Kendall's tau was first discussed by Kendall (1938). Underpinning the definition of Kendall's tau are the concepts of concordance and discordance.

Definition 6.5.1 *Let (X_i, Y_i) and (X_j, Y_j) denote two observations from a vector (X, Y) of continuous random variables. (X_i, Y_i) and (X_j, Y_j) are concordant if*

$$sgn(X_i - X_j) = sgn(Y_i - Y_j).$$

Correspondingly, a discordant pair is a pair, as defined above, where

$$sgn(X_i - X_j) = -sgn(Y_i - Y_j)$$

and the sign function, sgn, is defined as

$$sgn(x) = \begin{cases} 1 & if \quad x > 0 \\ -1 & if \quad x < 0 \\ 0 & if \quad x = 0. \end{cases}$$

The population version of Kendall's tau for a vector (X, Y) of continuous random variables, with joint distribution function $H(x, y)$, can now be defined as follows (see, e.g., Kruskall, 1958):

Definition 6.5.2 *Let (X_1, Y_1) and (X_2, Y_2) be i.i.d. random vectors, each with joint distribution function $H(x, y)$. The population version of Kendall's tau is then the probability of concordance minus the probability of discordance:*

$$\begin{aligned} \tau &= P(sgn(X_1 - X_2) = sgn(Y_1 - Y_2)) - \\ &\quad P(sgn(X_1 - X_2) = -sgn(Y_1 - Y_2)) \\ &= P((X_1 - X_2)(Y_1 - Y_2) > 0) - P((X_1 - X_2)(Y_1 - Y_2) < 0). \end{aligned}$$

The population version of Kendall's tau can also be expressed in terms of the copula C associated with the vector (X, Y) as follows (Schweizer and Wolff, 1981):

$$\tau = 4 \int_0^1 \int_0^1 C(u, v) dC(u, v) - 1. \tag{6.2}$$

If C is absolutely continuous, then (6.2) becomes

$$\tau = 1 - 4 \int_0^1 \int_0^1 \frac{\partial C}{\partial u}(u, v) \frac{\partial C}{\partial v}(u, v) dudv.$$

Table 6.1 given at the end of this chapter contains expressions for τ (where possible) for the various families of copulas discussed in Section 6.4.

Sample version of Kendall's tau

In order to estimate the parameter τ from a random sample of n observations (X_1, Y_1), (X_2, Y_2), $\ldots$, $(X_n Y_n)$ drawn from a vector (X, Y) of continuous random variables, one must find estimates for the probability of concordance and discordance. This can be done as follows (Kruskal, 1958; Noether, 1967; Kendall, 1970):

For each set of pairs $(X_i, Y_i), (X_j, Y_j)$ of sample observations, define the indicator variables:

$$I_{ij} = sgn(X_i - X_j)sgn(Y_i - Y_j),\ i = 1, 2, \ldots n;\ j = 1, 2, \ldots n.$$

The values assumed by I_{ij} are then:

$$I_{ij} = \begin{cases} 1 & \text{if these pairs are concordant} \\ -1 & \text{if these pairs are discordant} \\ 0 & \text{else.} \end{cases}$$

Now, since $I_{ij} = I_{ji}$ and $I_{ii} = 0$, there are only $\binom{n}{2}$ sets of pairs to be considered.
An estimator for τ is, therefore, given by:

$$\hat{\tau} = \sum_{1 \leq i} \sum_{< j \leq n} \frac{I_{ij}}{\binom{n}{2}}.$$

Kendall (1970), and many other authors, use $\hat{\tau}$ in yet another form, namely:

Let c denote the number of concordant pairs (corresponding to $I_{ij} = 1$) and d denote the number of discordant pairs (corresponding to $I_{ij} = -1$), then Kendall's tau for the sample is defined as

$$\hat{\tau} = \frac{c - d}{c + d} = \frac{c - d}{\binom{n}{2}}.$$

When ties exist in the data, the following adjusted formula is used for $\hat{\tau}$ (Kendall, 1970):

$$\hat{\tau} = \frac{c - d}{\sqrt{n(n-1)/2 - T} \cdot \sqrt{n(n-1)/2 - U}},$$

where $T = \frac{1}{2} \sum_t t(t-1)$ and t is the number of X observations that are tied at a given rank and $U = \frac{1}{2} \sum_u u(u-1)$, where u is the number of Y observations tied at a specific rank.

Remark: Many of the current statistical software can compute $\hat{\tau}$ for both untied and tied cases.

6.5.2 Spearman's rho

Spearman's rho, denoted by ρ_s, was first discussed by the psychologist Charles Spearman (1904). If X and Y are random variables with respective distribution functions $F(x)$ and $G(y)$, then the population version of Spearman's rho is defined to be the Pearson correlation coefficient between the random variables $F(X)$ and $G(Y)$:

$$\rho_s(X,Y) = \rho_p(F(X), G(Y)).$$

Because of this fact, Spearman's rho is sometimes referred to as the rank or grade coefficient of correlation.

Like Kendall's tau, ρ_s is also based on the concepts of concordance and discordance.

Definition 6.5.3 *Let $(X_1, Y_1), (X_2, Y_2)$ and (X_3, Y_3) be three independent random vectors with a common joint distribution $H(x, y)$. Spearman's rho coefficient associated with the vector (X, Y) distributed according to H, is then defined as (see Kruskall, 1958):*

$$\rho_s(X,Y) = 3[P((X_1 - X_2)(Y_1 - Y_3) > 0) - P((X_1 - X_2)(Y_1 - Y_3) < 0)].$$

The population version of Spearman's rho can also be expressed in terms of the copula C associated with the vector (X, Y) as follows (Schweizer and Wolff, 1981):

$$\begin{aligned}\rho_s(X,Y) &= 12\int_0^1\int_0^1 C(u,v)dudv - 3 \\ &= 12\int_0^1\int_0^1 uvdC(u,v) - 3.\end{aligned}$$

Table 6.1 contains expressions for ρ_s (where possible) for the various families of copulas discussed in Section 6.4.

Sample version of Spearman's rho

Let $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ be a random sample of n observations drawn from a bivariate distribution.

Let $\{R_i^x,\ i = 1, 2, \ldots, n\}$ and $\{R_i^y,\ i = 1, 2, \ldots, n\}$ denote the ranks of $\{X_i,\ i = 1, 2, ..., n\}$ and $\{Y_i,\ i = 1, 2, ..., n\}$, respectively. An estimator for ρ_s, denoted by $\hat{\rho}_s$, is then simply the Pearson correlation coefficient between these ranks:

$$\hat{\rho}_s = \frac{\sum_{i=1}^n (R_i^x - \overline{R}^x)(R_i^y - \overline{R}^y)}{\sqrt{\sum_{i=1}^n (R_i^x - \overline{R}^x)^2}\sqrt{\sum_{i=1}^n (R_i^y - \overline{R}^y)^2}}, \tag{6.3}$$

where $\overline{R}^x = \frac{1}{n}\sum_{i=1}^{n} R_i^x$ and $\overline{R}^y = \frac{1}{n}\sum_{i=1}^{n} R_i^y$. If there are no tied ranks in the data, then (6.3) becomes

$$\hat{\rho}_s = 1 - \frac{6\sum_{i=1}^{n} D_i^2}{n(n^2-1)}, \tag{6.4}$$

where D_i, $i = 1, 2, \ldots, n$, are the differences between the ranks of corresponding values of X and Y (see, e.g., Kruskall, 1958). If tied ranks exist, Pearson's correlation coefficient between the ranks, equation (6.3), has to be used to calculate Spearman's rho, instead of equation (6.4). In this case, each member of a group of tied observations is assigned the average of the ranks they would have if they were distinguishable. This method of assigning ranks to tied observations is referred to as the *midranks method* (see, e.g., Noether, 1967; Kendall, 1970).

Table 6.1: Kendall's tau and Spearman's rho for some families of copulas/bivariate distributions.

Family	Kendall's tau (τ)	Spearman's rho (ρ_s)
Normal	$\frac{2}{\pi}\arcsin(\theta)$	$\frac{6}{\pi}\arcsin\left(\frac{\theta}{2}\right)$
FGM	$\frac{2}{9}\theta$	$\frac{1}{3}\theta$
Plackett	No closed form	$\frac{\theta+1}{\theta-1}-\frac{2\theta\ln\theta}{(\theta-1)^2}$
Cuadras-Augé	$\frac{\theta}{2-\theta}$	$\frac{3\theta}{4-\theta}$
Ali-Mikhail-Haq	$\frac{3\theta-2}{3\theta}-\frac{2(1-\theta)^2}{3\theta^2}\ln(1-\theta)$	$\frac{12(1+\theta)}{\theta^2}\int_1^{1-\theta}\frac{\ln t}{1-t}dt$ $-\frac{24(1-\theta)}{\theta^2}\ln(1-\theta)-\frac{3(\theta+12)}{\theta}$
Clayton	$\frac{\theta}{\theta+2}$	No closed form
Raftery	$\frac{2\theta}{3-\theta}$	$\frac{\theta(4-3\theta)}{(2-\theta)^2}$

Chapter 7

A new semiparametric and nonparametric bootstrap test for Spearman's rho

In this chapter bootstrap-based tests for Spearman's rho are considered. We propose two new tests for ρ_s, viz. a semiparametric bootstrap test based on copulas and a nonparametric bootstrap test. We conclude the chapter by proposing a nonparametric bootstrap test for the equality of the Spearman's rho of two bivariate populations.

7.1 Introduction

Let $\mathbf{Z}_n = \big((X_1, Y_1), \ldots, (X_n, Y_n)\big)$ be a random sample from a population (X, Y) and denote by $H(x, y)$ the joint distribution function of (X, Y). We once again use the notation $F(x)$ and $G(y)$ for the marginal distribution functions of X and Y, respectively.

Suppose we wish to test the hypothesis:

$$H_0 : \rho_s(X,Y) = \rho_{s_0} \quad \text{vs.} \quad H_A : \rho_s(X,Y) > \rho_{s_0}.$$

The test rejects H_0 if and only if

$$T_n(\mathbf{Z}_n) \geq C_n(\alpha),$$

where

$$P_{H_0}(T_n(\mathbf{Z}_n) \geq C_n(\alpha)) \cong \alpha.$$

Possible test statistics for this testing scenario are

$$T_n^{(1)}(\mathbf{Z}_n) = \hat{\rho}_s(\mathbf{Z}_n) \quad \text{and} \quad T_n^{(2)}(\mathbf{Z}_n) = \sqrt{(n-3)}\left(z_{\hat{\rho}_s} - z_0 - \frac{\rho_{s0}}{2(n-1)}\right),$$

where $z_{\hat{\rho}_s}$ and z_0 denote Fisher's Z transformation (see Fisher, 1921) given by:

$$z_{\hat{\rho}_s} = \frac{1}{2}\ln\left(\frac{1+\hat{\rho}_s(\mathbf{Z}_n)}{1-\hat{\rho}_s(\mathbf{Z}_n)}\right) \text{ and } z_0 = \frac{1}{2}\ln\left(\frac{1+\rho_{s0}}{1-\rho_{s0}}\right).$$

If $H(x,y)$ is the bivariate normal distribution function, then $T_n^{(2)}(\mathbf{Z}_n)$ follows approximately a standard normal distribution (Fisher, 1970; Anderson, 1984) and $C_n(\alpha) \approx \Phi^{-1}(1-\alpha)$.
However, if $H(x,y)$ is not the bivariate normal distribution function, then the distribution of $T_n(\mathbf{Z}_n)$ is unknown.
Next, we propose a semiparametric bootstrap test for Spearman's rho.

7.2 Semiparametric bootstrap test

Again, let $C_\theta(u,v)$ be a parametric family of copulas indexed by a single parameter θ. Assume further that $F(x)$ and $G(y)$ are unknown nuisance parameters and that

$$H(x,y) = C_\theta(F(x), G(y)).$$

First, note that testing the hypothesis

$$H_0 : \rho_s(X,Y) = \rho_{s0} \quad \text{vs.} \quad H_A : \rho_s(X,Y) > \rho_{s0}$$

is equivalent to testing

$$H_0 : \theta = \theta_0 \quad \text{vs.} \quad H_A : \theta > \theta_0,$$

for some $\theta_0 = \theta_0(\rho_{s0})$.
The test rejects H_0 if and only if

$$\hat{\rho}_s(\mathbf{Z}_n) \geq C_n(\alpha),$$

where

$$P_{H_0}(\hat{\rho}_s(\mathbf{Z}_n) \geq C_n(\alpha)) \cong \alpha.$$

The critical value $C_n(\alpha)$ is unknown, since H is unknown (F and G are unknown). However, a bootstrap estimator $C_n(\alpha; \mathbf{Z}_n)$ can be derived for $C_n(\alpha)$.
The following procedure can be used to obtain $C_n(\alpha; \mathbf{Z}_n)$:

1) Simulate (U_i, V_i), $i = 1, 2, \ldots, n$ from C_{θ_0}, using the techniques discussed in Chapter 6.

2) Estimate $F(x)$ and $G(y)$ by $F_n(x)$ and $G_n(y)$, the e.d.f.'s of $\{X_i,\ i=1,2,\ldots,n\}$ and $\{Y_i,\ i=1,2,\ldots,n\}$, respectively.

3) Set $V_i^{x0*} = F_n^{-1}(U_i)$ and $V_i^{y0*} = G_n^{-1}(V_i),\ i=1,2,\ldots,n$.

This can be done as follows:

Obtain the order statistics $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$, then

$$V_i^{x0*} = \begin{cases} X_{(nU_i)} \text{ if } nU_i \text{ is an integer} \\ X_{(\lfloor nU_i \rfloor + 1)} \text{ if } nU_i \text{ is not an integer.} \end{cases}$$

The V_i^{y0*} can be obtained similarly.

(4) Define $C_n(\alpha; \mathbf{Z}_n)$ by

$$P_{H_0^*}^*(\hat{\rho}_s(\mathbf{Z}_n^{0*}) \geq C_n(\alpha; \mathbf{Z}_n)) \cong \alpha,$$

where

$$\mathbf{Z}_n^{0*} = ((V_1^{x0*}, V_1^{y0*}), (V_2^{x0*}, V_2^{y0*}), \ldots, (V_n^{x0*}, V_n^{y0*})).$$

We call this a semiparametric bootstrap test because a parametric family of copulas is assumed, but $F(x)$ and $G(y)$ are unknown nuisance parameters.
In the next section we propose a completely nonparametric bootstrap test for ρ_s.

7.3 Nonparametric bootstrap test

Again, assume that $F(x)$ and $G(y)$ are unknown nuisance parameters.
Note that testing the hypothesis

$$H_0 : \rho_s(X,Y) = \rho_{s0} \quad \text{vs.} \quad H_A : \rho_s(X,Y) > \rho_{s0}$$

is equivalent to testing

$$H_0 : \rho_p(F(X), G(Y)) = \rho_{s0} \quad \text{vs.} \quad H_A : \rho_p(F(X), G(Y)) > \rho_{s0}.$$

The test rejects H_0 if and only if

$$\hat{\rho}_s(\mathbf{Z}_n) \geq C_n(\alpha),$$

where

$$P_{H_0}(\hat{\rho}_s(\mathbf{Z}_n) \geq C_n(\alpha)) \cong \alpha.$$

Consider the following transformation of (X,Y):

$$
\begin{aligned}
X^0 &= \sqrt{3}(F(X)-0.5)\left[\sqrt{\frac{1+\rho_{s0}}{1+\rho_s(X,Y)}}+\sqrt{\frac{1-\rho_{s0}}{1-\rho_s(X,Y)}}\right]+ \\
&\quad \sqrt{3}(G(Y)-0.5)\left[\sqrt{\frac{1+\rho_{s0}}{1+\rho_s(X,Y)}}-\sqrt{\frac{1-\rho_{s0}}{1-\rho_s(X,Y)}}\right] \qquad (7.1)
\end{aligned}
$$

$$
\begin{aligned}
Y^0 &= \sqrt{3}(F(X)-0.5)\left[\sqrt{\frac{1+\rho_{s0}}{1+\rho_s(X,Y)}}-\sqrt{\frac{1-\rho_{s0}}{1-\rho_s(X,Y)}}\right]+ \\
&\quad \sqrt{3}(G(Y)-0.5)\left[\sqrt{\frac{1+\rho_{s0}}{1+\rho_s(X,Y)}}+\sqrt{\frac{1-\rho_{s0}}{1-\rho_s(X,Y)}}\right]. \qquad (7.2)
\end{aligned}
$$

Theorem 7.3.1 $\rho_p(X^0, Y^0) = \rho_{s0}$.

Proof. It follows, after some algebra, that $E(X^0Y^0) = \rho_{s0}$, $E(X^0) = E(Y^0) = 0$ and $E((X^0)^2) = E((Y^0)^2) = 1$. Hence,

$$\rho_p(X^0, Y^0) = \frac{Cov(X^0, Y^0)}{\sqrt{Var(X^0)}\sqrt{Var(Y^0)}} = \rho_{s0}.$$

Q.E.D.■

Consider data $\mathbf{Z}_n^0 = \{(V_i^{x0}, V_i^{y0}),\ i = 1, 2, \ldots, n\}$ where V_i^{x0} is given by the expression in (7.1) with:

- $F(X)$ replaced by $F_n(X_i) = \frac{R_i^x}{n}$ (if F is continuous), where $\{R_i^x,\ i = 1, 2, \ldots, n\}$ denote the ranks of $\{X_i,\ i = 1, 2, \ldots, n\}$.
- $G(Y)$ replaced by $G_n(Y_i) = \frac{R_i^y}{n}$ (if G is continuous), where $\{R_i^y,\ i = 1, 2, \ldots, n\}$ denote the ranks of $\{Y_i,\ i = 1, 2, \ldots, n\}$.
- $\rho_s(X, Y)$ replaced by $\hat{\rho}_s(\mathbf{Z}_n)$.

V_i^{y0} is defined similarly using equation (7.2).

Resampling under H_0^* corresponds to sampling with replacement from the pairs $\mathbf{Z}_n^0 = \{(V_i^{x0}, V_i^{y0}),\ i = 1, 2, \ldots, n\}$.

Refer to these resampled pairs as

$$\mathbf{Z}_n^{0*} = \{(V_i^{x0*}, V_i^{y0*}),\ i = 1, 2, \ldots, n\}.$$

The bootstrap critical value $C_n(\alpha; \mathbf{Z}_n)$ is now chosen such that

$$P^*_{H_0^*}(\hat{\rho}_p(\mathbf{Z}_n^{0*}) \geq C_n(\alpha; \mathbf{Z}_n)) \cong \alpha.$$

We can approximate $C_n(\alpha; \mathbf{Z}_n)$ by $\hat{C}_n(\alpha; \mathbf{Z}_n)$ as follows:

1) Obtain a first bootstrap sample $\mathbf{Z}_n^{0*}$ and calculate $\hat{\rho}_p(\mathbf{Z}_n^{0*})$ (Pearson's correlation between $\{V_i^{x0*},\ i = 1, 2, \ldots, n\}$ and $\{V_i^{y0*},\ i = 1, 2, \ldots, n\}$). Denote this by T_1^*. It is important to note that one needs to calculate $\hat{\rho}_p(\mathbf{Z}_n^{0*})$ and not $\hat{\rho}_s(\mathbf{Z}_n^{0*})$.

2) Independently repeat step 1) a number B times to obtain B bootstrap replications $T_1^*, T_2^*, \ldots, T_B^*$.

3) Obtain the order statistics $T_{(1)}^* \leq T_{(2)}^* \leq \cdots \leq T_{(B)}^*$.

4) Set $\hat{C}_n(\alpha; \mathbf{Z}_n) = T^*_{(\lfloor B(1-\alpha) \rfloor)}$.

Hence, reject the null hypothesis if and only if

$$\hat{\rho}_s(\mathbf{Z}_n) \geq \hat{C}_n(\alpha; \mathbf{Z}_n).$$

The bootstrap p-value for testing

$$H_0 : \rho_s(X, Y) = \rho_{s0} \quad \text{vs.} \quad H_A : \rho_s(X, Y) > \rho_{s0}$$

can be approximated by

$$\frac{1}{B} \sum_{b=1}^{B} I(T_b^* \geq \hat{\rho}_s(\mathbf{Z}_n)).$$

Chapter 8 contains the results of a Monte-Carlo study conducted to compare the performance of the test based on the assumption that $H(x, y)$ is the bivariate normal distribution (discussed in Section 7.1) with the semiparametric and nonparametric bootstrap tests. The semiparametric test was developed as a useful benchmark tool for measuring the performance of the nonparametric test.

7.4 Nonparametric bootstrap test for the equality of two Spearman's rho's

Let $\mathbf{Z}_n^{(1)} = ((X_1^{(1)}, Y_1^{(1)}), \ldots, (X_n^{(1)}, Y_n^{(1)}))$ be a random sample from a population $(X^{(1)}, Y^{(1)})$. Denote by $H_1(x,y)$ the joint distribution function of $(X^{(1)}, Y^{(1)})$ and by $F_1(x)$ and $G_1(y)$ the marginal distribution functions of $X^{(1)}$ and $Y^{(1)}$, respectively.
Let $\mathbf{Z}_m^{(2)} = ((X_1^{(2)}, Y_1^{(2)}), \ldots, (X_m^{(2)}, Y_m^{(2)}))$ be a random sample from a population $(X^{(2)}, Y^{(2)})$. Denote by $H_2(x,y)$ the joint distribution function of $(X^{(2)}, Y^{(2)})$ and by $F_2(x)$ and $G_2(y)$ the marginal distribution functions of $X^{(2)}$ and $Y^{(2)}$, respectively. Let $\mathbf{Z}_{n+m} = \left(\mathbf{Z}_n^{(1)}, \mathbf{Z}_m^{(2)}\right)$.
Consider testing the hypothesis:

$$H_0 : \rho_s(X^{(1)}, Y^{(1)}) = \rho_s(X^{(2)}, Y^{(2)}) \quad \text{vs.} \quad H_A : \rho_s(X^{(1)}, Y^{(1)}) > \rho_s(X^{(2)}, Y^{(2)}). \tag{7.3}$$

The test rejects H_0 if and only if

$$\hat{\rho}_s(\mathbf{Z}_n^{(1)}) - \hat{\rho}_s(\mathbf{Z}_m^{(2)}) \geq C_n(\alpha),$$

where

$$P_{H_0}\left(\hat{\rho}_s(\mathbf{Z}_n^{(1)}) - \hat{\rho}_s(\mathbf{Z}_m^{(2)}) \geq C_n(\alpha)\right) \cong \alpha.$$

The critical value $C_n(\alpha)$ is unknown, since $H_1(x,y)$ and $H_2(x,y)$ are unknown. However, a bootstrap estimator $C_n(\alpha; \mathbf{Z}_n^{(1)}, \mathbf{Z}_m^{(2)})$ can be derived for $C_n(\alpha)$:
First, note that testing (7.3) is equivalent to testing

$$H_0 : \rho_p(F_1(X^{(1)}), G_1(Y^{(1)})) = \rho_p(F_2(X^{(2)}), G_2(Y^{(2)}))$$

vs.

$$H_A : \rho_p(F_1(X^{(1)}), G_1(Y^{(1)})) > \rho_p(F_2(X^{(2)}), G_2(Y^{(2)})).$$

Consider the following transformation of $(X^{(1)}, Y^{(1)})$ and $(X^{(2)}, Y^{(2)})$:

$$\begin{aligned} X^{(1)0} &= \sqrt{3}(F_1(X^{(1)}) - 0.5)\left[\sqrt{\frac{1+\hat{\rho}_s(\mathbf{Z}_{n+m})}{1+\rho_s(X^{(1)}, Y^{(1)})}} + \sqrt{\frac{1-\hat{\rho}_s(\mathbf{Z}_{n+m})}{1-\rho_s(X^{(1)}, Y^{(1)})}}\right] + \\ &\quad \sqrt{3}(G_1(Y^{(1)}) - 0.5)\left[\sqrt{\frac{1+\hat{\rho}_s(\mathbf{Z}_{n+m})}{1+\rho_s(X^{(1)}, Y^{(1)})}} - \sqrt{\frac{1-\hat{\rho}_s(\mathbf{Z}_{n+m})}{1-\rho_s(X^{(1)}, Y^{(1)})}}\right] \end{aligned} \tag{7.4}$$

$$\begin{aligned} Y^{(1)0} &= \sqrt{3}(F_1(X^{(1)}) - 0.5)\left[\sqrt{\frac{1+\hat{\rho}_s(\mathbf{Z}_{n+m})}{1+\rho_s(X^{(1)}, Y^{(1)})}} - \sqrt{\frac{1-\hat{\rho}_s(\mathbf{Z}_{n+m})}{1-\rho_s(X^{(1)}, Y^{(1)})}}\right] + \\ &\quad \sqrt{3}(G_1(Y^{(1)}) - 0.5)\left[\sqrt{\frac{1+\hat{\rho}_s(\mathbf{Z}_{n+m})}{1+\rho_s(X^{(1)}, Y^{(1)})}} + \sqrt{\frac{1-\hat{\rho}_s(\mathbf{Z}_{n+m})}{1-\rho_s(X^{(1)}, Y^{(1)})}}\right] \end{aligned} \tag{7.5}$$

$$X^{(2)0} = \sqrt{3}(F_2(X^{(2)}) - 0.5)\left[\sqrt{\frac{1+\hat{\rho}_s(\mathbf{Z}_{n+m})}{1+\rho_s(X^{(2)},Y^{(2)})}} + \sqrt{\frac{1-\hat{\rho}_s(\mathbf{Z}_{n+m})}{1-\rho_s(X^{(2)},Y^{(2)})}}\right] + \sqrt{3}(G_2(Y^{(2)}) - 0.5)\left[\sqrt{\frac{1+\hat{\rho}_s(\mathbf{Z}_{n+m})}{1+\rho_s(X^{(2)},Y^{(2)})}} - \sqrt{\frac{1-\hat{\rho}_s(\mathbf{Z}_{n+m})}{1-\rho_s(X^{(2)},Y^{(2)})}}\right] \quad (7.6)$$

$$Y^{(2)0} = \sqrt{3}(F_2(X^{(2)}) - 0.5)\left[\sqrt{\frac{1+\hat{\rho}_s(\mathbf{Z}_{n+m})}{1+\rho_s(X^{(2)},Y^{(2)})}} - \sqrt{\frac{1-\hat{\rho}_s(\mathbf{Z}_{n+m})}{1-\rho_s(X^{(2)},Y^{(2)})}}\right] + \sqrt{3}(G_2(Y^{(2)}) - 0.5)\left[\sqrt{\frac{1+\hat{\rho}_s(\mathbf{Z}_{n+m})}{1+\rho_s(X^{(2)},Y^{(2)})}} + \sqrt{\frac{1-\hat{\rho}_s(\mathbf{Z}_{n+m})}{1-\rho_s(X^{(2)},Y^{(2)})}}\right]. \quad (7.7)$$

It follows from Theorem 7.3.1 that:

$$\rho_p(X^{(1)0}, Y^{(1)0}) = \rho_p(X^{(2)0}, Y^{(2)0}) = \hat{\rho}_s(\mathbf{Z}_{n+m}).$$

Denote by F_{1n}, G_{1n}, F_{2m} and G_{2m} the empirical d.f.'s of $\{X_i^{(1)},\ i = 1,2,\ldots,n\}$, $\{Y_i^{(1)},\ i = 1,2,\ldots,n\}$, $\{X_i^{(2)},\ i = 1,2,\ldots,m\}$ and $\{Y_i^{(2)},\ i = 1,2,\ldots,m\}$, respectively.

Consider data $\mathbf{Z}^0_{n,1} = \{(V^{x0}_{i,1}, V^{y0}_{i,1}),\ i = 1,2,\ldots n\}$ and $\mathbf{Z}^0_{m,2} = \{(V^{x0}_{i,2}, V^{y0}_{i,2}),\ i = 1,2,\ldots,m\}$, where $\{V^{x0}_{i,1},\ i = 1,2,\ldots,n\}$ are given by the expression in (7.4) with:

- $F_1(X^{(1)})$ replaced by $F_{1n}(X_i^{(1)}) = \dfrac{R^x_{i,1}}{n}$ (if F_1 is continuous), where $\{R^x_{i,1},\ i = 1,2,\ldots,n\}$ denote the ranks of $\{X_i^{(1)},\ i = 1,2,\ldots,n\}$.
- $G_1(Y^{(1)})$ replaced by $G_{1n}(Y_i^{(1)}) = \dfrac{R^y_{i,1}}{n}$ (if G_1 is continuous), where $\{R^y_{i,1},\ i = 1,2,\ldots,n\}$ denote the ranks of $\{Y_i^{(1)},\ i = 1,2,\ldots,n\}$.
- $\rho_s(X^{(1)}, Y^{(1)})$ replaced by $\hat{\rho}_s(\mathbf{Z}_n^{(1)})$.

$\{V^{y0}_{i,1},\ i = 1,2,\ldots,n\}$ are defined similarly using expression (7.5).
$\{V^{x0}_{i,2},\ i = 1,2,\ldots,m\}$ are given by the expression in (7.6) with:

- $F_2(X^{(2)})$ replaced by $F_{2m}(X_i^{(2)}) = \dfrac{R^x_{i,2}}{m}$ (if F_2 is continuous), where $\{R^x_{i,2},\ i = 1,2,\ldots,m\}$ denote the ranks of $\{X_i^{(2)},\ i = 1,2,\ldots,m\}$.

- $G_2(Y^{(2)})$ replaced by $G_{2m}(Y_i^{(2)}) = \dfrac{R_{i,2}^y}{m}$ (if G_2 is continuous), where $\{R_{i,2}^y,\ i = 1, 2, \ldots, m\}$ denote the ranks of $\{Y_i^{(2)},\ i = 1, 2, \ldots, m\}$.

- $\rho_s(X^{(2)}, Y^{(2)})$ replaced by $\hat{\rho}_s(\mathbf{Z}_m^{(2)})$.

$\{V_{i,2}^{y0}, i = 1, 2, \ldots, m\}$ are defined similarly using expression (7.7).

Resampling under H_0^* now corresponds to sampling with replacement from the pairs $\mathbf{Z}_{n,1}^0$ and $\mathbf{Z}_{m,2}^0$, separately. Refer to such resampled pairs as

$$\begin{aligned}\mathbf{Z}_{n,1}^{0*} &= \{(V_{i,1}^{x0*}, V_{i,1}^{y0*}),\ i = 1, 2, \ldots, n\} \text{ and}\\ \mathbf{Z}_{m,2}^{0*} &= \{(V_{i,2}^{x0*}, V_{i,2}^{y0*}),\ i = 1, 2, \ldots, m\}.\end{aligned}$$

Define $C_n(\alpha; \mathbf{Z}_n^{(1)}, \mathbf{Z}_m^{(2)})$ by

$$P_{H_0^*}^*\left(\hat{\rho}_p(\mathbf{Z}_{n,1}^{0*}) - \hat{\rho}_p(\mathbf{Z}_{m,2}^{0*}) \geq C_n(\alpha; \mathbf{Z}_n^{(1)}, \mathbf{Z}_m^{(2)})\right) \cong \alpha.$$

We can approximate $C_n(\alpha; \mathbf{Z}_n^{(1)}, \mathbf{Z}_m^{(2)})$ by $\hat{C}_n(\alpha; \mathbf{Z}_n^{(1)}, \mathbf{Z}_m^{(2)})$ as follows:

1) Obtain bootstrap samples $\mathbf{Z}_{n,1}^{0*}$ and $\mathbf{Z}_{m,2}^{0*}$. Calculate $T_1^* = \hat{\rho}_p(\mathbf{Z}_{n,1}^{0*}) - \hat{\rho}_p(\mathbf{Z}_{m,2}^{0*})$.

2) Independently repeat step 1) a number B times to obtain B bootstrap replications $T_1^*, T_2^*, \ldots, T_B^*$.

3) Obtain the order statistics $T_{(1)}^* \leq T_{(2)}^* \leq \cdots \leq T_{(B)}^*$.

4) Set $\hat{C}_n(\alpha; \mathbf{Z}_n^{(1)}, \mathbf{Z}_m^{(2)}) = T_{(\lfloor B(1-\alpha) \rfloor)}^*$.

The bootstrap p-value for testing

$$H_0 : \rho_s(X^{(1)}, Y^{(1)}) = \rho_s(X^{(2)}, Y^{(2)}) \quad \text{vs.} \quad H_A : \rho_s(X^{(1)}, Y^{(1)}) > \rho_s(X^{(2)}, Y^{(2)})$$

can be approximated by

$$\frac{1}{B}\sum_{b=1}^{B} I(T_b^* \geq \hat{\rho}_s(\mathbf{Z}_n^{(1)}) - \hat{\rho}_s(\mathbf{Z}_m^{(2)})).$$

Chapter 8 contains the results of a Monte-Carlo study where we compare the performance of this nonparametric bootstrap test with the following parametric test discussed in Neter et al. (1996): Let

$$T(\mathbf{Z}_n^{(1)}, \mathbf{Z}_m^{(2)}) = \frac{z_1 - z_2}{\sqrt{\frac{1}{n-3} + \frac{1}{m-3}}},$$

where z_1 and z_2 denotes Fisher's Z transformation given by

$$\begin{aligned} z_1 &= \frac{1}{2}\ln\left(\frac{1+\hat{\rho}_s(\mathbf{Z}_n^{(1)})}{1-\hat{\rho}_s(\mathbf{Z}_n^{(1)})}\right) \quad \text{and} \\ z_2 &= \frac{1}{2}\ln\left(\frac{1+\hat{\rho}_s(\mathbf{Z}_m^{(2)})}{1-\hat{\rho}_s(\mathbf{Z}_m^{(2)})}\right). \end{aligned}$$

If $H_1(x,y)$ and $H_2(x,y)$ are independent bivariate normal distribution functions, then under H_0, $T(\mathbf{Z}_n^{(1)},\mathbf{Z}_m^{(2)})$ is distributed approximately as a standard normal variable and $C_n(\alpha) \approx \Phi^{-1}(1-\alpha)$.

Chapter 8

Empirical studies

8.1 Introduction

Recall that simulation results for testing the mean in the univariate case were presented and discussed in Chapter 4. In this chapter we present the results of some further Monte-Carlo studies. Section 8.2 displays the simulation results for the following tests:

- The variance in the univariate case. (Discussed in Sections 3.2.2(e) and 3.3.2.)
- Spearman's rho for a bivariate distribution. (Discussed in Sections 7.2 and 7.3.)
- The equality of two Spearman's rho's. (Discussed in Section 7.4.)

All of these tests were evaluated by using evaluation Method I and Method II.
A nominal significance level of $\alpha = 0.05$ was used throughout and all bootstrap critical value calculations were based on $B = 1000$ independent bootstrap replications. All calculations were done using double precision arithmetic in FORTRAN and routines from the IMSL library.

For the tests evaluated by using Method I, all size (or power) estimates were calculated as the proportion of 20 000 Monte-Carlo samples that resulted in H_0 (or H_A) being rejected, using the associated bootstrap critical values. All standard errors of the estimated probabilities are less than or equal to $\sqrt{0.25/20\ 000} = 0.0035$.

For the tests evaluated by using Method II, the estimated sizes (or powers) were calculated according to

$$\overline{p} = \frac{1}{MC_1} \sum_{j=1}^{MC_1} p_j,$$

where

$$p_j = \frac{1}{MC_2} \sum_{i=1}^{MC_2} I\left(T_n(\mathbf{V}_{n,i}^0) \geq \hat{C}_n(\alpha; \mathbf{X}_n^j)\right), j = 1, 2, \ldots, MC_1,$$

for estimated sizes and

$$p_j = \frac{1}{MC_2} \sum_{i=1}^{MC_2} I\left(T_n(\mathbf{V}_{n,i}^A) \geq \hat{C}_n(\alpha; \mathbf{X}_n^j)\right), j = 1, 2, \ldots, MC_1,$$

for estimated powers. Here, T_n refers to the test statistic under consideration, whilst $\hat{C}_n(\alpha; \mathbf{X}_n^j)$ refers to the bootstrap critical value. The standard errors of the estimated probabilities were calculated according to

$$\sqrt{\frac{\frac{1}{MC_1} \sum_{j=1}^{MC_1} (p_j - \overline{p})^2}{MC_1}}.$$

In each of the Monte-Carlo studies we took $MC_1 = 10\ 000$ and $MC_2 = 5\ 000$. All standard errors were found to be negligibly small and are not reported in the tables.

In Section 8.3 we present the results of a Monte-Carlo study where the bootstrap estimate of power is compared to the Monte-Carlo approximation of the true power for testing the mean in the univariate case (refer to Section 3.2.2(a)).

The chapter concludes with a few concluding remarks concerning bootstrap hypothesis testing and future research in this area.

8.2 Monte-Carlo results for various tests

8.2.1 The variance in the univariate case

Consider testing the hypothesis

$$H_0 : \sigma^2 = \sigma_0^2 \quad \text{vs.} \quad H_A : \sigma^2 > \sigma_0^2. \tag{8.1}$$

In this section we present the results of a simulation study where we compare the estimated sizes of the tests (R), (W) and (T). Here, (R) refers to the test where the bootstrap critical value is obtained by resampling from transformed data in order to "mimic" H_0 (see expression (3.10)), (W) refers to the test where the bootstrap critical value is obtained by resampling from the original data (see expression (3.11)) and (T) refers to the test where the bootstrap critical value is obtained by using the exponentially *tilted* version of the e.d.f. (see Section 3.3.2). Sample sizes $n = 30, 50, 100, 200$ and 400 were used.

The results displayed in Tables 8.1–8.5 were based on the statistic defined in Section 3.2.2(e), viz.

$$T_n(\mathbf{X}_n) = \frac{nS_n^2(\mathbf{X}_n)}{\sigma_0^2}.$$

Data were generated from four distributions:

- Normal distribution with mean 0 and variance σ^2;
- Uniform distribution with mean 0 and variance σ^2;
- Double exponential distribution with mean 0 and variance σ^2;
- Exponential distribution with mean 1 and variance σ^2.

The σ_0^2 column in Tables 8.2–8.5 refers to the pseudo test data $\mathbf{V}_n^0$ generated from a distribution with parameter σ_0^2.

Table 8.1: Estimated size for testing the variance using evaluation Method I.

Distribution	σ_0^2	n	(R)	(W)	(T)
Normal	1.0	30	0.089	0.000	0.057
		50	0.078	0.000	0.054
		100	0.069	0.000	0.054
		200	0.058	0.000	0.051
		400	0.054	0.000	0.051
Uniform	1.0	30	0.107	0.000	0.052
		50	0.092	0.000	0.052
		100	0.082	0.000	0.051
		200	0.074	0.000	0.052
		400	0.061	0.000	0.050
Double exponential	2.0	30	0.092	0.000	0.061
		50	0.075	0.000	0.058
		100	0.061	0.000	0.056
		200	0.055	0.000	0.055
		400	0.050	0.000	0.054
Exponential	1.0	30	0.099	0.000	0.072
		50	0.082	0.000	0.071
		100	0.066	0.000	0.065
		200	0.054	0.000	0.062
		400	0.049	0.000	0.059

Table 8.2: Estimated size for testing the variance using Method II (normal).

		Data from: $\sigma^2 = 1.0$			Data from: $\sigma^2 = 1.2$		
σ_0^2	n	(R)	(W)	(T)	(R)	(W)	(T)
1.0	30	0.091	0.201	0.082	0.091	0.093	0.070
	50	0.078	0.176	0.070	0.078	0.060	0.060
	100	0.067	0.162	0.060	0.067	0.027	0.055
	200	0.061	0.145	0.056	0.061	0.009	0.053
	400	0.056	0.136	0.053	0.056	0.001	0.052

Table 8.3: Estimated size for testing the variance using Method II (uniform).

		Data from: $\sigma^2 = 1.0$			Data from: $\sigma^2 = 1.1$		
σ_0^2	n	(R)	(W)	(T)	(R)	(W)	(T)
1.0	30	0.079	0.166	0.056	0.078	0.087	0.044
	50	0.071	0.154	0.053	0.071	0.060	0.042
	100	0.064	0.142	0.052	0.064	0.032	0.042
	200	0.059	0.138	0.051	0.060	0.014	0.041
	400	0.056	0.136	0.050	0.057	0.004	0.041

Table 8.4: Estimated size for testing the variance using Method II (double exponential).

		Data from: $\sigma^2 = 2.0$			Data from: $\sigma^2 = 2.2$		
σ_0^2	n	(R)	(W)	(T)	(R)	(W)	(T)
2.0	30	0.111	0.248	0.110	0.111	0.194	0.111
	50	0.096	0.214	0.097	0.096	0.165	0.091
	100	0.082	0.189	0.079	0.082	0.120	0.078
	200	0.070	0.169	0.068	0.070	0.084	0.068
	400	0.062	0.157	0.061	0.062	0.053	0.064

Table 8.5: Estimated size for testing the variance using Method II (exponential).

		Data from: $\sigma^2 = 1.0$			Data from: $\sigma^2 = 1.2$		
σ_0^2	n	(R)	(W)	(T)	(R)	(W)	(T)
	30	0.129	0.268	0.168	0.130	0.196	0.155
	50	0.113	0.247	0.138	0.112	0.159	0.122
1.0	100	0.095	0.207	0.104	0.095	0.105	0.100
	200	0.081	0.188	0.085	0.081	0.065	0.086
	400	0.072	0.171	0.072	0.071	0.033	0.078

Various conclusions can be drawn from Tables 8.1 – 8.5:

Method I

(1) The estimated sizes of the test (W) are 0, even for small sample sizes. This agrees with the result of Theorem 3.2.1.

(2) The estimated sizes of the test (T) are in most of the cases closer to the nominal significance level than those of (R), especially for smaller sample sizes.

Method II

(3) When the data $\mathbf{X}_n$ are generated from a distribution with the parameter specified by the *null hypothesis*, (W) perform extremely poorly. Even for very large sample sizes the estimated sizes do not come close to the nominal significance level. If the data $\mathbf{X}_n$ are generated from a distribution with the parameter specified by the *alternative hypothesis*, the estimated sizes of the test (W) converge to 0 as $n \to \infty$.

(4) It is clear that the test (R) produces the same results when the data are generated from a distribution with the parameter specified by the null and alternative hypotheses. This finding is not surprising, because $C_n^R(\alpha; \mathbf{X}_n)$ is scale-invariant. However, this is not the case for the test (T) (see, for example, Tables 8.3 and 8.5) and the test (W).

(5) In most cases (T) performs slightly better than (R), when the data $\mathbf{X}_n$ are generated from a distribution with the parameter specified by the null hypothesis. The exception being the case when the underlying distribution is exponential (see Table 8.5).

It is clear from both evaluation methods that neither the test (R) nor the test (T) perform satisfactorily for small sample sizes. This is especially true when the underlying distribution is skew and/or has heavy tails. However, testing the variance is a challenging problem, not only for the bootstrap, but also for most other techniques. These findings are in agreement with the findings of Schenker (1985) as far as the construction of bootstrap confidence intervals for the variance is concerned.

However, the statistic considered above is a pivot under normality, but is non-pivotal for other distributions. An asymptotic pivotal statistic is given by

$$T_n(\mathbf{X}_n) = \frac{\sqrt{n}(S_n^2(\mathbf{X}_n) - \sigma_0^2)}{\sqrt{\hat{\mu}_4 - S_n^4(\mathbf{X}_n)}},$$

where

$$\hat{\mu}_4 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^4.$$

From Table 8.6 below, it is clear that the estimated sizes when this statistic is used, are closer to the nominal significance level compared to those when the non-pivotal statistic (Tables 8.1-8.5) is used. This is once again in agreement with the results obtained for the non-pivotal and pivotal statistics used in Chapter 4 for testing the *mean* of a population.

Table 8.6: Estimated size for testing the variance using Method II.

Double exponential Data from: $\sigma^2 = 2$			Exponential Data from: $\sigma^2 = 1$		
σ_0^2	n	(R)	σ_0^2	n	(R)
2.0	30	0.031	1.0	30	0.025
	50	0.037		50	0.030
	100	0.041		100	0.036
	200	0.045		200	0.040
	400	0.047		400	0.043

8.2.2 Spearman's rho

Consider testing the hypothesis

$$H_0 : \rho_s = \rho_{s_0} \quad \text{vs.} \quad H_A : \rho_s > \rho_{s_0}. \tag{8.2}$$

In this section we present the results of a simulation study where we compare the estimated sizes and estimated powers of the tests (S), (N) and (Norm) for the hypothesis in (8.2). Here, (S) refers to the semiparametric test based on copulas (see Section 7.2), (N) refers to the nonparametric bootstrap test (see Section 7.3) and (Norm) refers to the test based on the normal critical values (see Section 7.1). Estimated powers are only presented in cases where estimated sizes of the tests were comparable. Sample sizes $n = 20, 30, 50$ and 100 were used.
Data were generated from two copulas:

- Cuadras-Augé with Spearman's correlation ρ_s;
- Plackett with Spearman's correlation ρ_s;

and from two bivariate distributions:

- Bivariate normal distribution with Spearman's correlation ρ_s;
- Raftery's bivariate exponential distribution with Spearman's correlation ρ_s.

For the semiparametric test we used $F(x) = 1 - e^{-x}$ and $G(y) = 1 - e^{-y}$ as the marginal distributions. The semiparametric test was only evaluated for the data generated from the two copulas.

The ρ_{s_0} (or ρ_{s_A}) column in Tables 8.11–8.15 refers to the pseudo test data generated from a copula/distribution with parameter ρ_{s_0} (or ρ_{s_A}).

Table 8.7: Estimated size for testing Spearman's rho using evaluation Method I.

Copula	ρ_{s_0}	n	(S)	(N)	(Norm)
Cuadras-Augé	$\rho_{s_0} = 0.1$	20	0.052	0.045	0.064
		30	0.050	0.045	0.060
		50	0.054	0.048	0.059
		100	0.054	0.051	0.058
	$\rho_{s_0} = 0.5$	20	0.051	0.042	0.114
		30	0.050	0.042	0.105
		50	0.052	0.041	0.098
		100	0.048	0.042	0.092
Plackett	$\rho_{s_0} = 0.1$	20	0.050	0.045	0.050
		30	0.051	0.047	0.050
		50	0.050	0.046	0.049
		100	0.053	0.050	0.052
	$\rho_{s_0} = 0.5$	20	0.050	0.056	0.067
		30	0.051	0.054	0.062
		50	0.051	0.053	0.064
		100	0.051	0.047	0.063

Table 8.8: Estimated power for testing Spearman's rho using evaluation Method I.

Copula	ρ_{s_A}	n	(S)	(N)	(Norm)
Plackett	$\rho_{s_A} = 0.2$	20	0.108	0.095	0.106
		30	0.132	0.127	0.130
		50	0.171	0.167	0.171
		100	0.267	0.261	0.266
	$\rho_{s_A} = 0.3$	20	0.218	0.206	0.215
		30	0.294	0.271	0.292
		50	0.407	0.389	0.408
		100	0.654	0.641	0.652

Table 8.9: Estimated size for testing Spearman's rho using evaluation Method I.

Distribution	ρ_{s_0}	n	(N)	(Norm)
Normal	$\rho_{s_0} = 0.1$	20	0.048	0.051
		30	0.048	0.052
		50	0.045	0.047
		100	0.050	0.050
	$\rho_{s_0} = 0.5$	20	0.051	0.053
		30	0.051	0.054
		50	0.049	0.051
		100	0.048	0.053
Exponential	$\rho_{s_0} = 0.1$	20	0.046	0.060
		30	0.046	0.058
		50	0.047	0.053
		100	0.047	0.052
	$\rho_{s_0} = 0.5$	20	0.056	0.077
		30	0.054	0.075
		50	0.054	0.076
		100	0.057	0.076

Table 8.10: Estimated power for testing Spearman's rho using evaluation Method I.

Copula	ρ_{s_A}	n	(N)	(Norm)
Normal	$\rho_{s_A} = 0.6$	20	0.138	0.143
		30	0.171	0.176
		50	0.233	0.245
		100	0.381	0.394
	$\rho_{s_A} = 0.7$	20	0.316	0.334
		30	0.448	0.470
		50	0.631	0.665
		100	0.897	0.912

Table 8.11: Estimated size for testing Spearman's rho using evaluation Method II.

Copula	ρ_{s0}	n	Data from: $\rho_s = 0.1$ (S)	Data from: $\rho_s = 0.1$ (N)	Data from: $\rho_s = 0.2$ (S)	Data from: $\rho_s = 0.2$ (N)
Cuadras-Augé	$\rho_{s0} = 0.1$	20	0.051	0.053	0.051	0.046
		30	0.051	0.052	0.051	0.045
		50	0.051	0.052	0.051	0.045
		100	0.051	0.051	0.051	0.045
Plackett	$\rho_{s0} = 0.1$	20	0.051	0.052	0.051	0.050
		30	0.051	0.052	0.051	0.050
		50	0.051	0.051	0.051	0.049
		100	0.051	0.051	0.051	0.049

Table 8.12: Estimated size for testing Spearman's rho using evaluation Method II.

Distribution	ρ_{s0}	n	Data from: $\rho_s = 0.1$ (N)	Data from: $\rho_s = 0.2$ (N)
Normal	$\rho_{s0} = 0.1$	20	0.052	0.052
		30	0.052	0.051
		50	0.051	0.050
		100	0.051	0.050
Exponential	$\rho_{s0} = 0.1$	20	0.053	0.050
		30	0.053	0.050
		50	0.052	0.048
		100	0.052	0.048

Table 8.13: Estimated size for testing Spearman's rho using evaluation Method II.

Copula	ρ_{s_0}	n	Data from: $\rho_s = 0.5$ (S)	Data from: $\rho_s = 0.5$ (N)	Data from: $\rho_s = 0.6$ (S)	Data from: $\rho_s = 0.6$ (N)
Cuadras-Augé	$\rho_{s_0} = 0.5$	20	0.050	0.060	0.050	0.045
		30	0.050	0.055	0.050	0.040
		50	0.051	0.052	0.051	0.040
		100	0.051	0.051	0.051	0.041
Plackett	$\rho_{s_0} = 0.5$	20	0.049	0.052	0.050	0.046
		30	0.050	0.050	0.050	0.045
		50	0.050	0.050	0.051	0.043
		100	0.051	0.049	0.051	0.043

Table 8.14: Estimated size for testing Spearman's rho using evaluation Method II.

Distribution	ρ_{s_0}	n	Data from: $\rho_s = 0.5$ (N)	Data from: $\rho_s = 0.6$ (N)
Normal	$\rho_{s_0} = 0.5$	20	0.051	0.049
		30	0.051	0.049
		50	0.050	0.048
		100	0.050	0.049
Exponential	$\rho_{s_0} = 0.5$	20	0.058	0.055
		30	0.057	0.053
		50	0.057	0.052
		100	0.055	0.052

Table 8.15: Estimated power for testing Spearman's rho using evaluation Method II.

Copula	ρ_s	n	Data from: $\rho_s = 0.1$ (S)	Data from: $\rho_s = 0.1$ (N)	Data from: $\rho_s = 0.2$ (S)	Data from: $\rho_s = 0.2$ (N)
Cuadras-Augé	$\rho_s = 0.2$	20	0.115	0.118	0.115	0.104
		30	0.137	0.138	0.137	0.126
		50	0.175	0.175	0.175	0.163
		100	0.259	0.258	0.259	0.241
	$\rho_s = 0.3$	20	0.222	0.225	0.221	0.206
		30	0.287	0.288	0.288	0.268
		50	0.403	0.403	0.404	0.383
		100	0.628	0.626	0.629	0.602
Plackett	$\rho_s = 0.2$	20	0.110	0.112	0.111	0.108
		30	0.128	0.127	0.127	0.124
		50	0.173	0.173	0.173	0.167
		100	0.262	0.261	0.263	0.255
	$\rho_s = 0.3$	20	0.216	0.217	0.216	0.208
		30	0.286	0.285	0.285	0.277
		50	0.413	0.411	0.412	0.402
		100	0.654	0.652	0.655	0.644

From Tables 8.7 – 8.15 we can conclude the following:

Method I

(1) All estimated sizes of the tests (S) are very close to the nominal significance level, as one would expect.

(2) The test (Norm) performs well (in terms of estimated sizes) when the underlying bivariate distribution/copula has a weak Spearman's correlation ($\rho_s = 0.1$). However, if the underlying distribution has a stronger Spearman's correlation ($\rho_s = 0.5$) the test (Norm) tends to be "liberal".

(3) The estimated sizes of (N) are also close to the nominal significance level, even if the underlying distribution has a relatively strong Spearman's correlation ($\rho_s = 0.5$).

(4) The performances of the three tests are very similar with regard to estimated powers (see Tables 8.8 and 8.10).

Method II

(5) All estimated sizes of (S) are very close to the nominal significance level. These results hold for the data generated from a copula with the parameter specified by the null and alternative hypothesis.

(6) Most of the estimated sizes of (N) are close to the nominal significance level. These results also hold for the data generated from a bivariate distribution/copula with the parameter specified by the null and alternative hypotheses.

(7) In terms of estimated powers, the performances of (S) and (N) are almost identical (see Table 8.15).

Conclusion: Based on the overall good performance with regard to estimated size and power, we recommend that the nonparametric test (N) be used for testing the hypothesis

$$H_0 : \rho_s = \rho_{s_0} \quad \text{vs.} \quad H_A : \rho_s > \rho_{s_0}.$$

Remark: Naturally, if one knows the underlying copula, the semiparametric test should be the best choice for testing the hypothesis in (8.2).

8.2.3 The equality of two Spearman's rho's

Consider the hypothesis

$$H_0 : \rho_{s_1} = \rho_{s_2} \quad \text{vs.} \quad H_A : \rho_{s_1} > \rho_{s_2}. \tag{8.3}$$

In this section we present the results of a Monte-Carlo study where we compare the estimated sizes of the tests based on (N) and (Norm) for the hypothesis in (8.3). Here, (N) refers to the nonparametric bootstrap test (see Section 7.4) and (Norm) refers to the test based on the normal critical values (see Section 7.4). Sample sizes $n = m = 20, 30, 50$ and 100 were used.

Data were generated from two copulas:

- FGM with Spearman's correlation ρ_{s_1} and ρ_{s_2} ;
- Cuadras-Augé with Spearman's correlation ρ_{s_1} and ρ_{s_2};

and from two bivariate distributions:

- Bivariate normal distribution with Spearman's correlation ρ_{s_1} and ρ_{s_2};
- Raftery's bivariate exponential distribution with Spearman's correlation ρ_{s_1} and ρ_{s_2}.

The ρ_s column in Tables 8.17 – 8.18 refers to the pseudo test data generated from a copula/ distribution with parameter ρ_{s_1} and ρ_{s_2}, respectively.

Table 8.16: Estimated size for testing the equality of two Spearman's rho's using evaluation Method I.

Copula	ρ_s	n	(N)	(Norm)
FGM	$\rho_{s_1} = 0.1$, $\rho_{s_2} = 0.1$	20	0.051	0.061
		30	0.047	0.058
		50	0.051	0.056
		100	0.051	0.056
Cuadras-Augé	$\rho_{s_1} = 0.1$, $\rho_{s_2} = 0.1$	20	0.048	0.060
		30	0.050	0.056
		50	0.052	0.057
		100	0.053	0.056
	$\rho_{s_1} = 0.5$, $\rho_{s_2} = 0.5$	20	0.057	0.104
		30	0.052	0.095
		50	0.051	0.090
		100	0.054	0.096
Normal	$\rho_{s_1} = 0.1$, $\rho_{s_2} = 0.1$	20	0.046	0.048
		30	0.051	0.050
		50	0.051	0.051
		100	0.051	0.050
	$\rho_{s_1} = 0.5$, $\rho_{s_2} = 0.5$	20	0.055	0.056
		30	0.054	0.056
		50	0.052	0.054
		100	0.052	0.053
Exponential	$\rho_{s_1} = 0.1$, $\rho_{s_2} = 0.1$	20	0.048	0.051
		30	0.050	0.055
		50	0.050	0.054
		100	0.051	0.053
	$\rho_{s_1} = 0.5$, $\rho_{s_2} = 0.5$	20	0.056	0.077
		30	0.056	0.076
		50	0.055	0.072
		100	0.055	0.071

Table 8.17: Estimated size for testing the equality of two Spearman's rho's using evaluation Method II.

Copula	ρ_s	n	Data from: $\rho_{s_1} = 0.1$, $\rho_{s_2} = 0.1$ (N)	Data from: $\rho_{s_1} = 0.2$, $\rho_{s_2} = 0.1$ (N)
FMG	$\rho_{s_1} = 0.1$, $\rho_{s_2} = 0.1$	20	0.057	0.058
		30	0.055	0.056
		50	0.053	0.055
		100	0.052	0.054
Cuadras-Augé	$\rho_{s_1} = 0.1$, $\rho_{s_2} = 0.1$	20	0.057	0.057
		30	0.055	0.054
		50	0.053	0.052
		100	0.052	0.052
Normal	$\rho_{s_1} = 0.1$, $\rho_{s_2} = 0.1$	20	0.056	0.058
		30	0.055	0.057
		50	0.053	0.054
		100	0.052	0.054
Exponential	$\rho_{s_1} = 0.1$, $\rho_{s_2} = 0.1$	20	0.057	0.057
		30	0.055	0.056
		50	0.054	0.055
		100	0.053	0.053

Table 8.18: Estimated size for testing the equality of two Spearman's rho's using evaluation Method II.

Copula	ρ_s	n	Data from: $\rho_{s_1} = 0.5$, $\rho_{s_2} = 0.5$ (N)	Data from: $\rho_{s_1} = 0.6$, $\rho_{s_2} = 0.5$ (N)
Cuadras-Augé	$\rho_{s_1} = 0.5$, $\rho_{s_2} = 0.5$	20	0.068	0.078
		30	0.064	0.071
		50	0.058	0.065
		100	0.055	0.059
Normal	$\rho_{s_1} = 0.5$, $\rho_{s_2} = 0.5$	20	0.064	0.076
		30	0.062	0.075
		50	0.059	0.071
		100	0.055	0.065
Exponential	$\rho_{s_1} = 0.5$, $\rho_{s_2} = 0.5$	20	0.072	0.083
		30	0.068	0.080
		50	0.064	0.074
		100	0.061	0.070

Various conclusions can be drawn from Tables 8.16 – 8.18 :

Method I

(1) From Table 8.16 it is clear that the test (Norm) performs satisfactorily (in terms of estimated sizes) when the underlying bivariate distributions/copulas have a weak Spearman's correlation ($\rho_{s_1} = \rho_{s_2} = 0.1$). However, if the underlying distributions have a stronger Spearman's correlation ($\rho_{s_1} = \rho_{s_2} = 0.5$), (Norm) tends to be too "liberal", especially for the Cuadras-Augé copula and the bivariate exponential distribution.

(2) The estimated sizes of (N) are all very close to the nominal significance level.

Method II

(3) If the underlying distributions/copulas have a weak Spearman's correlation, all estimated sizes of (N) are very close to the nominal significance level. These results hold for the data generated from a copula with the parameter specified by the null *and* alternative hypotheses

(see Table 8.17). However, if the underlying distributions/copulas have a stronger Spearman's correlation, the nonparametric test seems to be slightly "liberal" (see Table 8.18).

Conclusion: Based on the overall good performance with regard to estimated size, we recommend that the non-parametric test (N) be used for testing the hypothesis

$$H_0 : \rho_{s_1} = \rho_{s_2} \quad \text{vs.} \quad H_A : \rho_{s_1} > \rho_{s_2}.$$

8.3 Evaluation of the performance of the bootstrap critical value and the bootstrap estimate of power for the mean

In this section Monte-Carlo simulations will be used to determine some properties of the bootstrap critical value $C^R_{n,P}(\alpha; \mathbf{X}_n)$, such as the mean squared error (MSE) and bias. Recall that $C^R_{n,P}(\alpha; \mathbf{X}_n)$ is the bootstrap estimator of the critical value $C_n(\alpha)$ for testing (see Section 3.2.2(a))

$$H_0 : \mu = \mu_0 \quad \text{vs.} \quad H_A : \mu > \mu_0, \tag{8.4}$$

and is defined by

$$P^*_{H_0^*}\left(\frac{\sqrt{n}(\overline{X}^*_n - \overline{X}_n)}{S_n(\mathbf{X}^*_n)} \geq C^R_{n,P}(\alpha; \mathbf{X}_n)\right) \cong \alpha.$$

We will also illustrate some properties of the bootstrap estimator of power (see expression (3.7)) for the test in (8.4), at a specific alternative μ_A.

First, we introduce some new notation and discuss some Monte-Carlo approximations:

Let $\widetilde{C}_n(\alpha)$ denote the Monte-Carlo approximation of $C_n(\alpha)$. It is then possible to calculate $\widetilde{C}_n(\alpha)$ as follows:

1) Obtain $\mathbf{X}_n = (X_1, X_2, \ldots, X_n)$ from a distribution with parameter μ_0.

2) Calculate $T_{n,P}(\mathbf{X}_n)$. Denote this by T_1. Recall that $T_{n,P}(\mathbf{X}_n) = \frac{\sqrt{n}(\overline{X}_n - \mu_0)}{S_n(\mathbf{X}_n)}$.

3) Independently repeat steps 1) and 2) 1 000 000 times to obtain $T_1, T_2, \ldots, T_{1\ 000\ 000}$.

4) Obtain the order statistics: $T_{(1)} \leq T_{(2)} \leq \cdots \leq T_{(1\ 000\ 000)}$.

5) Set $\widetilde{C}_n(\alpha) = T_{(\lfloor 1\ 000\ 000(1-\alpha)\rfloor)}$.

The value of $MSE(C^R_{n,P}(\alpha; \mathbf{X}_n))$ can now be approximated by

$$\widehat{MSE} = \frac{1}{MC}\sum_{i=1}^{MC}(\hat{C}^R_{n,P}(\alpha; \mathbf{X}^i_n) - \widetilde{C}_n(\alpha))^2,$$

while the value of $Bias(C^R_{n,P}(\alpha;\mathbf{X}_n))$ can be approximated by

$$\widehat{Bias} = \frac{1}{MC}\sum_{i=1}^{MC}(\hat{C}^R_{n,P}(\alpha;\mathbf{X}^i_n)) - \widetilde{C}_n(\alpha).$$

The estimated variance of $C^R_{n,P}(\alpha;\mathbf{X}_n)$ is given by

$$\widehat{Var} = \widehat{MSE} - \widehat{Bias}^2.$$

Next, let P^A_{boot} denote the bootstrap estimator of the power for the test in (8.4), at a specific alternative μ_A. Let $\widetilde{P}^A_{MC}$ denote the Monte-Carlo approximation of P^A, the true power of the test at a specific alternative μ_A. We can calculate $\widetilde{P}^A_{MC}$ as follows:

1) Obtain $\mathbf{X}_n = (X_1, X_2, \ldots, X_n)$ from a distribution with parameter μ_A.

2) Calculate $T_{n,P}(\mathbf{X}_n)$. Denote this by T_1.

3) Independently repeat steps 1) and 2) 500 000 times to obtain $T_1, T_2, \ldots, T_{500\ 000}$.

4) $\widetilde{P}^A_{MC} = \frac{1}{MC}\sum_{i=1}^{500\ 000} I(T_i \geq \widetilde{C}_n(\alpha))$.

We can now approximate $MSE(P^A_{boot})$ and $Bias(P^A_{boot})$ by

$$\frac{1}{MC}\sum_{i=1}^{MC}(\widehat{P}^A_{boot}(i) - \widetilde{P}^A_{MC})^2 \quad \text{and} \quad \frac{1}{MC}\sum_{i=1}^{MC}\widehat{P}^A_{boot}(i) - \widetilde{P}^A_{MC}.$$

Data were generated from four distributions:

- Normal distribution with mean 0 and variance 1;
- Double exponential distribution with mean 0 and variance 2;
- Exponential distribution with mean 1 and variance 1;
- Chi-squared distribution with mean 3 and variance 6.

Sample sizes $n = 20, 30, 50$ and 100 were used and we chose $MC = 40\ 000$. All standard errors were found to be negligibly small. The bootstrap power estimates were averaged over the 40 000 data sets for which the power was evaluated at numerous alternatives to form bootstrap estimated power curves. It is important to note that the difference (at a specific alternative) between the bootstrap estimated power curve and the Monte-Carlo approximation of the true power curve, represents the estimated bias of P^A_{boot}. In Figures 8.1, 8.4, 8.7 and 8.10, the dotted lines represent the bootstrap estimated power curves, while the solid lines represent the Monte-Carlo approximation of the true power curves.

Normal distribution: $H_0 : \mu = 0$ vs. $H_A : \mu > 0$

Table 8.19: Estimated Bias, Variance and MSE of $C^R_{n,P}(\alpha; \mathbf{X}_n)$

n	$\widehat{Bias}$	$\widehat{Var}$	$\widehat{MSE}$
20	0.011	0.026	0.026
30	0.004	0.013	0.013
50	-0.002	0.006	0.007
100	-0.003	0.005	0.005

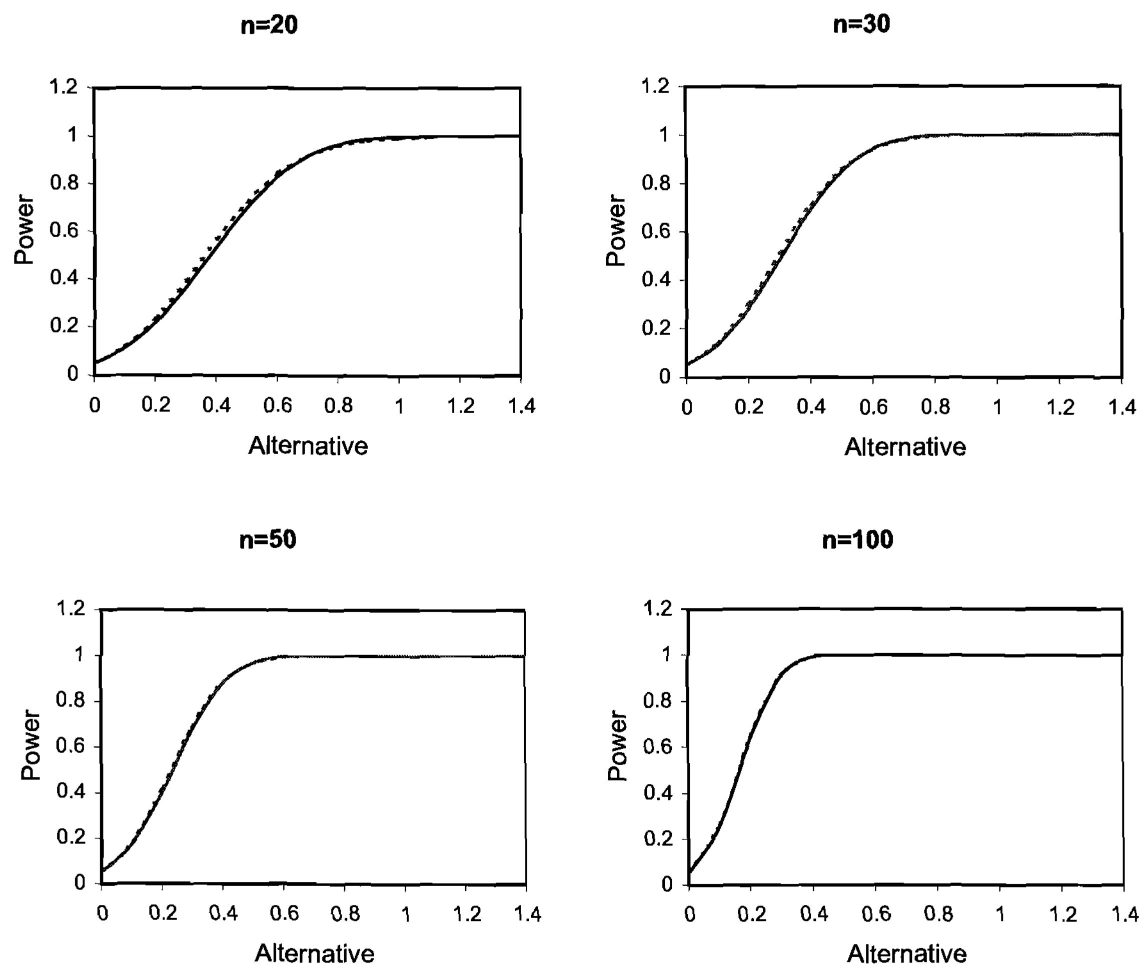

Figure 8.1: Bootstrap estimated power curves.

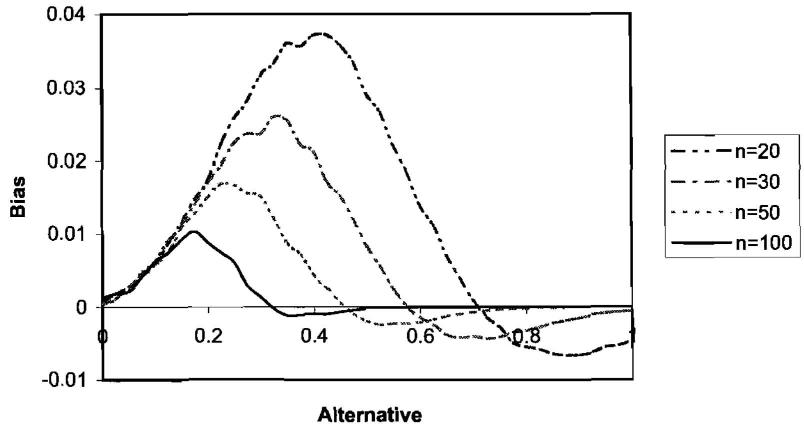

Figure 8.2: Estimated bias of P^A_{boot}

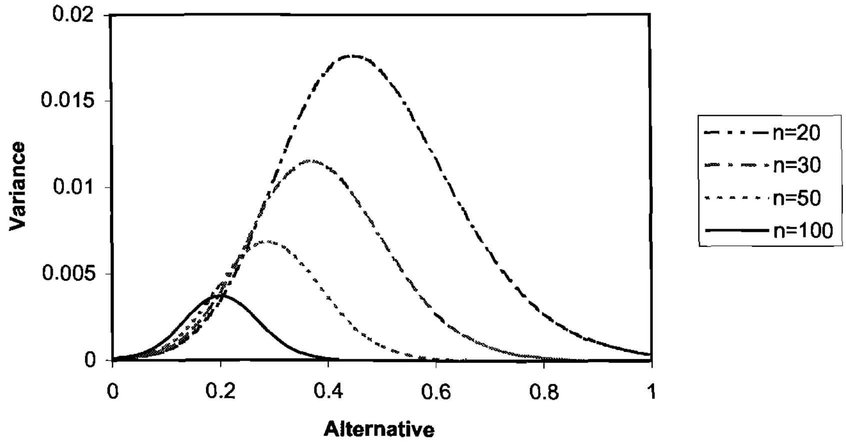

Figure 8.3: Estimated variance of P^A_{boot}

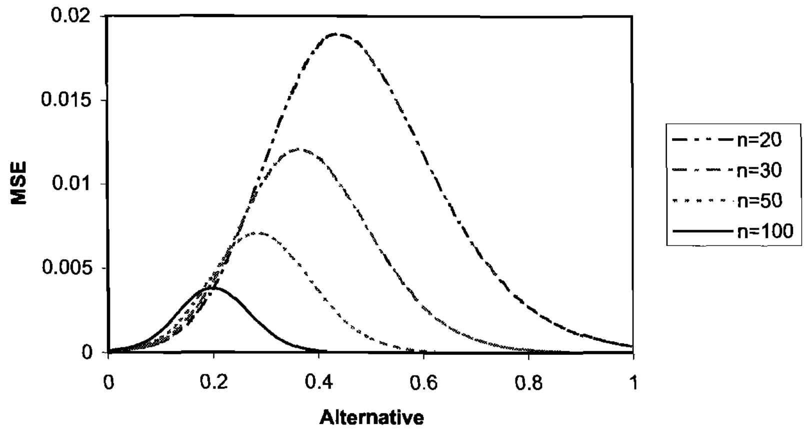

Figure 8.4: Estimated MSE of P^A_{boot}

Double exponential distribution: $H_0 : \mu = 0$ vs. $H_A : \mu > 0$

Table 8.20: Estimated Bias, Variance and MSE of $C^R_{n,P}(\alpha; \mathbf{X}_n)$

n	$\widehat{Bias}$	$\widehat{Var}$	$\widehat{MSE}$
20	0.062	0.096	0.100
30	0.033	0.046	0.047
50	-0.009	0.020	0.020
100	0.001	0.009	0.009

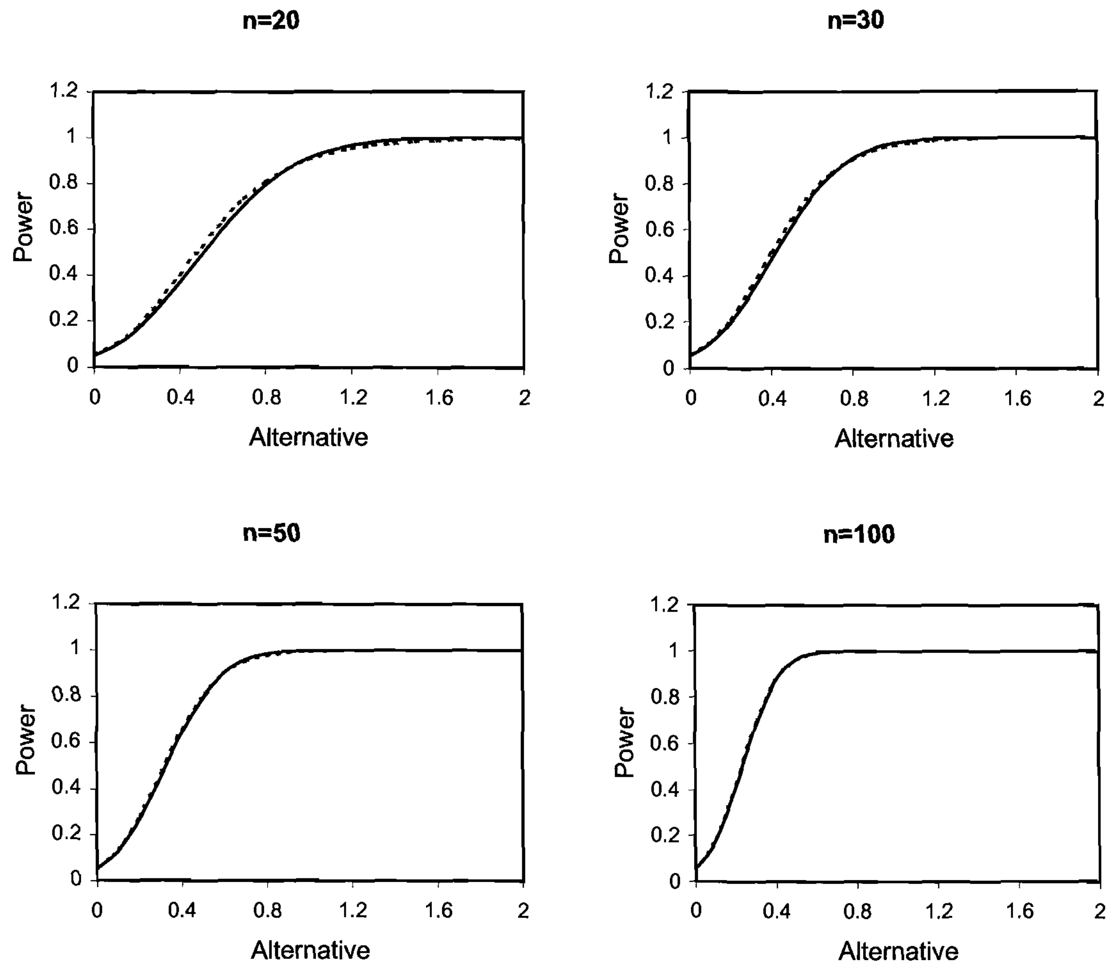

Figure 8.5: Bootstrap estimated power curves.

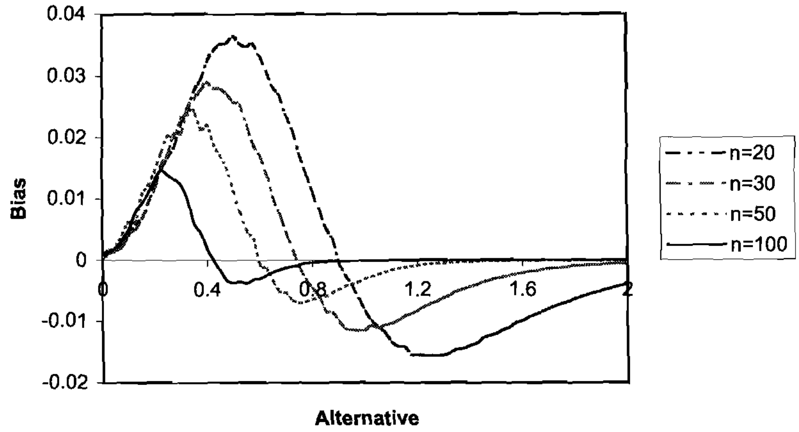

Figure 8.6: Estimated bias of P^A_{boot}

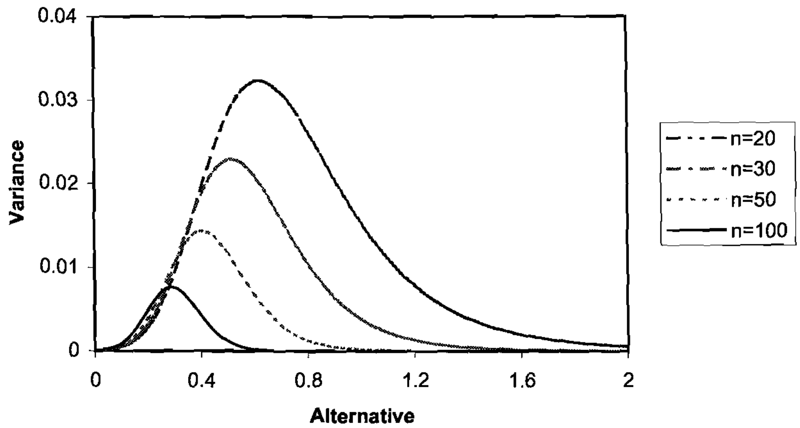

Figure 8.7: Estimated variance of P^A_{boot}

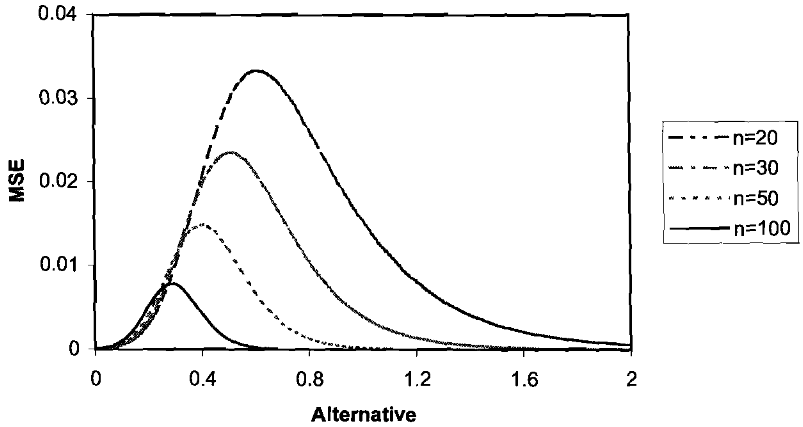

Figure 8.8: Estimated MSE of P^A_{boot}

Exponential distribution: $H_0 : \mu = 1$ vs. $H_A : \mu > 1$

Table 8.21: Estimated Bias, Variance and MSE of $C^R_{n,P}(\alpha; \mathbf{X}_n)$

n	$\widehat{Bias}$	$\widehat{Var}$	$\widehat{MSE}$
20	0.072	0.013	0.018
30	0.045	0.009	0.011
50	0.025	0.006	0.007
100	0.009	0.005	0.005

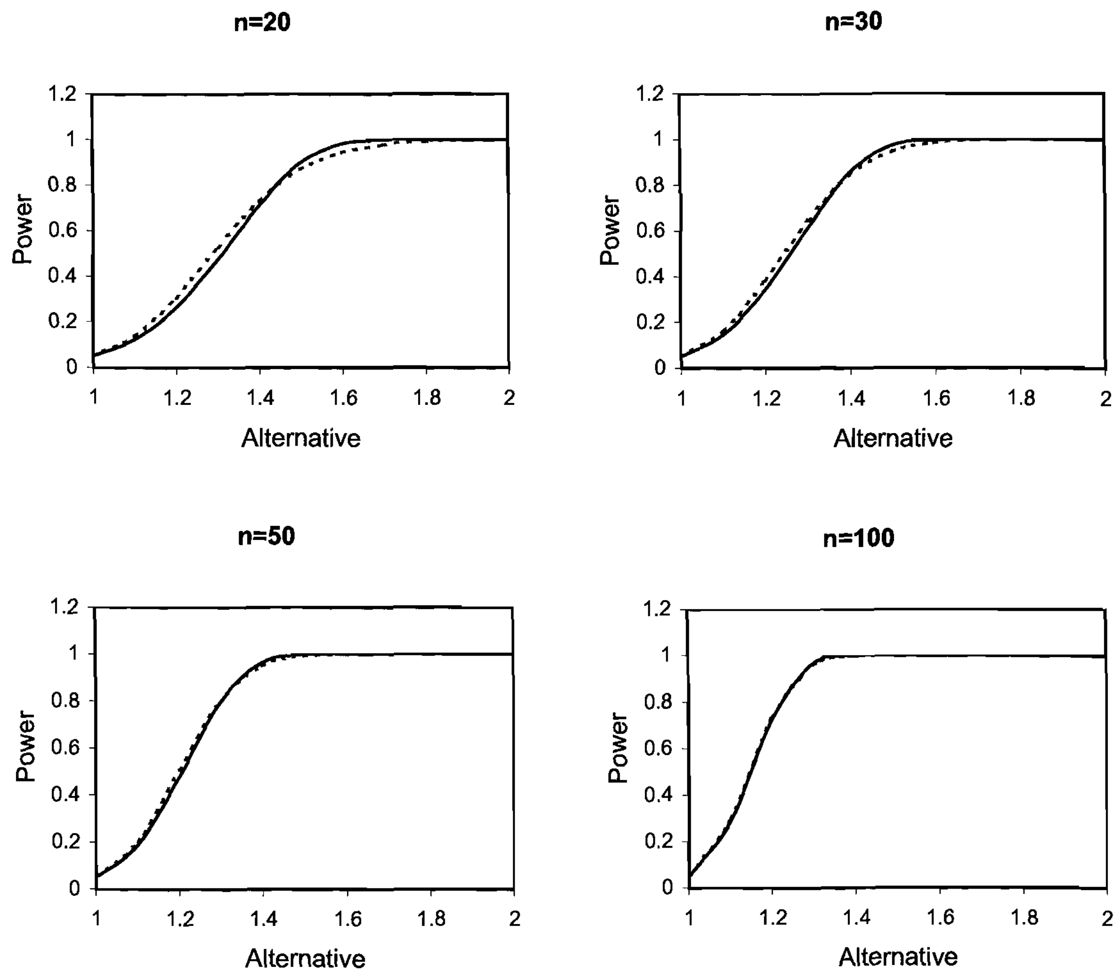

Figure 8.9: Bootstrap estimated power curves.

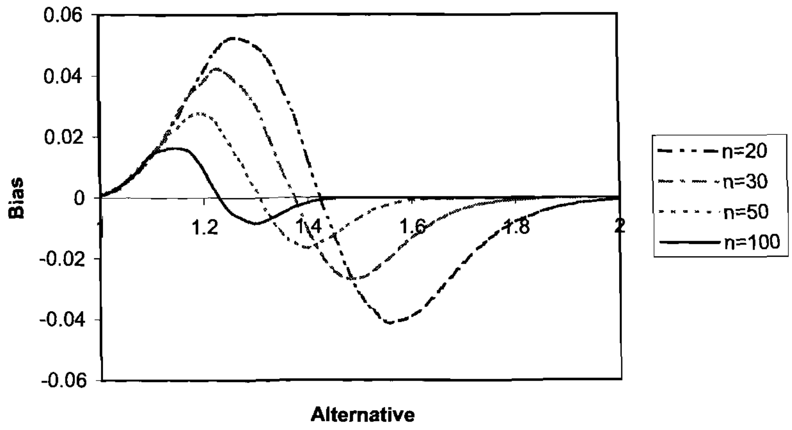

Figure 8.10: Estimated bias of P^A_{boot}

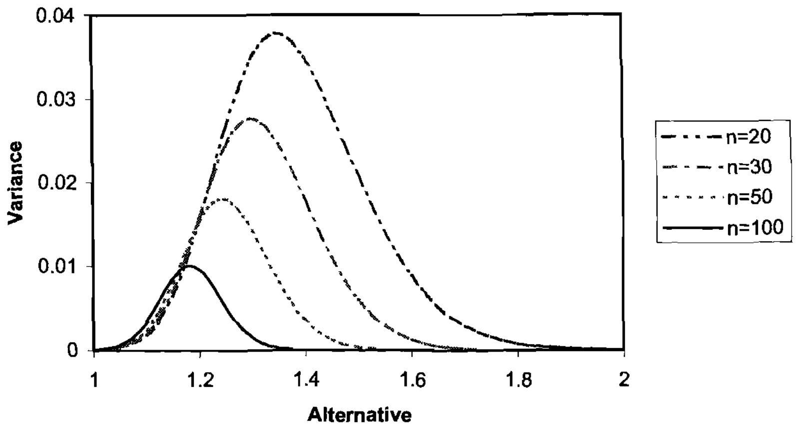

Figure 8.11: Estimated variance of P^A_{boot}

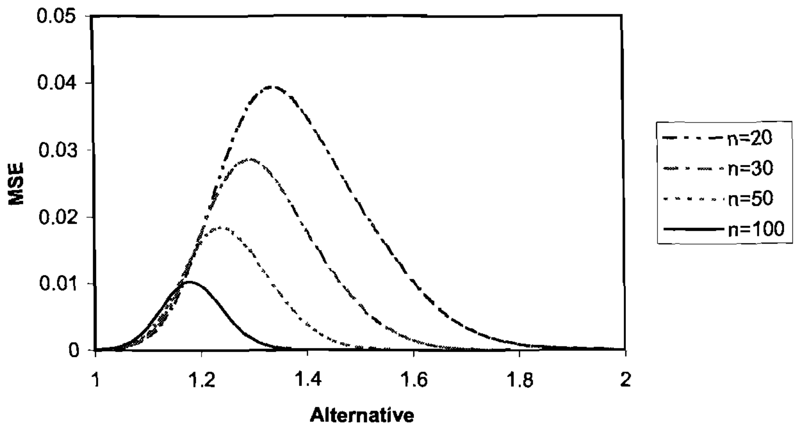

Figure 8.12: Estimated MSE of P^A_{boot}

Chi-squared distribution: $H_0 : \mu = 3$ vs. $H_A : \mu > 3$

Table 8.22: Estimated Bias, Variance and MSE of $C^R_{n,P}(\alpha; \mathbf{X}_n)$

n	$\widehat{Bias}$	$\widehat{Var}$	$\widehat{MSE}$
20	0.060	0.013	0.017
30	0.037	0.010	0.011
50	0.020	0.007	0.007
100	0.006	0.005	0.005

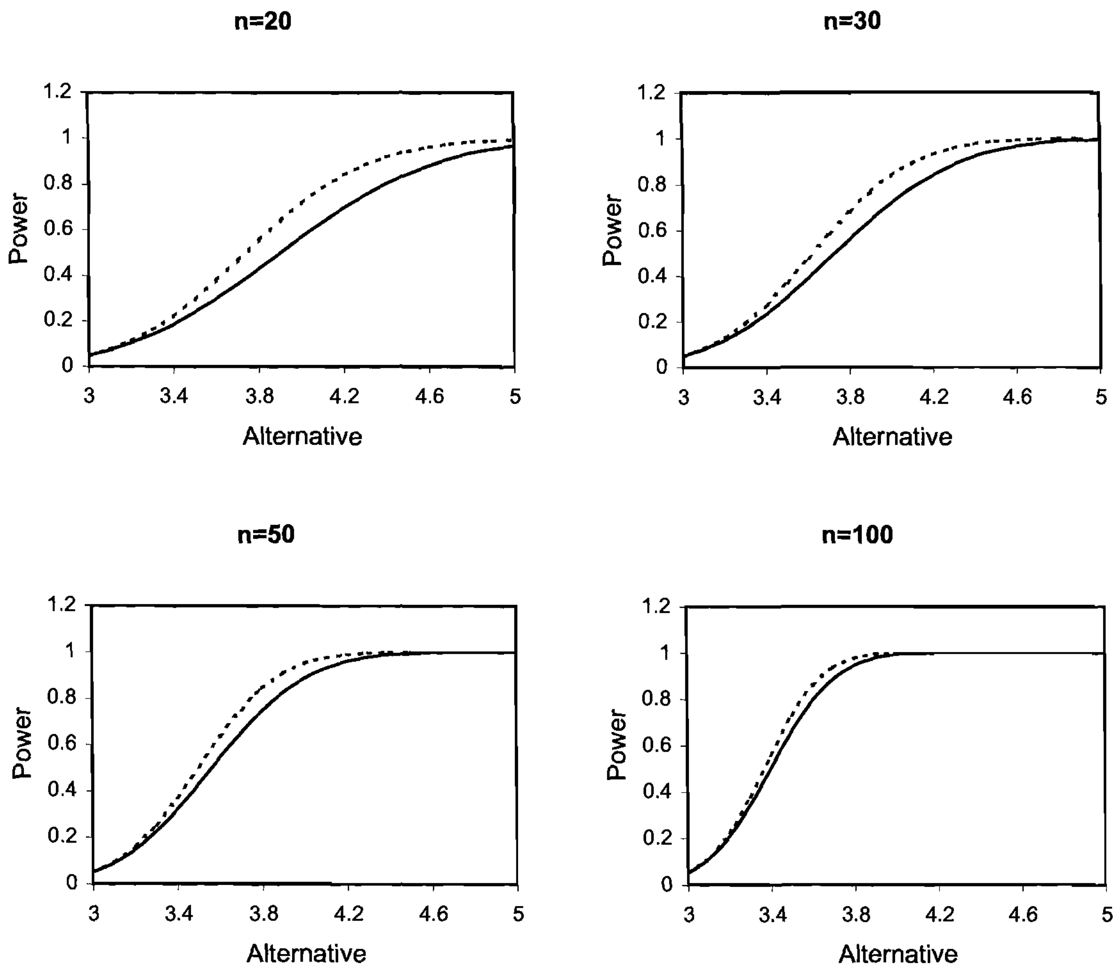

Figure 8.13: Bootstrap estimated power curves.

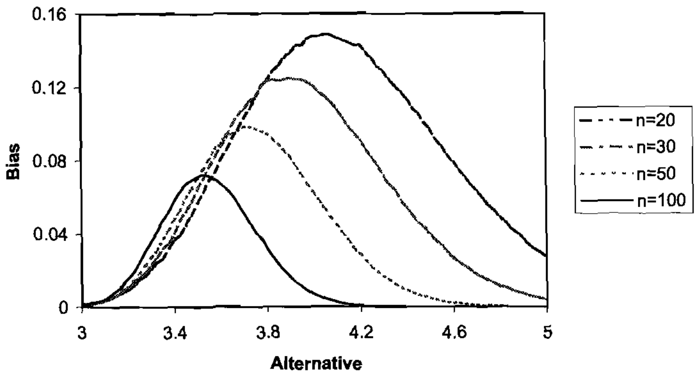

Figure 8.14: Estimated bias of P^{A}_{boot}

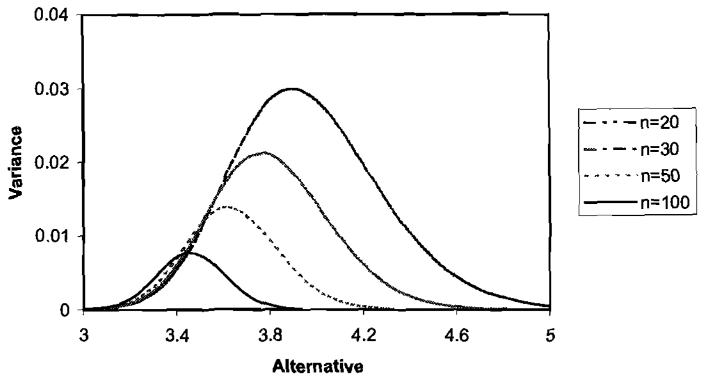

Figure 8.15: Estimated variance of P^{A}_{boot}

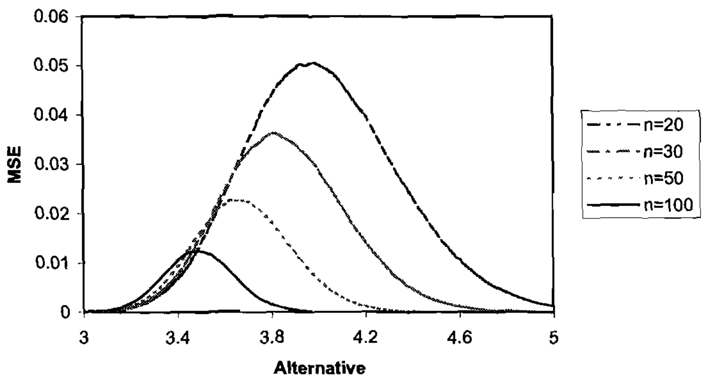

Figure 8.16: Estimated MSE of $\widehat{P}^{A}_{boot}$

From Figures 8.1 – 8.16 we can conclude the following:

(1) In most cases the bootstrap estimated power curves lie very close to the Monte-Carlo power curves.

(2) In the case where the data are generated from the chi-squared distribution, the bootstrap seems, on average, to overestimate the power slightly, even for large n. This can be seen from Figure 8.14

(3) When the data are generated from the double exponential and exponential distributions, the bootstrap overestimates the power (on average) up to a certain point, whereafter it slightly underestimates the power. This can be seen from Figures 8.6 and 8.10.

Note the interesting shapes formed by the estimated MSE, $Bias$ and Var graphs for the estimated powers. These shapes can possibly be explained by the fact that the estimated power values for alternatives which produce true power values close to 0.5 have a greater variation and bias than those estimated powers for alternatives which produce true powers close to 0 or 1.

Overall, the simulation results of this section present a positive case for using the bootstrap to estimate the critical value and the power, when testing for the mean in the univariate case. Recall that similar positive results were also obtained in the simulation studies of Chapter 4, where we investigated the performance of bootstrap critical values with regard to sizes of the relevant tests. Similar simulations can be done, e.g., when testing for the variance or Spearman's correlation coefficient.

8.4 Concluding remarks and future research

One of the reasons for the success of the bootstrap methodology can be attributed to the ease with which it can be applied in practice. As was shown previously, accurate and simple Monte-Carlo procedures can be used to obtain approximations for bootstrap estimates (see, e.g., Sections 2.2 and 3.2.1). However, in the case of bootstrap hypothesis testing, care should be taken to ensure that the correct procedures are followed (as was explained throughout this dissertation).

Future research

Suppose $X_1, X_2, \ldots, X_p$ are p random variables with variance-covariance matrix $\mathbf{\Sigma}$, i.e., $\mathbf{\Sigma}$ is a $(p \times p)$ matrix with elements $\sigma_{ij} = Cov(X_i, X_j)$ in the i^{th} row and j^{th} column, $i, j = 1, 2, \ldots, p$. Denote by $\mathbf{\Sigma}_0$ the $(p \times p)$ hypothesized matrix for the population variance-covariance matrix and write $\mathbf{X} = (X_1, X_2, \ldots, X_p)$. Consider decomposing the matrices $\mathbf{\Sigma}$ and $\mathbf{\Sigma}_0$ as follows:

$$\mathbf{\Sigma} = \mathbf{A}\mathbf{A}^T \quad \text{and}$$

$$\mathbf{\Sigma}_0 = \mathbf{B}\mathbf{B}^T,$$

where $\mathbf{A}$ and $\mathbf{B}$ are $(p \times p)$ matrices (one could, for example, make use of a Cholesky-decomposition to obtain these matrices).

If $\mathbf{\Sigma}^{-1}$ exists, then it is possible to transform $\mathbf{X}$ in such a way that it has the specified variance-covariance matrix $\mathbf{\Sigma}_0$. The following theorem describes this transformation.

Theorem 8.4.1 *If $\mathbf{\Sigma}^{-1}$ exists and $Cov(\mathbf{X}^T) = \mathbf{\Sigma}$, $\mathbf{\Sigma} = \mathbf{A}\mathbf{A}^T$ and $\mathbf{\Sigma}_0 = \mathbf{B}\mathbf{B}^T$, then $Cov(\mathbf{B}\mathbf{A}^{-1}\mathbf{X}^T) = \mathbf{\Sigma}_0$.*

Proof.

$$\begin{aligned}
Cov(\mathbf{B}\mathbf{A}^{-1}\mathbf{X}^T) &= \mathbf{B}\mathbf{A}^{-1}\mathbf{\Sigma}(\mathbf{B}\mathbf{A}^{-1})^T \\
&= \mathbf{B}\mathbf{A}^{-1}\mathbf{A}\mathbf{A}^T(\mathbf{A}^{-1})^T\mathbf{B}^T \\
&= \mathbf{B}\mathbf{B}^T \\
&= \mathbf{\Sigma}_0
\end{aligned}$$

Q.E.D.■

In Section 3.2.2 (e) a bootstrap test for testing the variance in the univariate case was discussed whereby the data were transformed to have a sample variance equal to σ_0^2. This transformation (given by expression (3.9)) is a special case of Theorem 8.4.1 with $p = 1$.

In Section 3.2.2 (g) a bootstrap test for testing the hypothesis

$$H_0 : \rho_p = \rho_0 \quad \text{vs.} \quad H_A : \rho_p > \rho_0,$$

was considered. Recall that $\mathbf{Z}_n = \big((X_1, Y_1), \ldots, (X_n, Y_n)\big)$ denoted a bivariate random sample from a population (X, Y) with unknown joint distribution function $H(x, y)$ and variance-covariance matrix

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_x^2 & \rho_p\sigma_x\sigma_y \\ \rho_p\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}.$$

In order for the data to have a sample Pearson's correlation equal to ρ_0, Martin (2007) proposed the transformation given by expressions (3.15) and (3.16). Another possible transformation can be obtained by using Theorem 8.4.1:

Let

$$\mathbf{S} = \begin{bmatrix} S_n^2(\mathbf{X}_n) & \hat{\rho}_p S_n(\mathbf{X}_n)S_n(\mathbf{Y}_n) \\ \hat{\rho}_p S_n(\mathbf{X}_n)S_n(\mathbf{Y}_n) & S_n^2(\mathbf{Y}_n) \end{bmatrix}$$

denote the sample variance-covariance matrix and let

$$\mathbf{\Sigma}_0 = \begin{bmatrix} S_n^2(\mathbf{X}_n) & \rho_0 S_n(\mathbf{X}_n)S_n(\mathbf{Y}_n) \\ \rho_0 S_n(\mathbf{X}_n)S_n(\mathbf{Y}_n) & S_n^2(\mathbf{Y}_n) \end{bmatrix},$$

where $\mathbf{X}_n = (X_1, X_2, \ldots, X_n)$ and $\mathbf{Y}_n = (Y_1, Y_2, \ldots, Y_n)$. Using similar arguments as before, the matrix $\mathbf{S}$ can be decomposed as follows:

$$\mathbf{S} = \mathbf{C}\mathbf{C}^T,$$

where $\mathbf{C}$ is a (2×2) matrix.

Assume that $\mathbf{S}^{-1}$ exists and consider the following transformation of $\{(X_i, Y_i), i = 1, 2, \ldots, n\}$:

$$\begin{bmatrix} V_i^{x0} \\ V_i^{y0} \end{bmatrix} = \mathbf{B}\mathbf{C}^{-1} \begin{bmatrix} X_i \\ Y_i \end{bmatrix}, \quad i = 1, 2, \ldots, n.$$

It now follows from Theorem 8.4.1 that the sample Pearson's correlation between $\{V_i^{x0},\ i = 1, 2, \ldots, n\}$ and $\{V_i^{y0},\ i = 1, 2, \ldots, n\}$ is equal to ρ_0. Resampling can now be done from the pairs $(V_1^{x0}, V_1^{y0}), (V_2^{x0}, V_2^{y0}), \ldots, (V_n^{x0}, V_n^{y0})$.

Future research will entail using Theorem 8.4.1 to transform multivariate data to have a specified variance-covariance structure. This will enable one to perform bootstrap-based tests for situations involving the variance-covariance matrix, for example, testing for an equal correlation structure in principle component analysis, hypotheses relating to factor analysis and canonical correlation analysis.

In addition to the above-mentioned research problems, another challenging problem would be to develop bootstrap-based tests for other measures of association (such as Kendall's tau), using similar arguments as those used to develop the test for Spearman's rho in Chapter 7.

Bibliography

Ali, M. M., Mikhail, N. N. and Haq, M. S. (1978). A class of bivariate distributions including the bivariate logistics, *Journal of Multivariate Analysis* **8**: 405–412.

Allen, D. L. (1997). Hypothesis testing using an L1-distance bootstrap, *The American Statistician* **51**: 145–150.

Anderson, N. H., Hall, P. and Titterington, D. M. (1994). Two sample test statistics for measuring discrepancies between two multivariate probability density functions using kernel-based density estimates, *Journal of Multivariate Analysis* **50**: 41–54.

Anderson, T. W. (1984). *An Introduction to Multivariate Analysis*, second edn, John Wiley & Sons, New York.

Athreya, K. B. (1987). Bootstrap of the mean in the infinite variance case, *Annals of Statistics* **15**: 724–731.

Athreya, K. B. and Fukuchi, J. (1997). Confidence intervals for endpoints of a c.d.f. via bootstrap, *Journal of Statistical Planning and Inference* **58**: 299–320.

Athreya, K. B., Fukuchi, J. and Lahiri, S. N. (1999). On the bootstrap and the moving block bootstrap for the maximum of a stationary process, *Journal of Statistical Planning and Inference* **76**: 1–18.

Basawa, I. V., Mallik, A. K., McCormick, W. P., Reeves, J. H. and Taylor, R. L. (1991). Bootstrapping unstable first-order autoregressive processes, *Annals of Statistics* **19**: 1098–1101.

Bühlmann, P. (1995). The blockwise bootstrap for general empirical processes of stationary sequences., *Stochastic Processes and their applications* **58**: 247–265.

Bühlmann, P. (1997). Sieve bootstrap for time series, *Bernoulli* **3**: 123–148.

Bühlmann, P. (1998). Sieve bootstrap for smoothing in nonstationary time series., *Annals of Statistics* **26**: 48–83.

Bühlmann, P. (2002). Bootstraps for time series, *Statistical Science* **17**: 52–72.

Bühlmann, P. and Künsch, H. R. (1995). The blockwise bootstrap for general parameters of a stationary time series, *Scandinavian Journal of Statistics* **22**: 35–54.

Bühlmann, P. and Künsch, H. R. (1999). Block length selection in the bootstrap for time series, *Computational Statistics and Data Analysis* **31**: 295–310.

Beran, R. (1987). Prepivoting to reduce level error of confidence sets, *Biometrika* **74**: 457–468.

Beran, R. (1988). Prepivoting test statistics: A bootstrap view of asymptotic refinements, *Journal of the American Statistical Association* **83**(403): 687–697.

Bickel, P. J. and Bühlmann, P. (1999). A new mixing notion and functional central limit theorems for a sieve bootstrap in time series, *Bernoulli* **5**: 413–446.

Bickel, P. J. and Freedman, D. A. (1981). Some asymptotic theory for the bootstrap, *Annals of Statistics* **9**: 1196–1217.

Bickel, P. J. and Ren, J. J. (1995). The m out of n bootstrap and goodness of fit tests with double censored data, *in* H. Rieder (ed.), *Robust Statistics, Data Analysis and Computer Intensive methods*, Springer, New York.

Billingsley, P. (1968). *Convergence of Probability Measures*, John Wiley & Sons, New York.

Boos, D. D. and Brownie, C. (1989). Bootstrap methods for testing homogeneity of variances, *Technometrics* **31**: 69–82.

Boos, D. D. and Brownie, C. (2004). Comparing variances and other measures of dispersion, *Statistical Science* **19**: 571–578.

Boos, D., Janssen, P. and Veraverbeke, N. (1989). Resampling from centered data in the two-sample problem, *Journal of Statistical Planning and Inference* **21**: 327–345.

Booth, J. G. and Hall, P. (1994). Monte Carlo approximations and the iterated bootstrap, *Biometrika* **81**: 331–340.

Booth, J. G. and Sarkar, S. (1998). Monte Carlo approximation of bootstrap variances, *American Statistician* **52**: 354–357.

Burden, R. L. and Faires, J. D. (1989). *Numerical Analysis*, 4th edn, PWS-KENT Publishing Company, Boston.

Cao, R. and van Keilegom, I. (2006). Empirical likelihood tests for two-sample problems via nonparametric density estimation, *Canadian Journal of Statistics* **34**: 61–77.

Carpenter, J. and Bithell, J. (2000). Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians, *Statistics in Medicine* **19**: 1141–1164.

Chapman, P. and Hinkley, D. V. (1986). The double bootstrap, pivots and confidence limits, *Technical Report 34*, Center for Statistical Sciences, University of Texas at Austin.

Choi, E. and Hall, P. (2000). Bootstrap confidence regions computed from autoregressions of arbitrary order, *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* **62**: 461–477.

Chung, K. and Lee, S. M. (2001). Optimal bootstrap sample size in construction of percentile confidence bounds, *Scandinavian Journal of Statistics* **28**: 225–239.

Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence, *Biometrika* **65**: 141–151.

Cramér, H. (1945). *Mathematical Methods of Statistics*, Princeton University Press, Princeton.

Cuadras, C. M. and Augé, J. (1981). A continuous general multivariate distribution and its properties, *Communications in Statistics A- Theory and Methods* **10**: 339–353.

Datta, S. (1996). On asymptotic properties of bootstrap AR(1) processes, *Journal of Statistical Planning and Inference* **53**: 361–374.

Davidson, R. and Flachaire, E. (2001). The wild bootstrap, tamed at last., *GREQAM. Document de Travail 99A32, revised.*

Davidson, R. and MacKinnon, J. G. (1992). A new form of the information matrix tests, *Econometrica* **60**: 145–157.

Davidson, R. and MacKinnon, J. G. (1999). Bootstrap testing in nonlinear models, *International Economic Review* **40**: 487–508.

Davidson, R. and MacKinnon, J. G. (2004). *Econometric Theory and Methods*, Oxford University Press, New York.

Davidson, R. and MacKinnon, J. G. (2007). Improving the reliability of bootstrap tests with the fast double bootstrap, *Computational Statistics and Data Analysis* **51**: 3259–3281.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap methods and their application*, Cambridge University Press, Cambridge.

Davison, A. C., Hinkley, D. V. and Worton, B. J. (1992). Bootstrap likelihoods, *Biometrika* **79**: 113–130.

de Angelis, D. and Young, G. A. (1992). Smoothing the bootstrap, *Internat. Statist. Rev* **60**: 45–56.

de Beer, C. F. and Swanepoel, J. W. H. (1989). A modified Durbin-Watson test for serial correlation in multiple regression under nonnormality using the bootstrap, *Journal of Statistical Computation and Simulation* **33**: 75–81.

de Jongh, D. N., Nankervis, J. C., Savin, N. E. and Whiteman, C. H. (1992). Integration versus trend stationary in time series, *Econometrica* **60**: 423–433.

Deheuvels, P., Mason, D. M. and Shorack, G. R. (1993). Some results on the influence of extremes on the bootstrap, *Ann. Inst. H. Poincare (Probab. Statist)* **29**: 83–103.

Devroye, L. (1986). *Non-Uniform Random Variate Generation*, Springer, New York.

DiCiccio, T. J. and Efron, B. (1996). Bootstrap confidence intervals (with discussion), *Statistical Science* **11**: 189–228.

Dufour, J. M. and Khalaf, L. (2002). Simulation based finite and large sample tests in multivariate regressions, *Journal of Econometrics* **111**: 303–322.

Efron, B. (1979). Bootstrap methods: another look at the jackknife, *Annals of Statistics* **7**: 1–26.

Efron, B. (1981). Nonparametric standard errors and confidence intervals, *The Canadian Journal of Statistics* **9**(2): 139–172.

Efron, B. and Tibshirani, R. (1985). The bootstrap method for assessing statistical accuracy, *Behaviormetrika* **17**: 1–35.

Efron, B. and Tibshirani, R. J. (1993). *An introduction to the bootstrap*, Chapman and Hall, New York.

Farlie, D. J. G. (1960). The performance of some correlation coefficients for a general bivariate distribution, *Biometrika* **47**: 307–323.

Fisher, N. I. (1997). Copulas, *in* S. Kotz, C. B. Read and D. L. Banks (eds), *Encyclopedia of Statistical Sciences*, Wiley, New York, pp. 159–163.

Fisher, N. I. and Hall, P. (1990). On bootstrap hypothesis testing, *The Australian Journal of Statistics* **32**: 177–190.

Fisher, R. A. (1921). On the Probable error of a coefficient of correlation deduced from a small sample, *Metron* **1**: 3–32.

Fisher, R. A. (1970). *Statistical methods for research workers*, fourteenth edn, Hafner Publishing Company.

Fishman, G. S. (1996). *Monte Carlo: concepts, algorithms and applications*, Springer-Verlag, New York.

Fredricks, G. A. and Nelsen, R. B. (2007). On the relationship between spearman's rho and kendall's tau for pairs of continuous random variables, *Journal of Statistical Planning and Inference* **137**: 2143–2150.

Götze, F. and Künsch, H. R. (1996). Second-order correctness of the blockwise bootstrap for stationary observations, *Annals of Statistics* **24**: 1914–1933.

Götze, F. and Rackauskas, A. (1999). Adaptive choice of bootstrap sample size, *A Festschrift for W.R. van Zwet*, I.M.S publications.

Good, P. (2000). *Permutation tests: A practical guide to resampling methods for testing hypothesis*, Springer, New York.

Hall, P. (1988). Theoretical comparison of bootstrap confidence intervals, *Annals of Statistics* **50**: 35–45.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*, Springer, New York.

Hall, P. and DiCiccio, T. J. (1989). On smoothing and the bootstrap, *Annals of Statistics* **17**: 692–704.

Hall, P. and LePage, R. (1996). On bootstrap estimation of the distribution of the studentized mean, *Ann. Inst. Statist. Math.* **48**: 403–421.

Hall, P. and Martin, M. A. (1988). On the bootstrap and two-sample problems, *The Australian Journal of Satistics* **30A**: 179–192.

Hall, P. and Martin, M. A. (1989). A note on the accuracy of bootstrap percentile method confidence intervals for a quantile, *Statistics and Probability Letters* **8**: 197–200.

Hall, P. and Titterington, D. M. (1989). The effect of simulation order on level accuracy and power of monte-carlo tests, *Journal of the Royal Statistical Society, Series B* **51**: 459–467.

Hall, P. and Wilson, S. R. (1991). Two guidelines for bootstrap hypothesis testing, *Biometrics* **47**: 179–192.

Hall, P., Horowitz, J. L. and Jing, B. Y. (1995). On blocking rules for the bootstrap with dependent data, *Biometrika* **82**: 561–574.

Hall, P., Martin, M. A. and Schucany, W. R. (1988). Better nonparametric bootstrap confidence intervals for the correlation coefficient, *Journal of Statistical Computation and Simulation* **33**: 161–172.

Heiman, G. and Kreiss, J. P. (1996). Bootstrapping general first order autoregression, *Statistics and Probability Letters* **30**: 87–98.

Helmers, R. and Wegkamp, M. (1998). Wild bootstrapping in finite populations with auxiliary information, *Scandinavian Journal of Statistics* **25**: 383–399.

Horowitz, J. L. (2001). The bootstrap and hypothesis testing in econometrics, *Journal of Econometrics* **100**: 37–40.

Huang, J. S., Sen, P. K. and Shao, J. (1996). Bootstrapping a sample quantile when the density has a jump, *Statistica Sinica* **6**: 299–309.

Hurvich, C. M. and Tsai, C. (1989). Regression and time series model selection in small samples, *Biometrika* **76**: 297–307.

Hutchinson, T. P. and Lai, C. D. (1990). *Continuous Bivariate Distributions, Emphasising Applications*, Rumsby Scientific Publishing, Adelaide.

James, G. S. (1951). The comparison of several groups of observations when the ratios of the population variances are unknown, *Biometrika* **38**: 324–329.

Janssen, P. (1997). Bootstrapping U-statistics, *South African Statistical Journal* **31**: 185–216.

Janssen, P., Swanepoel, J. W. H. and Veraverbeke, N. (2001). Efficiency of linear regression estimators based on presmoothing, *Communications in Statistics – Theory and Methods* **30**: 2079–2097.

Janssen, P., Swanepoel, J. W. H. and Veraverbeke, N. (2002). The modified bootstrap error process for Kaplan-Meier quantiles, *Statistics and Probability Letters* **58**: 31–39.

Jimenez-Gamero, M. D., Munoz-Garca, J. and Pino-Mejas, R. (2003). Bootstrapping parameter estimated degenerate U and V statistics, *Statistics and Probability Letters* **61**: 61–70.

Jing, B. Y. (1995). Two-sample empirical likelihood method, *Statistics and Probability Letters* **24**: 315–319.

Joe, H. (1997). *Multivariate Models and Dependence*, Chapman & Hall, London.

Johnson, M. E. (1986). *Multivariate Statistical Simulation*, Wiley, New York.

Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations, *Annals of Statistics* **17**: 1217–1241.

Kendall, M. G. (1938). A new measure of rank correlation, *Biometrika* **30**: 81–93.

Kendall, M. G. (1970). *Rank correlation methods*, 4 edn, Griffin, London.

Knight, K. (1989). On the bootstrap of the sample mean in the infinite variance case, *Annals of Statistics* **17**: 1168–1175.

Kruskall, W. H. (1958). Ordinal measures of association, *Journal of the American Statistical Association* **53**: 814–861.

Lahiri, S. N. (1996). On Edgeworth expansion and moving block bootstrap for studentized M-estimators in multiple linear regression models, *Journal of Multivariate Analysis* **56**: 42–59.

Lee, S. M. S. (1999). On a class of m out of n bootstrap confidence intervals, *Journal of the Royal Statistical Society, series B* **61**: 901–911.

Lehman, E. L. (1966). Some concepts of dependence, *Annals of Mathematical Statistics* **37**: 1137–1153.

Lehmann, E. L. (1975). *Nonparametrics: Statistical Methods Based on Ranks*, Holden-Day, San Francisko.

Liu, R. Y. (1988). Bootstrap procedures under some non-i.i.d. models, *Annals of Statistics* **16**: 1696–1708.

Loh, W. Y. (1987). Calibrating confidence coefficient, *Journal of the American Statistical Association* **82**: 155–162.

MacKinnon, J. G. (2002). Bootstrap inference in econometrics, *Canadian Journal of Economics* **35**: 615–645.

Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models, *Annals of Statistics* **21**: 255–285.

Martin, M. A. (2007). Bootstrap hypothesis testing for some common statistical problems: A critical evaluation of size and power properties, *Computational Statistics and Data Analysis* **51**: 6321–6342.

Nankervis, J. C. and Savin, N. E. (1996). The level and power of the bootstrap t test in the AR(1) model with trend, *Journal of Business and Economic Statistics* **14**: 161–168.

Nelsen, R. B. (2006). *An Introduction to Copulas*, 2nd edn, Springer, New York.

Neter, J., Kutner, M. H., Nachtsheim, C. J. and Wasserman, W. (1996). *Applied Linear Statistical Models*, fourth edn, McGraw-Hill, New York.

Noether, G. E. (1967). *Elements of Nonparametric Statistics*, John Wiley & Sons, New York.

Noreen, E. W. (1989). *Computer Intensive Methods for testing Hypothesis: An introduction*, Wiley, New York.

Paparoditis, E. and Politis, D. N. (2005). Bootstrap hypothesis testing in regression models, *Statistics and Probability Letters* **74**: 356–365.

Park, Y. J. (2003). Boostrap unit root tests, *Econometrica* **71**: 1845–1895.

Plackett, R. L. (1965). A class of bivariate distributions, *Journal of the American Statistical Association* **60**: 516–522.

Polansky, A. M. (2001). Bandwidth selection for the smoothed bootstrap percentile method, *Computational Statistics and Data Analysis* **36**: 339–349.

Raftery, A. E. (1984). A continuous multivariate exponential distribution, *Communications in Statistics - Theory and Methods* **13**: 947–965.

Rao, C. R. (1989). Statistics and truth, putting chance to work, International Co-operative Publishing House, Burtonsville.

Rayner, R. K. (1990). Bootstrapping p values and power in the first-order autoregression: A Monte-Carlo investigation, *Journal of Business and Economic Statistics* **8**: 251–263.

Romano, J. P. (1988). A bootstrap revival of some nonparametric distance tests, *Journal of the American Statistical Association* **83**: 698–708.

Sakov, A. (1998). *Using the m out of n bootstrap in hypothesis testing*, PhD thesis, University of California, Berkeley.

Sakov, A. and Bickel, P. J. (2000). An edgeworth expansion for the m out of n bootstrapped median, *Statistics and Probability Letters* **49**: 217–223.

Schenker, N. (1985). Qualms about bootstrap confidence intervals, *Journal of the American Statistical Association* **80**: 360–361.

Schweizer, B. and Sklar, A. (1983). *Probabilistic Metric Spaces*, North-Holland, New York.

Schweizer, B. and Wolff, E. F. (1981). On nonparametric measures of dependence for random variables, *Annals of Statistics* **9**: 879–885.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*, John Wiley & Sons, New York.

Shao, J. and Tu, D. (1995). *The Jackknife and Bootstrap*, Springer, New York.

Shibata, R. (1980). An optimal autoregressive spectral estimate, *Annals of Statistics* **9**: 300–306.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges, *Publ Inst Statist Univ Paris* **8**: 229–231.

Spearman, C. (1904). The proof and measurement of association between two things, *American Journal of Psychology* **15**: 72–101.

Sriram, T. N. (1994). Invalidity of bootstrap for critical branching processes with immigration, *Annals of Statistics* **22**: 1013–1023.

Stern, S. (1997). Simulation-based estimaton, *Journal of Economic Literature* **35**: 2006–2039.

Stewart, K. G. (1997). Exact testing in multivariate regression, *Econometric Reviews* **16**: 321–352.

Swanepoel, J. W. H. (1986). A note on proving that the (modified) bootstrap works, *Communications in Statistics – Theory and methods* **15**: 3193–3203.

Swanepoel, J. W. H. (1990). A review of bootstrap methods, *South African Statistical Journal* **24**: 1–34.

Swanepoel, J. W. H. and van Wyk, J. W. J. (1986). The bootstrap applied to power spectral density function estimation, *Biometrika* **73**: 135–141.

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

van Giesbergen, N. P. A. and Kiviet, J. F. (2002). How to implement the bootstrap in static or stable dynamic regression models: Test statistic versus confidence interval approach, *Journal of Econometrics* **108**: 133–156.

Veraverbeke, N. (1997). Bootstrapping in survival analysis, *South African Statistical Journal* **31**: 217–258.

Westfall, P. H. and Young, S. S. (1993). *Resampling-based Multiple Testing: Examples and Methods for p-value Adjustments.*, Wiley, New York.

Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis, *Annals of Statistics* **14**: 1261–1295.

Young, A. (1986). Conditional data-based simulations: Some examples from geometrical statistics., *International Statistics Review* **54**: 1–13.

Young, A. (1988). Resampling tests of statistical hypothesis, *Proceeding of the Eighth Biannual Symposium on Computational Statistics*, Physica-verlag, Heidelberg.

Zarepour, M. and Knight, K. (1999). Bootstrapping unstable first order autoregressive process with errors in the domain of attraction of stable law, *Stochastic Models* **15**: 11–27.