# Contributions to the $m$-out-of-$n$ bootstrap

L. Santana, M.Sc.

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

**Promotor: Prof. J. W. H. Swanepoel**

December 2009
Potchefstroom

# Acknowledgements

I would like to thank the following individuals and acknowledge their contributions to my studies.

- ⋆ Prof. J.W.H. Swanepoel, for his support, guidance, and indefatigable enthusiasm.
- ⋆ J.S. Allison, G. Koekemoer, W.D. Schutte, S. Jansen van Vuuren and L. Boshoff, for their help with many small (and not so small) problems. Also, they have my gratitude for their willingness to lend an ear (sometimes both) to my seemingly endless list of complaints and crack-pot notions.
- ⋆ Prof. C.J. Swanepoel, for her help with the translation of the summary and her words of advice.
- ⋆ The computers. I couldn't have done it without you guys.
- ⋆ My parents, brothers, and my entire family, for their unwavering support and love during my studies.
- ⋆ To God, for allowing me the opportunity and privilege to study.

# Summary

The traditional bootstrap is a sophisticated resampling procedure that has received a great deal of attention in the literature over the past three decades. This thesis focuses on a variation of the traditional bootstrap, called the $m$-out-of-$n$ bootstrap, where one resamples fewer than $n$ observations from a sample of size $n$. It is shown, by referring to various sources in the literature, that this modification of the traditional bootstrap has many desirable properties, not least of which is that it rectifies certain inconsistencies suffered by the traditional bootstrap.

The aim of this thesis is twofold: First, to explore the collection of applications of the $m$-out-of-$n$ bootstrap and examine their usefulness, and second, to contribute to this collection by developing new applications, providing the necessary tools to apply them correctly, and to obtain estimators for the resample size, $m$.

The first few chapters of this thesis are a literature study which examines the development of the theory underlying the traditional bootstrap and the $m$-out-of-$n$ bootstrap as well as considering the practical applications of these two techniques. Included is a discussion of situations where the traditional bootstrap method fails to produce consistent results, but where the $m$-out-of-$n$ bootstrap is consistent under minimal conditions.

Once the basic theoretical background of the $m$-out-of-$n$ bootstrap has been established, a new methodology for applying the $m$-out-of-$n$ bootstrap in point estimation problems is presented. A contrast is made between the naive application of the $m$-out-of-$n$ bootstrap and the new methodology by referring to the new method as the 'corrected $m$-out-of-$n$ ' bootstrap.

The use of the $m$-out-of-$n$ bootstrap is considered for two new areas of application:

- ⋆ First, a new method for point estimation of parameters based on BRAGGing (Bootstrap Robust AGGregating) estimation methods is proposed using both the original naive $m$-out-of-$n$ bootstrap methodology, as well as the newer, corrected $m$-out-of-$n$ bootstrap methodology. The estimation of the resample size for this estimation problem is also addressed by considering Cornish-Fisher and other expansions.

- ⋆ Second, the application of the $m$-out-of-$n$ bootstrap to hypothesis testing is considered. Two new data-based choices of the resample size, $m$, are proposed in this setup. The first estimator is based on a bootstrap estimate of the size of the test when using a bootstrap critical value, and the second is based on the probability structure of the $p$-values of a test under the null hypothesis.

In both of these new areas of application, the data-dependent choices of $m$ are theoretically and numerically motivated, the former being accomplished through the use of comprehensive mathematical arguments and the latter through the use of extensive Monte-Carlo simulations.

# Uittreksel


Die tradisionele $n$-uit-$n$ skoenlusprosedure is 'n gesofistikeerde hersteekproefnemingsmetode wat ruim aandag gekry het in die literatuur die afgelope drie dekades. Hierdie verhandeling fokus op 'n variasie van die tradisionele metode, naamlik die $m$-uit-$n$ skoenlus, waar daar minder as $n$ waarnemings geneem word uit 'n steekproef van grootte $n$. Deur na verskillende bronne uit die literatuur te verwys, word aangetoon dat hierdie wysiging van die tradisionele skoenlus talle gewenste eienskappe besit, onder meer dat dit sekere nie-konsekwenthede eie aan die tradisionele skoenlus, herstel.

Die doel van die verhandeling is tweevoudig: Eerstens, om die versameling toepassings van die $m$-uit-$n$ skoenlus te vind, te ondersoek en die nut daarvan te bepaal, en tweedens, om by te dra tot die versameling deur nuwe toepassings te ontwikkel, die nodige tegnieke en gereedskap daar te stel en dit korrek toe te pas om sodoende beramers vir die hersteekproefgroottes $m$ te bepaal.

Die eerste paar hoofstukke van die verhandeling is 'n literatuurstudie waarin die ontwikkeling van die tradisionele skoenlus en die $m$-uit-$n$ skoenlus bestudeer word, asook die praktiese uitvoerbaarheid van die twee tegnieke. Ingesluit is 'n bespreking van situasies waar die tradisionele metode faal in die daarstelling van konsekwente resultate, maar waar die $m$-uit-$n$ skoenlus wel konsekwente resultate behaal onder minimale voorwaardes.

Sodra die teorietiese agtergrond vir die $m$-uit-$n$ skoenlus gevestig is, word 'n nuwe metodologie daargestel om die $m$-uit-$n$ skoenlus in puntberamingsprobleme toe te pas. Kontraste word uitgewys tussen die naïewe toepassing van die $m$-uit-$n$ skoenlus en die nuwe metodologie, deur na die nuwe metode te verwys as die "gekorrigeerde $m$-uit-$n$ skoenlus".

Die $m$-uit-$n$ skoenlus word aangewend op twee nuwe toepassingsgebiede:

⋆ Eerstens word 'n nuwe metode vir puntberaming van parameters voorgestel, gebaseer op die sogenaamde BRAGGing ("Bootstrap Robust AGGregating") beramingsmetodes, deur gebruik te maak van beide die oorspronklike naïewe $m$-uit-$n$ skoenlusmetodologie en die nuwe gekorrigeerde skoenlusmetodologie. Die beraming van die hersteekproefgrootte vir hierdie beramingssprobleem word ook aangespreek deur gebruikmaking van Cornish-Fisher en ander ontwikkelings.

⋆ Tweedens word die toepassing van die $m$-uit-$n$ skoenlus op hipotesetoetsing aangepak. Twee nuwe data-afhanklike keuses van die hersteekproefgroote, $m$, word voorgestel in die hipotesetoetsingopset. Die eerste beramer word gegrond op die skoenlusberamer van die toetsgrootte as 'n skoenluskritieke waarde gebruik word, en die tweede beramer word gebaseer op die waarskynlikheidstrukture van die $p$-waardes van die toets onder die nulhipotese.

In albei van hierdie nuwe toepassingsgebiede word die data-afhanklike keuses van $m$ beide teoreties en numeries gemotiveer. Die teoretiese doelwit word bereik deur die gebruik van omvattende wiskundige argumente en numeries word uitgebreide Monte-Carlo simulasies uitgevoer.

# Contents

# Chapter 1

# Introduction

## 1.1 Overview

First discussed in Efron (1979), the bootstrap is a resampling procedure that, over the last 30 years, has been on the receiving end of a great deal of attention in statistical literature. The reason for this can possibly be ascribed to the large number of attractive properties of this procedure, namely,

- ⋆ the bootstrap is easily applied by the practitioner,
- ⋆ it requires very few, non-restrictive, assumptions, and
- ⋆ it can be applied to a vast number of situations where traditional theory can become unwieldy, almost regardless of the theoretical complexity of the problem being considered.

However, the procedure is not without its flaws, since while it is applicable in a large number of situations, there are cases where it fails. These cases are referred to as the *non-regular cases* (Shao and Tu 1995). The remedies for these non-regular cases typically involve changing either the statistic being used or the scheme used to resample the data. The simplest of these remedies uses the latter solution, and involves resampling fewer observations than appear in the original sample, in particular it samples $m$ observations from the $n$ original observations. This modification of the bootstrap has been called the $m$-out-of-$n$ bootstrap, but it has not enjoyed the same degree of exposure in the literature shared by its cousin, the traditional $n$-out-of-$n$ bootstrap . In this thesis overviews of both the traditional bootstrap and the $m$-out-of-$n$ bootstrap are provided.

The $m$-out-of-$n$ bootstrap, while being a very useful techinique for rectifying the inconsistency problems suffered by the traditional bootstrap, does unfortunately require that we know the re-sample size $m$. Data-dependent choices of $m$ differ depending on the problems being considered. Some of these choices have been discussed for a handful of these problems in the literature. This thesis contributes to this collection of methods for selecting $m$ by considering two new problems: point estimation using robust aggregating methods, and bootstrap hypothesis testing.

## 1.2 Objectives

- ⋆ Provide an overview of the traditional nonparametric bootstrap by looking at various applications of the technique.
- ⋆ Discuss the consistency of the traditional bootstrap, focusing in particular on the situations where the bootstrap is *not* consistent.

- ⋆ Describe various modifications that can be made to the traditional bootstrap by altering the resampling scheme.
- ⋆ Provide an overview of the $m$-out-of-$n$ bootstrap.
- ⋆ Investigate the literature on the remedies to the bootstrap consistency failures, focusing on those remedies that make use of the $m$-out-of-$n$ bootstrap and then briefly discussing the consistency of the results obtained from these methods.
- ⋆ Review the literature on the choice of the resample size $m$ when using the $m$-out-of-$n$ bootstrap applied in some settings.
- ⋆ Provide an overview of Edgeworth expansions and Cornish-Fisher expansions.
- ⋆ Briefly describe bootstrap aggregating and bootstrap robust aggregating.
- ⋆ Develop a new method, based on bootstrap robust aggregating, for the estimation of a parameter using the $m$-out-of-$n$ bootstrap approach. A number of data-dependent choices of $m$ for this technique are also derived.
- ⋆ Conduct simulation studies to determine the effectiveness of the new estimation method and of the data-based choices of $m$ which have been developed.
- ⋆ Briefly describe the concepts related to bootstrap hypothesis testing. The application of the $m$-out-of-$n$ bootstrap for hypothesis testing is also discussed.
- ⋆ Possible data-dependent choices of $m$, when the $m$-out-of-$n$ bootstrap is applied to hypothesis testing, are discussed. The methods are based on the calculation of bootstrap critical values and bootstrap $p$-values.

## 1.3 Thesis outline

After this introductory chapter, the thesis begins by looking at a basic review of the traditional bootstrap in Chapter 2. This chapter discusses all of the necessary basic techniques and notation that one needs in order to understand the subsequent chapters.

Chapter 3 discusses the consistency of the bootstrap. The chapter begins by declaring some notation that is used throughout the chapter. The consistency of the bootstrap is then investigated by considering the various cases in the literature where it has been found to be inconsistent. Each of the cases is defined fairly thoroughly before the bootstrap consistency result is stated.

In Chapter 4 the $m$-out-of-$n$ bootstrap is discussed by first considering the variations of the $m$-out-of-$n$ bootstrap that are mentioned in various sources (see for example, Bickel, Götze and van Zwet (1997), Politis, Romano and Wolf (1999), etc.) The consistency of the $m$-out-of-$n$ bootstrap is discussed by comparing the inconsistent results which arose in Chapter 3 with the results obtained from the $m$-out-of-$n$ bootstrap. A framework for the 'correct' implementation of the $m$-out-of-$n$ bootstrap is then provided and the chapter concludes with a short literature study of the assorted data-based methods used to select $m$ when applying the $m$-out-of-$n$ bootstrap in a handful of scenarios. For each of these choices an algorithm is provided.

A nonparametric point estimation method which employs BAGGing, BRAGGing and the $m$-out-of-$n$ bootstrap is developed in Chapter 5. In this chapter the construction of an estimator using the BRAGGing techniques found in Swanepoel (1988) and Berrendero (2007) is discussed. In addition to this, potential data-dependent choices of $m$ in this new estimation procedures are derived by making use of Cornish-Fisher expansions and population moment estimators. The derivations found in this chapter are quite lengthy, making their inclusion in the chapter quite awkward. The full derivations of theoretical statments which appear in this chapter are presented in Appendix D. A number of data-based choices of $m$ derived from these results are proposed in this chapter and a simulation study is conducted to determine the adequacy of each one.

The final chapter in this thesis, Chapter 6, provides a more detailed description of bootstrap hypothesis testing than is provided in Chapter 2. The $m$-out-of-$n$ bootstrap is considered with the primary aim of selecting the resample size in a practical application of hypothesis testing. Two different possible data-dependent methods of choosing $m$ are developed; the first is based on the calculation of bootstrap critical values and the second is based on the calculation of bootstrap $p$-values . Both choices are extentively evaluated by making use of Monte-Carlo simulation studies.

## 1.4 Notation

Some of the notation which is used in this thesis is briefly summarized here.

| **Symbol** | **Meaning** |
|---|---|
| $\stackrel{a.s.}{\to}$ | Almost sure convergence. |
| $\stackrel{p}{\to}$ | Convergence in probability. |
| $\stackrel{\mathcal{D}}{\to}$ | Convergence in distribution or weak convergence. |
| $\mathbb{R}$ | Real numbers. |
| $\mathbb{N}$ | Natural numbers. |
| $\rho$ | General distance measure for functions. |
| $\rho_\infty$ | Distance measure for functions generated by the sup norm. |
| $\mathrm{P}(\cdot)$ | Probability measure. |
| $\mathrm{P}^*(\cdot)$ | Bootstrap probability measure. |
| $\mathrm{E}(\cdot)$ | Expected value. |
| $\mathrm{E}^*(\cdot)$ | Bootstrap expected value. |
| $\mathrm{Var}(\cdot)$ | Variance. |
| $\mathrm{Var}^*(\cdot)$ | Bootstrap variance. |

# Chapter 2

# The traditional bootstrap

## 2.1 Introduction

The bootstrap method is a sub-branch of a much larger collection of methods broadly known as resampling methods. Included in this set of methods are the Jackknife and cross-validation methods, to name but two. Clearly the idea of resampling has been around for many years, but interest was renewed with the advent of powerful computer hardware which allowed, for the first time, for viable and practical applications of Monte-Carlo simulation methods. Resampling was explored from a different perspective in an article by Bradley Efron (Efron 1979) where he first introduced what has since been labeled "the bootstrap" *. A possible reason why this name was associated with this method is because the method appears to be able to obtain results concerning the sampling distribution of random variables based solely on a single set of sample data, making it almost appear as if we get 'something for nothing'.

Efron and Tibshirani (1993) define the bootstrap method as *a computer based method for assigning measures of accuracy to statistical estimates.* However, this definition is rather restrictive for two reasons:

1. the bootstrap method is capable of more than simply assigning measures of accuracy (such as standard error) to statistical measures, since one can also calculate measures of precision (such as bias), among other things;
2. it is not always necessary to make use of a computer to calculate these measures since occasionally these results can be obtained analytically.

A slightly more informative (but less succinct) definition might be the following: *The bootstrap is a technique that can estimate population parameters and distributional properties of statistics by substituting the population mechanism used to obtain the parameter with an empirical equivalent. These estimates can be obtained analytically, but they are mostly obtained through the use of resampling and Monte-Carlo methods carried out on a computer.*

The bootstrap discussed in this chapter will be called the *traditional bootstrap* to distinguish it from other types of bootstraps which will also be discussed in this thesis. The traditional bootstrap enjoys a wide variety of applications including estimation of standard error and bias, construction of confidence intervals, and hypothesis testing. These topics will be briefly covered in this chapter. In addition to this, many of these applications require one to iterate the bootstrap with what is commonly called the "double bootstrap". This technique will also be briefly discussed.

*The etymology of the word "bootstrap" apparently derives from the concept of being able to pull oneself up out of a hole by tugging on one's bootstraps (Efron and Tibshirani 1993). This apparently impossible feat was accomplished by the title character in the novel "The Adventures of Baron Munchausen" by Rudolph Erich Raspe.

The chapter that follows this one will pick up on these ideas with a topic which is quite pertinent to this thesis, i.e., situations where the bootstrap can fail. This discussion will include some necessary conditions for these failures to occur.

## 2.2 Notation

The notation used in bootstrap calculations is quite peculiar and can be slightly confusing to the uninitiated. The following list is a summary of the basic notation that is used in the basic application of the bootstrap and also in the application of the double bootstrap. Further iterations of the bootstrap (beyond the double bootstrap) require more notation, but the extension of the listed notation is logical (typically only requiring the addition of more 'stars'). This notation will be used in this and all subsequent chapters.

- ⋆ Let $\boldsymbol{X}_n = \{X_1, X_2, \ldots, X_n\}$ denote independent and identically distributed (i.i.d) random variables (R.V.'s) from some unknown distribution function, $F$, i.e., $\boldsymbol{X}_n = \{X_1, X_2, \ldots, X_n\}$ is a random sample of size $n$ drawn from $F$.

- ⋆ Let $\boldsymbol{X}_n^* = \{X_1^*, X_2^*, \ldots, X_n^*\}$ denote a sample of size $n$ drawn independently from $\hat{F}$, where $\hat{F}$ is an estimator of the true distribution function $F$.

- ⋆ Let $\boldsymbol{X}_n^{**} = \{X_1^{**}, X_2^{**} \ldots, X_n^{**}\}$ denote a sample of size $n$ drawn independently from $\hat{F}^*$, where $\hat{F}^*$ is an estimator of the distribution function $\hat{F}$.

- ⋆ Let $\theta$ be a population parameter of interest. This parameter is sometimes given as a functional, say $t$, of the unknown distribution function $F$. The parameter can then be written as $\theta = t(F)$.

- ⋆ Let $\hat{\theta}_n = \hat{\theta}_n(X_1, X_2, \ldots, X_n)$ be the estimate for the population parameter $\theta$ based on sample data. Occasionally this statistic will be expressed in terms of a functional, say $t$, of an empirical estimate of $F$, denoted $\hat{F}$, which is based on the observed sample $X_1, X_2, \ldots, X_n$. The statistic is then given by $\hat{\theta}_n = t(\hat{F})$.

- ⋆ Let $\hat{\theta}_n^* = \hat{\theta}_n(X_1^*, X_2^*, \ldots, X_n^*)$ be the bootstrap statistic based on 'resampled' sample data. Occasionally this bootstrap statistic will be expressed in terms of a functional, say $t$, of an empirical estimate of $\hat{F}$, denoted $\hat{F}^*$, based on the bootstrap sample $X_1^*, X_2^*, \ldots, X_n^*$. The statistic is then given by $\hat{\theta}_n^* = t(\hat{F}^*)$.

- ⋆ Let $\hat{\theta}_n^{**} = \hat{\theta}_n(X_1^{**}, X_2^{**} \ldots, X_n^{**})$ be the double bootstrap statistic, based on $X_1^{**}, X_2^{**} \ldots, X_n^{**}$ obtained by resampling from $X_1^*, X_2^*, \ldots, X_n^*$.

- ⋆ Let $\mathrm{P}(\cdot)$ denote the probability operator

- ⋆ Let $\mathrm{P}^*(\cdot)$ denote the bootstrap probability operator. The relationship between the probability operators P and P* is:

$$\mathrm{P}^*\left(X_i^* \leq x\right) = \mathrm{P}\left(X_i^* \leq x | X_1, X_2, \ldots, X_n\right).$$

  That is, the bootstrap probability operator is a conditional probability where one conditions on the sample data.

- ⋆ Let $\mathrm{P}^{**}(\cdot)$ denote the double bootstrap probability operator.

- ⋆ Let $\mathrm{E}(\cdot)$ denote the expectation operator when working with the sample data and statistics.

- ⋆ Let $\mathrm{E}^*(\cdot)$ denote the expectation operator when working with resampled (bootstrap) data and statistics. Under the "star" version of expectation, the sample data and sample statistics are viewed as constant. The following describes the relationship between E and E*:

$$\mathrm{E}^*\left(X_i^*\right) = \mathrm{E}\left(X_i^* | X_1, X_2, \ldots, X_n\right).$$

- ⋆ Let $\mathrm{E}^{**}(\cdot)$ denote the expectation operator when working with double bootstrap data and statistics. Under the "star-star" version of expectation, the bootstrap sample data and bootstrap sample statistics are viewed as constant.

- ⋆ Let $\mathrm{Var}(\cdot)$ denote the variance operator when working with sample data or statistics.

- ⋆ Let $\mathrm{Var}^*(\cdot)$ denote the variance operator when working with resampled (bootstrap) data and statistics. Again, under the "star" version of variance, the sample data and sample statistics are viewed as constant. The following describes the relationship between Var and Var*:

$$\mathrm{Var}^*\left(X_i^*\right) = \mathrm{Var}\left(X_i^* | X_1, X_2, \ldots, X_n\right).$$

- ⋆ Let $\mathrm{Var}^{**}(\cdot)$ denote the variance operator when working with double bootstrap data and statistics. Again, under the "star-star" version of variance, the bootstrap sample data and bootstrap sample statistics are viewed as constant.

## 2.3 Basic concepts: Applying the bootstrap

In this section the basic concepts necessary to apply the bootstrap method will be discussed. The two methods for obtaining estimates include deriving exact expressions for the estimates through the plug-in principle (discussed next) and Monte-Carlo simulations executed on a computer.

### 2.3.1 The plug-in principle

A fundamental concept underlying the correct usage of the bootstrap is that of the *plug-in principle.* Efron and Tibshirani (1993) describe the use of the plug-in principle by making reference to, what they call, the "Bootstrap World" and the "Real World". The plug-in principle is the mechanism that allows one to "shift" from the real world to the bootstrap world as represented graphically in Figure 2.1.

Colourful descriptions aside, these concepts help to illustrate how the method works. Efron and Tibshirani (1993) reasoned that a set of sample data, $X_1, X_2, \ldots, X_n$, generated from some unknown probability distribution, say, $F$, situated in the real world, can be viewed as a 'population' or pseudo-population in the bootstrap world. The 'shifting' between the real world and the bootstrap world occurs when one replaces the unknown probability structure, $F$, with an empirical equivalent. Therefore, when one shifts into the bootstrap world using the plug-in principle, one shifts into a situation where the probability structure and population parameters are known. Estimates of real world elements are then simply the corresponding bootstrap world equivalents.

Consider the situation where $\theta$ is some parameter of interest and that this parameter can be expressed as $\theta = t(F)$, i.e., it can be expressed as some functional, $t$, of the unknown distribution function $F$. The plug-in principle then asserts that the bootstrap estimator of the parameter $\theta$ is simply:

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

where $\hat{F}$ is some empirical estimate of the true distribution function, $F$. In other words, the plug-in principle involves estimating the parameter $\theta$, a functional of the unknown distribution function $F$, by simply applying the same functional to an estimated distribution function, $\hat{F}$.

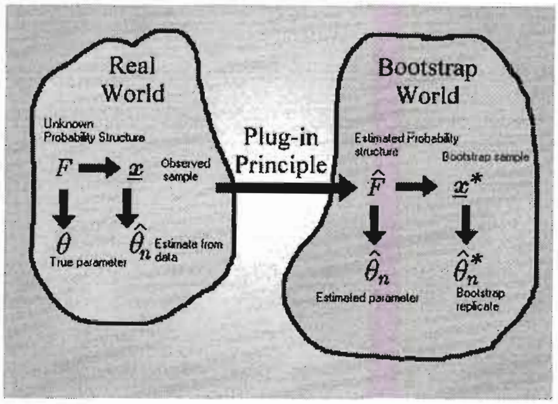


**Figure 2.1:** *Schematic representation of the Plug-in principle with the bootstrap*

The choice of $\hat{F}$ is usually taken to be the *empirical distribution function* (EDF), defined as:

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{I}(X_i \leq x) , \tag{2.1}$$

where $I$ is the indicator function and is defined as follows:

$$\mathrm{I}(A) = \begin{cases} 1, & \text{if } A \text{ occurs} \\ 0, & \text{if } A^c \text{ occurs.} \end{cases}$$

Different choices of $\hat{F}$ lead to different types of bootstrap resampling schemes. However, for this thesis the EDF $F_n$ will be chosen to be the primary estimate of the distribution function $F$.

The reason why the EDF is chosen as the default approximation for the distribution function $F$ is twofold:

1. First, the EDF has many desirable properties as an estimator for $F$. Primary among these is the fact that, according to the Glivenko-Cantelli theorem, $F_n$ converges uniformly, with probability 1, to $F$ as $n$ becomes large.

2. Second, drawing samples *independently* from the EDF reduces to drawing samples *with replacement* from the original sample.

Thus, Monte-Carlo simulations which rely on the EDF have a solid asymptotic basis and are also easy to implement in practice.

### Iterating the plug-in principle (Double bootstrap)

One should note that the bootstrap estimates of real world parameters are just sample statistics in the real world. This means that if, after the plug-in principle is applied once, interest lies in determining the *unknown* distributional properties of the bootstrap estimators (such as the expected value or variance of the bootstrap estimator), then one can iterate the plug-in principle to obtain these estimates. The real and bootstrap world ideas found in Efron and Tibshirani (1993) can be applied once again. However, the shifting now occurs from the real world into the bootstrap world and then into the "Double Bootstrap" world. Figure 2.2 represents this shifting graphically. Quantities that were obtained from the initial application of the plug-in principle are

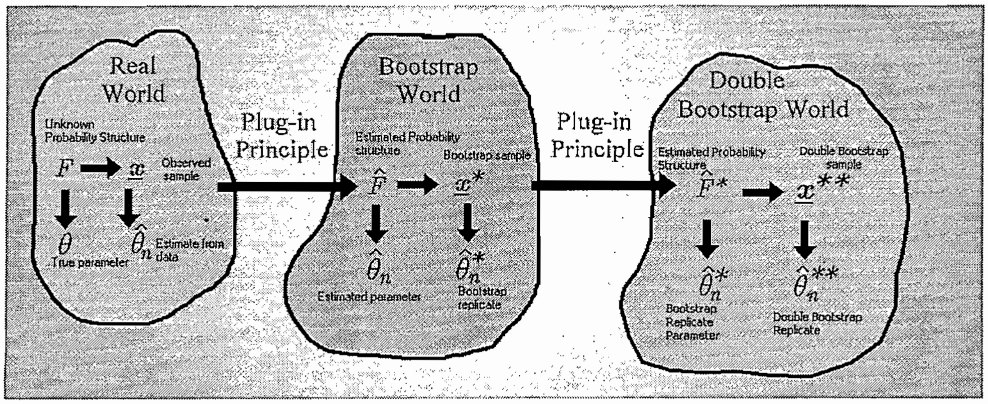


**Figure 2.2:** *Schematic representation of the Plug-in principle with the double bootstrap*

now 'estimated' again using a second application of the plug-in principle. These new statistics are based on the EDF, $F_n^*$ of the resampled data, $X_1^*, X_2^*, \ldots, X_n^*$. Define $F_n^*$ as

$$F_n^*(x) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{I}(X_i^* \leq x) \ .$$

Once again, if the parameter of interest is $\theta = t(F)$, the estimator of this parameter is the bootstrap or plug-in estimate $\hat{\theta}_n = t(F_n)$, then the double bootstrap 'estimator' of $\hat{\theta}_n$ is $\hat{\theta}_n^* = t(F_n^*)$. Some examples of this double bootstrap will be explored in Section 2.4.

### 2.3.2 Simulation

The practical implementation of the bootstrap method will often involve a Monte-Carlo simulation carried out on a computer, but it is sometimes possible to apply the bootstrap without writing a single line of code. This is because it is possible (in some circumstances) to obtain explicit expressions for the estimates. The simplest example of this is to make use of the bootstrap method to obtain an estimate for the population mean. If it is assumed that the population mean $\mu$ is the mean of the entire population data, and that it can be expressed in functional form as $\mu = t(F)$, then by applying the plug-in principle, an estimate would simply be the mean of the sample data expressed as $\bar{X}_n = t(F_n)$. Thus, the sample mean is the bootstrap estimate of the population mean. Taking this idea further we can easily show that the bootstrap estimator of $\mathrm{Var}(\bar{X}_n)$ is equal to $S_n^2/n$, where $S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$. In the trivial example presented above, the population distribution function was replaced with the EDF; this is the basic idea behind the plug-in principle and it is the essence of the bootstrap method. Naturally, more complex examples will not reduce to simple expressions like it did with the sample mean, but fortunately the results in these cases can easily be approximated using Monte-Carlo simulations.

Monte-Carlo simulations in general involves repeatedly drawing samples from a specified, known distribution and then calculating the statistic of interest for each of these generated samples. However, in the context of the bootstrap Monte-Carlo simulations, repeated samples will be drawn from the EDF defined in (2.1).

The Monte-Carlo algorithms that will be discussed include a number of terms which need to be defined:

⋆ The number of iterations specified in the algorithm are known as the *number of bootstrap replications* and is usually denoted by $B$.

⋆ The samples drawn at each iteration are called the *bootstrap samples.*

⋆ The statistics calculated at each iteration of the simulation (i.e., for each bootstrap sample) are known as *bootstrap replications* or *bootstrap statistics.*

The following algorithm now briefly describes how a basic bootstrap Monte-Carlo simulation will be conducted in order to create $B$ bootstrap replications of a statistic $\hat{\theta}_n$.

**Basic Monte-Carlo bootstrap algorithm:**

1. Generate a bootstrap sample of $n$ independent observations, $X_1^*, X_2^*, \ldots, X_n^*$, from the EDF, $F_n$, i.e., sample with replacement from $X_1, X_2, \ldots, X_n$.
2. Calculate the statistic $\hat{\theta}_n^* = \hat{\theta}_n(X_1^*, X_2^*, \ldots, X_n^*)$ for the sample generated in step (1).
3. Independently repeat steps (1) and (2) $B$ times. The statistic calculated in step (2) in the $b^{\text{th}}$ iteration will be denoted by $\hat{\theta}_{n,b}^*$. The result is the following set of bootstrap replications: $\hat{\theta}_{n,1}^*, \hat{\theta}_{n,2}^*, \ldots, \hat{\theta}_{n,B}^*$.

A histogram of the generated bootstrap replications obtained in the above algorithm can be used as an approximation of the sampling distribution of the statistic $\hat{\theta}_n$.

At this point it seems important to state that the above algorithm is used to approximate the *ideal* bootstrap estimate of the sampling distribution of the statistic $\hat{\theta}_n$. In other words, the Monte-Carlo result can only provide an approximation of the estimate and not the estimate itself. Only when the number of iterations is increased to infinity (or when every possible combination of samples are drawn from the original sample) does the answer approach the ideal bootstrap estimate value. Fortunately, it has been documented (Efron and Tibshirani 1993) that using repetitions as small as $B = 1000$ can still produce sufficiently accurate results for certain applications of the bootstrap such as for standard error estimation. However, modern technology allows larger numbers of replications to be calculated without a significant increase in computational time. Thus one can make accurate approximations of the ideal bootstrap estimate rather quickly.

## 2.4 Application of the bootstrap

In this section various applications of the bootstrap will be briefly discussed. In each case the 'plug-in' expression for the estimate will be given and the Monte-Carlo algorithm which can be used to approximate the estimate will be provided.

### 2.4.1 Estimating standard error

The standard error of some statistic $\hat{\theta}_n = t(F_n)$ based on the sample data $X_1, X_2, \ldots, X_n$ can be expressed as $\text{SE}(\hat{\theta}_n) = \sqrt{\text{Var}(\hat{\theta}_n)} = \sqrt{\text{Var}(t(F_n))}$. The bootstrap estimate of this quantity is obtained by making use of the plug-in principle, and an ideal bootstrap estimate of the standard error of $\hat{\theta}_n$ is given by $\text{SE}^*(\hat{\theta}_n^*) = \sqrt{\text{Var}^*(\hat{\theta}_n^*)} = \sqrt{\text{Var}^*(t(F_n^*))}$, (Efron and Tibshirani 1993, Davison and Hinkley 1997).

The estimated standard error of a general statistic, $\hat{\theta}_n$ can be then approximated using a Monte-Carlo simulation. Note that the ideal bootstrap estimated standard error of the statistic is denoted by $\text{SE}^*(\hat{\theta}_n^*)$, while the Monte-Carlo approximation of this quantity is denoted by $\text{SE}_B^*(\hat{\theta}_n^*)$.

**Monte-Carlo bootstrap algorithm for $\mathrm{SE}^*_B(\hat{\theta}^*_n)$ :**

1. Generate a bootstrap sample of $n$ independent observations, $X^*_1, X^*_2, \ldots, X^*_n$, from the EDF, $F_n$, i.e., sample with replacement from $X_1, X_2, \ldots, X_n$.
2. Calculate the statistic $\hat{\theta}^*_n = \hat{\theta}_n(X^*_1, X^*_2, \ldots, X^*_n)$ for the sample generated in step (1).
3. Independently repeat steps (1) and (2) $B$ times. The statistic calculated in step (2) in the $b^{\text{th}}$ iteration will be denoted by $\hat{\theta}^*_{n,b}$. We obtain the following bootstrap replications: $\hat{\theta}^*_{n,1}, \hat{\theta}^*_{n,2}, \ldots, \hat{\theta}^*_{n,B}$.
4. Now, calculate:

$$\mathrm{SE}^*_B(\hat{\theta}^*_n) = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}(\hat{\theta}^*_{n,b} - \hat{\theta}^*_{n,\bullet})^2} \ ,$$

where

$$\hat{\theta}^*_{n,\bullet} = \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}^*_{n,b} \ .$$

### 2.4.2 Double bootstrap: Estimating the standard error of bootstrap standard error

Once one has an estimate of the standard error of a statistic $\hat{\theta}_n$ (obtained using the bootstrap), it is possible to also obtain the standard error of this bootstrap estimator. This technique requires iteratively applying the plug-in principle, also called the double bootstrap as explained previously. The algorithm for the double bootstrap (given below) is derived by applying the plug-in principle to the standard error of the bootstrap estimate of standard error (Efron and Tibshirani 1993). The standard error of the bootstrap estimate of standard error is given by

$$\mathrm{SE}\left(\mathrm{SE}^*(\hat{\theta}^*_n)\right) = \sqrt{\mathrm{Var}\left(\sqrt{\mathrm{Var}^*(\hat{\theta}^*_n)}\right)},$$

and the plug-in principle applied to this quantity yields the following expression:

$$\mathrm{SE}^*\left(\mathrm{SE}^{**}(\hat{\theta}^{**}_n)\right) = \sqrt{\mathrm{Var}^*\left(\sqrt{\mathrm{Var}^{**}(\hat{\theta}^{**}_n)}\right)},$$

where $\hat{\theta}^{**}_n = \hat{\theta}_n(X^{**}_1, X^{**}_2 \ldots, X^{**}_n)$.

If the ideal double bootstrap estimate for the standard error of the bootstrap standard error is represented by $\tilde{\sigma}_{SE} = \mathrm{SE}^*\left(\mathrm{SE}^{**}(\hat{\theta}^{**}_n)\right)$ then the simulation approximation will be given by $\tilde{\sigma}_{SE_B}$.

The algorithm for applying this double bootstrap is provided below.

**Approximating the standard error of the estimated bootstrap standard error of $\hat{\theta}_n$:**

1. Generate a bootstrap sample $X_1^*, X_2^*, \ldots, X_n^*$ from the EDF, $F_n$, i.e., sample with replacement from $X_1, X_2, \ldots, X_n$.

   a. Generate a double bootstrap sample $X_1^{**}, X_2^{**} \ldots, X_n^{**}$ from the EDF $F_n^*$, i.e., sample with replacement from $X_1^*, X_2^*, \ldots, X_n^*$.

   b. Calculate $\hat{\theta}_n^{**} = \hat{\theta}_n(X_1^{**}, X_2^{**} \ldots, X_n^{**})$.

   c. Repeat steps (a) and (b) $R$ times. The statistic calculated in step (b) in the $r^{\text{th}}$ iteration will be denoted by $\hat{\theta}_{n,r}^{**}$. We obtain the following double bootstrap replications $\hat{\theta}_{n,1}^{**}, \hat{\theta}_{n,2}^{**}, \ldots \hat{\theta}_{n,R}^{**}$.

   d. Now calculate:

$$\hat{\sigma}_{SE_R} = \sqrt{\frac{1}{R-1}\sum_{r=1}^{R}(\hat{\theta}_{n,r}^{**} - \hat{\theta}_{n,\bullet}^{**})^2} \ ,$$

   where

$$\hat{\theta}_{n,\bullet}^{**} = \frac{1}{R}\sum_{r=1}^{R}\hat{\theta}_{n,r}^{**} \ .$$

2. Independently repeat step (1) $B$ times. The statistic calculated in step (1) in the $b^{\text{th}}$ iteration will be denoted by $\hat{\sigma}_{SE_R,b}$. We obtaining the following bootstrap replications: $\hat{\sigma}_{SE_R,1}, \hat{\sigma}_{SE_R,2}, \ldots, \hat{\sigma}_{SE_R,B}$.

3. Now, calculate:

$$\tilde{\sigma}_{SE_B} = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}(\hat{\sigma}_{SE_R,b} - \hat{\sigma}_{SE_R,\bullet})^2} \ ,$$

   where

$$\hat{\sigma}_{SE_R,\bullet} = \frac{1}{B}\sum_{b=1}^{B}\hat{\sigma}_{SE_R,b} \ .$$

### 2.4.3 Estimating bias

Let the bias of some estimator $\hat{\theta}_n$, which estimates the population parameter $\theta$, be denoted by:

$$\begin{aligned}\beta &= \mathrm{E}(\hat{\theta}_n) - \theta \\ &= \mathrm{E}(s(\boldsymbol{X}_n)) - t(F). \end{aligned} \tag{2.2}$$

The function $s(\cdot)$ is a function applied to the sample data to obtain $\hat{\theta}_n$, such that $\hat{\theta}_n$ is not necessarily the plug-in estimate of $\theta$. Note that the plug-in estimate of $\theta = t(F)$ would have been denoted $\hat{\theta}_n = t(F_n)$. The distinction between $s(\boldsymbol{X}_n)$ and $t(F_n)$ is made because the estimator $\hat{\theta}_n$ does not *have* to be the plug-in estimate for $\theta$. Of course, if $\hat{\theta}_n$ is the plug-in estimate, then $\hat{\theta}_n = s(\boldsymbol{X}_n) = t(F_n)$.

In order to estimate the quantity in (2.2) with the bootstrap, the plug-in principle is applied, i.e.,

$$\begin{aligned}\hat{\beta} &= \mathrm{E}^*(\hat{\theta}_n^*) - \hat{\theta}_n \\ &= \mathrm{E}^*(s(\boldsymbol{X}_n^*)) - t(F_n). \end{aligned} \tag{2.3}$$

Note that if the estimator $\hat{\theta}_n$ is indeed the plug-in estimate of $\theta$ then the above expression becomes

$$\hat{\beta} = \mathrm{E}^*(t(F_n^*)) - t(F_n). \tag{2.4}$$

A good feature of the plug-in principle is that, even though the plug-in estimates $t(F_n)$ are not necessarily completely unbiased for $t(F)$ (consider the plug-in estimate for the standard deviation), they do tend to have small biases (Efron and Tibshirani 1993, p. 125).

Both equations (2.3) and (2.4) are the ideal bootstrap results as discussed in Efron and Tibshirani (1993) and Davison and Hinkley (1997). As usual, these values have to be estimated by making use of a Monte Carlo algorithm and this approximation will be denoted by $\hat{\beta}_B$. The following algorithm estimates the ideal bootstrap estimate of bias:

**Monte-Carlo bootstrap algorithm for bootstrap bias**

1. Calculate the plug-in estimate of the parameter $\theta = t(F)$, i.e., calculate $t(F_n)$.
2. Generate a bootstrap sample of $n$ independent observations, $X_1^*, X_2^*, \ldots, X_n^*$, from the EDF, $F_n$, i.e., sample with replacement from $X_1, X_2, \ldots, X_n$.
3. Calculate the bootstrap statistic $\hat{\theta}_n^* = s(\boldsymbol{X}_n^*)$ for the sample generated in step (2).
4. Independently repeat steps (2) and (3) $B$ times. The statistic calculated in step (2) in the $b^{\text{th}}$ iteration will be denoted by $\hat{\theta}_{n,b}^*$. The result is that we obtain the following bootstrap replications: $\hat{\theta}_{n,1}^*, \hat{\theta}_{n,2}^*, \ldots, \hat{\theta}_{n,B}^*$.
5. Calculate the approximate estimated bias as:

$$\hat{\beta}_B = \hat{\theta}_{n,\bullet}^* - t(F_n),$$

where $\hat{\theta}_{n,\bullet}^* = \frac{1}{B}\sum_{b=1}^{B} \hat{\theta}_{n,b}^*$.

Other improved algorithms are available for approximating the bootstrap bias, see Efron and Tibshirani (1993) and Davison and Hinkley (1997).

### 2.4.4 Estimation of sampling distributions

The traditional bootstrap is, in general, able to estimate the sampling distribution of a random variable $R_n(\mathbf{X}_n; F)$, which may now depend on the unknown distribution function $F$, using the same methods that have already been discussed. The distribution of the random variable is given by:

$$G(x) = P(R_n(\mathbf{X}_n; F) \leq x) \qquad x \in \mathbb{R}.$$

The bootstrap estimator of $G(x)$ is obtained by once again applying the plug-in principle, i.e.,

$$\hat{G}(x) = P^*(R_n(\mathbf{X}_n^*; F_n) \leq x) \qquad x \in \mathbb{R}.$$

Consider the following example: If the random variable in question is defined as $R_n(\mathbf{X}_n; F) = \sqrt{n}(\bar{X}_n - \mu)/S_n(\mathbf{X}_n)$, where $\bar{X}_n$ and $S_n(\mathbf{X}_n)$ are the sample mean and sample standard deviation respectively, then the bootstrap statistic becomes

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

$\hat{G}(x)$, the bootstrap estimate of the sampling distribution of $R_n(\mathbf{X}_n; F)$, can be approximated using $\hat{G}_B(x)$. The Monte-Carlo simulation algorithm for calculating $\hat{G}_B(x)$ is:

**Basic Monte-Carlo bootstrap algorithm:**

1. Generate a bootstrap sample of $n$ independent observations, $X_1^*, X_2^*, \ldots, X_n^*$, from the EDF, $F_n$, i.e., sample with replacement from $X_1, X_2, \ldots, X_n$.
2. Calculate the statistic $R_n^* = R_n(\boldsymbol{X}_n^*; F_n)$ for the sample generated in step (1).
3. Independently repeat steps (1) and (2) $B$ times. The statistic calculated in step (2) in the $b^{\text{th}}$ iteration will be denoted by $R_{n,b}^*$. The result is the following set of bootstrap replications: $R_{n,1}^*, R_{n,2}^*, \ldots, R_{n,B}^*$.
4. Finally, calculate
$$\hat{G}_B(x) = \frac{1}{B}\sum_{b=1}^{B} \mathrm{I}\left(R_{n,b}^* \le x\right).$$

**Remarks:**

1. The bootstrap estimator $\hat{G}(x)$ for $G(x)$ can be shown to be asymptotically valid for many statistics. The proof of this validity typically involves proving whether or not the maximum difference between the two distributions $\hat{G}(x)$ and $G(x)$ converges to zero in probability or almost surely as $n \to \infty$, i.e.,
$$\sup_{x\in\mathbb{R}} |\hat{G}(x) - G(x)| = o(1),$$
almost surely (or in probability). The estimator is said to be *first-order accurate* if $o(1)$ can be replaced by $O(n^{-1/2})$. If the rate of convergence is of the order $o(n^{-1/2})$, then the estimator is said to be *second-order accurate.* Recall that a simple normal approximation is usually only first-order accurate, while the bootstrap estimator is usually second-order accurate.
2. Statistical literature contains many examples of statistics that have been shown to be either first or second-order accurate. Some examples include $L$-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, Muñoz-García and Pino-Mejías (2003)).

### 2.4.5 Bootstrap confidence intervals

In the following section the five most common ways in which the bootstrap can be applied to construct confidence intervals for a parameter are investigated. The extension to confidence upper and lower bounds is arbitrarily easy and will not be discussed. The five methods are:

1. The Basic or Backwards Percentile Method.

2. The Bias-Corrected Percentile Method ($BC$).
3. The Accelerated Bias-Corrected Percentile Method ($BC_a$).
4. The Bootstrap-t Interval (*Bootstrap-t*).
5. The Hybrid Percentile Method †.

All of these methods involve the estimation of percentiles in one way or another. That is, the quantiles of the distribution of the statistic are used in the interval in some way. These quantiles are typically estimated by using the distribution function of the bootstrap statistic. All the simulation algorithms which are then used to approximate these intervals will thus involve arranging the bootstrap replications in ascending order and then choosing the element that occurs at a certain index, i.e., the bootstrap distribution's approximate quantiles. The index is calculated as some function of the number of bootstrap replications, $B$, and the chosen significance level, $\alpha$. Methods 1, 2, 3 and 5 make use of this method almost directly while the method listed as number 4 makes use of the more traditional Student-$t$ concept (i.e., it employs the quantiles from some Studentized distribution).

In order to explain these methods the following distribution functions must be defined first. Let $G$ denote the distribution function of the statistic $\hat{\theta}_n = \hat{\theta}_n(X_1, X_2, \ldots, X_n)$, i.e.,

$$G(x) = \mathrm{P}(\hat{\theta}_n \leq x), \qquad x \in \mathbb{R}.$$

Making use of the plug-in principle, let $\hat{G}$ denote the distribution function of the bootstrap statistic $\hat{\theta}_n^* = \hat{\theta}_n(X_1^*, X_2^*, \ldots, X_n^*)$, i.e.,

$$\hat{G}(x) = \mathrm{P}^*(\hat{\theta}_n^* \leq x), \qquad x \in \mathbb{R}. \tag{2.5}$$

### The basic percentile method

Percentile confidence interval methods are based on the percentiles of the distribution function $G$, i.e., the lower and upper bounds of the interval for the parameter $\theta$ can be obtained by simply using the quantiles $G^{-1}(\alpha)$ and $G^{-1}(1-\alpha)$ (DiCiccio and Efron 1996). These quantiles can be estimated by using $\hat{G}$, the plug-in estimate of $G$. The resulting interval is called the *basic or backwards percentile* interval. The basic percentile $100(1-\alpha)\%$ confidence interval for $\theta$ is then defined as:

$$I_b = \left[\hat{G}^{-1}\left(\frac{\alpha}{2}\right); \hat{G}^{-1}\left(1 - \frac{\alpha}{2}\right)\right] .$$

The following Monte-Carlo algorithm can be used to approximate this estimated interval:

† Note: The names 'Backward' and 'Hybrid' are adopted since they are also used in Chung and Lee (2001).

**Approximating the basic percentile confidence interval through simulation**

1. Generate $X_1^*, X_2^*, \ldots, X_n^*$ independently from the EDF, $F_n$, i.e., generate $X_1^*, X_2^*, \ldots, X_n^*$ by sampling with replacement from $X_1, X_2, \ldots, X_n$.
2. Calculate $\hat{\theta}_n^* = \hat{\theta}_n(X_1^*, X_2^*, \ldots, X_n^*)$.
3. Repeat steps 1 and 2 $B$ times obtaining $\hat{\theta}_{n,1}^*, \hat{\theta}_{n,2}^*, \ldots, \hat{\theta}_{n,B}^*$.
4. Obtain the order statistics $\hat{\theta}_{n,(1)}^* \leq \hat{\theta}_{n,(2)}^* \leq \cdots \leq \hat{\theta}_{n,(B)}^*$.
5. The interval is then:
$$I_{b,B} = \left[\hat{\theta}_{n,(r)}^*; \hat{\theta}_{n,(s)}^*\right] ,$$
where
$$r = \left\lfloor (B+1)\cdot\left(\frac{\alpha}{2}\right)\right\rfloor \quad \text{and} \quad s = \left\lfloor (B+1)\cdot\left(1-\frac{\alpha}{2}\right)\right\rfloor .$$

**The bias-corrected ($BC$) percentile method**

According to Davison and Hinkley (1997) the basic percentile method suffers from bias and so a correction was introduced to correct for this bias. The biased basic percentile method can be corrected by altering the quantiles calculated for the interval. The bias-corrected percentile $100(1-\alpha)\%$ confidence interval for $\theta$ is given by:

$$I_{bc} = \left[\hat{G}^{-1}\left\{\Phi\left(2z_0 - z\left(\frac{\alpha}{2}\right)\right)\right\}; \hat{G}^{-1}\left\{\Phi\left(2z_0 + z\left(\frac{\alpha}{2}\right)\right)\right\}\right] ,$$

where $\Phi$ is the standard normal distribution function, $z(\frac{\alpha}{2})$ is defined as the standard normal quantiles such that $\Phi\left(z\left(\frac{\alpha}{2}\right)\right) = 1 - \frac{\alpha}{2}$. The quantity $z_0$ is known as the bias correction. The bias correction can roughly be seen as a measure of the median discrepancy, or bias, between the quantities $\hat{\theta}_n^*$ and $\hat{\theta}_n$ in normal units (Efron and Tibshirani 1993). It can be estimated by $\hat{z}_0$ as follows:

$$\Phi(\hat{z}_0) = \hat{G}(\hat{\theta}_n) \qquad \text{or} \qquad \hat{z}_0 = \Phi^{-1}(\hat{G}(\hat{\theta}_n)).$$

A detailed description of the reasoning behind this estimator can be found in Efron and Tibshirani (1993) and Efron (1987).

To approximate this interval the same algorithm for the basic percentile is used. The only difference is that the last step is replaced with the following:

**Approximating the $BC$ percentile confidence interval through simulation**

5. The interval is then:

$$I_{bc,B} = \left[\hat{\theta}^*_{n,(r)}; \hat{\theta}^*_{n,(s)}\right] ,$$

where

$$r = \left\lfloor (B+1) \cdot \Phi\left(2\hat{z}_0 - z\left(\frac{\alpha}{2}\right)\right)\right\rfloor$$

and

$$s = \left\lfloor (B+1) \cdot \Phi\left(2\hat{z}_0 + z\left(\frac{\alpha}{2}\right)\right)\right\rfloor ,$$

where $\hat{z}_0 = \Phi^{-1}\left(\hat{G}\left(\hat{\theta}_n\right)\right)$, and the expression $\hat{G}(\hat{\theta}_n)$ appearing in the definition of $\hat{z}_0$, can be approximated by:

$$\hat{G}(\hat{\theta}_n) = \frac{1}{B}\sum_{b=1}^{B} \mathrm{I}(\hat{\theta}^*_{n,b} \leq \hat{\theta}_n) .$$

### The accelerated bias-corrected ($BC_a$) percentile method

In addition to correcting for bias, it is also possible to correct for skewness. The accelerated bias-corrected percentile method does this by not only including the properties of the bias-corrected method, but also by adjusting for any problems arising from skewness. The method is described in Efron and Tibshirani (1993) and Efron (1987). The interval, denoted $I_{bca}$, is given by:

$$I_{bca} = \left[\hat{G}^{-1}\left(\tilde{\alpha}_1\right); \hat{G}^{-1}\left(\tilde{\alpha}_2\right)\right] ,$$

where

$$\tilde{\alpha}_1 = \Phi\left(\frac{(z_0 - z\left(\frac{\alpha}{2}\right))}{[1 + a(z\left(\frac{\alpha}{2}\right) - z_0)]} + z_0\right),$$

and

$$\tilde{\alpha}_2 = \Phi\left(\frac{(z_0 + z\left(\frac{\alpha}{2}\right))}{[1 - a(z\left(\frac{\alpha}{2}\right) + z_0)]} + z_0\right),$$

The parameter $a$ can be estimated by $\hat{a}$ defined as:

$$\hat{a} = \frac{1}{6} \cdot \frac{\sum_{i=1}^{n} K_i^3}{\{\sum_{i=1}^{n} K_i^2\}^{\frac{3}{2}}} . \qquad (2.6)$$

One possible choice of $K_i$ is the Jackknife influence function of the original statistic $\hat{\theta}_n$, i.e.,

$$K_i = (n-1)(\hat{\theta}_{n,[\bullet]} - \hat{\theta}_{n-1,[i]}), \qquad \forall i = 1, 2, \ldots n ,$$

where $\hat{\theta}_{n-1,[i]} = \hat{\theta}_{n-1}(X_1, X_2, \ldots X_{i-1}, X_{i+1}, \ldots X_n)$, $\hat{\theta}_{n-1,[i]}$ is calculated from the original sample data with the $i^{\text{th}}$ element "deleted", and $\hat{\theta}_{n,[\bullet]} = \frac{1}{n}\sum_{i=1}^{n} \hat{\theta}_{n-1,[i]}$ (Efron 1992). Other choices for $K_i$ can be found in the literature (Efron 1987).

Once again the interval can be estimated using the basic percentile Monte-Carlo algorithm described in the previous sections with the exception that the last step is replaced with the following:

**Approximating the $BC_a$ percentile confidence interval through simulation**

5. The interval is then:

$$I_{bca,B} = \left[\hat{\theta}^*_{n,(r)}; \hat{\theta}^*_{n,(s)}\right] ,$$

where

$$r = \lfloor (B+1) \cdot \tilde{\alpha}_1 \rfloor ,$$

$$s = \lfloor (B+1) \cdot \tilde{\alpha}_2 \rfloor .$$

$$\tilde{\alpha}_1 = \Phi\left(\frac{(\hat{z}_0 - z\left(\frac{\alpha}{2}\right))}{[1 + \hat{a}(z\left(\frac{\alpha}{2}\right) - \hat{z}_0)]} + \hat{z}_0\right),$$

and

$$\tilde{\alpha}_2 = \Phi\left(\frac{(\hat{z}_0 + z\left(\frac{\alpha}{2}\right))}{[1 - \hat{a}(z\left(\frac{\alpha}{2}\right) + \hat{z}_0)]} + \hat{z}_0\right),$$

where $\hat{z}_0 = \Phi^{-1}\left(\hat{G}\left(\hat{\theta}_n\right)\right)$, and the expression $\hat{G}(\hat{\theta}_n)$ appearing in the definition of $\hat{z}_0$, can be approximated by:

$$\hat{G}(\hat{\theta}_n) = \frac{1}{B}\sum_{b=1}^{B} \mathrm{I}(\hat{\theta}^*_{n,b} \leq \hat{\theta}_n) .$$

Obtain $\hat{a}$ by performing the calculation as described in equation (2.6).

For additional information on this topic see Davison and Hinkley (1997).

### The bootstrap-$t$ interval

The bootstrap-$t$ interval is related to the Student-$t$ interval for the sample mean,

$$\left[\bar{X}_n - t_{n-1}\left(1 - \frac{\alpha}{2}\right) S_n/\sqrt{n}; \bar{X}_n - t_{n-1}\left(\frac{\alpha}{2}\right) S_n/\sqrt{n}\right],$$

where $S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$ and $t_{n-1}(\cdot)$ is the quantile function for the Student $t$-distribution with $n-1$ degrees of freedom. This interval works very well for the sample mean and when the underlying data are normal. The bootstrap-$t$ interval attempts to mimic the above interval by replacing the sample mean with any estimator, the $t$ quantile function with a critical value obtained from the bootstrap (or bootstrap critical value) and the standard error with a bootstrap estimate of the standard error.

Let $\theta = t(F)$ be some parameter of interest. Let $\hat{\theta}_n = t(F_n)$ be the plug-in estimate of $\theta$, and let $\hat{\sigma}_n = \hat{\mathrm{SE}}(\hat{\theta}_n)$ be the estimated standard error of $\hat{\theta}_n$. Through Studentizing one can then construct a $100(1-\alpha)\%$ confidence interval for $\theta$ by first considering the following quantity:

$$Z = \frac{\hat{\theta}_n - \theta}{\hat{\mathrm{SE}}(\hat{\theta}_n)}, \tag{2.7}$$

where the quantity $\hat{\mathrm{SE}}(\hat{\theta}_n)$ is possibly obtained by making use of the bootstrap. Applying the plug-in principle to the Studentized statistic (2.7), the following expression is obtained:

$$Z^* = \frac{\hat{\theta}^*_n - \hat{\theta}_n}{\hat{\mathrm{SE}}^*(\hat{\theta}^*_n)},$$

where $\widehat{\mathrm{SE}}^*(\hat{\theta}_n^*)$ is the standard error of $\hat{\theta}_n^*$. In other words, the bootstrap statistic is centered around the parameter estimate $\hat{\theta}_n$ and then divided by its own (estimated) standard error, $\widehat{\mathrm{SE}}^*(\hat{\theta}_n^*)$. (Note that when estimating this statistic in a bootstrap simulation it may be necessary to use the *double* bootstrap to calculate the value of the standard error, $\hat{\sigma}_n^* = \widehat{\mathrm{SE}}^*(\hat{\theta}_n^*)$).

Let $\hat{H}(x)$ be defined as the distribution function of $Z^*$, i.e.,

$$\hat{H}(x) = \mathrm{P}^*(Z^* \leq x).$$

Once we have this Studentized statistic, we need to find the value $\hat{t}(\alpha)$ such that it satisfies the following expression:

$$\hat{H}(\hat{t}(\alpha)) = \mathrm{P}^*(Z^* \leq \hat{t}(\alpha)) = \alpha,$$

that is, we define $\hat{t}(\alpha)$

$$\hat{t}(\alpha) = \hat{H}^{-1}(\alpha).$$

The $100(1-\alpha)\%$ bootstrap-$t$ equal-tailed, two-sided confidence interval for $\theta$, denoted $I_t$ (not to be confused with the notation for the indicator function), is given by:

$$I_t = \left[\hat{\theta}_n - \hat{H}^{-1}\left(1 - \frac{\alpha}{2}\right) \cdot \hat{\sigma}_n ; \hat{\theta}_n - \hat{H}^{-1}\left(\frac{\alpha}{2}\right) \cdot \hat{\sigma}_n\right] ,$$

or

$$I_t = \left[\hat{\theta}_n - \hat{t}\left(1 - \frac{\alpha}{2}\right) \cdot \hat{\sigma}_n ; \hat{\theta}_n - \hat{t}\left(\frac{\alpha}{2}\right) \cdot \hat{\sigma}_n\right] ,$$

where $\hat{\sigma}_n$ is the estimated standard error of $\hat{\theta}_n$ (possibly estimated using the bootstrap). The algorithm for approximating this bootstrap-$t$ interval is provided next.

**Approximating the bootstrap-t confidence interval**

1. Calculate the statistics $\hat{\theta}_n = \hat{\theta}_n(X_1, X_2, \ldots, X_n)$ and $\hat{\sigma}_n = \widehat{\mathrm{SE}}(\hat{\theta}_n) = \sqrt{\widehat{\mathrm{Var}}(\hat{\theta}_n)}$, where $\hat{\sigma}_n$ could be obtained by using some non-parametric procedure like the bootstrap if a closed form is not available.

2. Generate $X_1^*, X_2^*, \ldots, X_n^*$ from the EDF, $F_n$, i.e., generate $X_1^*, X_2^*, \ldots, X_n^*$ by sampling with replacement from $X_1, X_2, \ldots, X_n$.

3. Calculate $\hat{\theta}_n^* = \hat{\theta}_n(X_1^*, X_2^*, \ldots, X_n^*)$.

4. Repeat steps 2 and 3 $B$ times obtaining $\hat{\theta}_{n,1}^*, \hat{\theta}_{n,2}^* \ldots \hat{\theta}_{n,B}^*$.

5. Calculate

$$Z_b^* = \frac{\hat{\theta}_{n,b}^* - \hat{\theta}_n}{\hat{\sigma}_{n,b}^*}, \quad b = 1, 2, \ldots B.$$

6. From $Z_1^*, Z_2^*, \ldots, Z_B^*$, obtain the order statistics $Z_{(1)}^* \leq Z_{(2)}^* \leq \cdots \leq Z_{(B)}^*$.

7. The interval is then:

$$I_{t,B} = \left[\hat{\theta}_n - Z_{(r)}^* \cdot \hat{\sigma}_n ; \hat{\theta}_n - Z_{(s)}^* \cdot \hat{\sigma}_n\right] ,$$

where

$$r = \left\lfloor (B+1) \cdot \left(1 - \frac{\alpha}{2}\right) \right\rfloor \quad \text{and} \quad s = \left\lfloor (B+1) \cdot \left(\frac{\alpha}{2}\right) \right\rfloor .$$

### The hybrid percentile interval

The hybrid percentile method is a mixture of the ideas underlying the bootstrap-$t$ method and the basic percentile method.

Define the distribution of $\hat{\theta}_n - \theta$ as $K(x) = \mathrm{P}(\hat{\theta}_n - \theta \leq x)$. To determine this interval one must first find the quantiles $c_{(\frac{\alpha}{2})}$ and $d_{(1-\frac{\alpha}{2})}$ defined as $K(c_{(\frac{\alpha}{2})}) = \mathrm{P}(\hat{\theta}_n - \theta \leq c_{(\frac{\alpha}{2})}) = \frac{\alpha}{2}$ and $K(d_{(1-\frac{\alpha}{2})}) = \mathrm{P}(\hat{\theta}_n - \theta \leq d_{(1-\frac{\alpha}{2})}) = 1 - \frac{\alpha}{2}$. This means that

$$\mathrm{P}\left(c_{(\frac{\alpha}{2})} \leq \hat{\theta}_n - \theta \leq d_{(1-\frac{\alpha}{2})}\right) = 1 - \alpha$$

$$\Rightarrow \mathrm{P}\left(\hat{\theta}_n - d_{(1-\frac{\alpha}{2})} \leq \theta \leq \hat{\theta}_n - c_{(\frac{\alpha}{2})}\right) = 1 - \alpha.$$

This in turn means that the interval for $\theta$ will have the form

$$\left[\hat{\theta}_n - d_{(1-\frac{\alpha}{2})}; \hat{\theta}_n - c_{(\frac{\alpha}{2})}\right]. \tag{2.8}$$

However, the values $c_{(\frac{\alpha}{2})}$ and $d_{(1-\frac{\alpha}{2})}$ are unknown (since $F$ is unknown) and are estimated by $\hat{c}_{(\frac{\alpha}{2})}$ and $\hat{d}_{(1-\frac{\alpha}{2})}$, given by:

$$\hat{c}_{(\frac{\alpha}{2})} = \hat{G}^{-1}\left(\frac{\alpha}{2}\right) - \hat{\theta}_n, \tag{2.9}$$

and

$$\hat{d}_{(1-\frac{\alpha}{2})} = \hat{G}^{-1}\left(1 - \frac{\alpha}{2}\right) - \hat{\theta}_n. \tag{2.10}$$

This leads to the estimated interval:

$$I_h = \left[\hat{\theta}_n - \hat{d}_{(1-\frac{\alpha}{2})}; \hat{\theta}_n - \hat{c}_{(\frac{\alpha}{2})}\right]. \tag{2.11}$$

Therefore if the estimated values of $c$ and $d$ found in equations (2.9) and (2.10) are substituted into the interval (2.11), the following interval is obtained:

$$I_h = \left[\hat{\theta}_n - \left(\hat{G}^{-1}\left(1 - \frac{\alpha}{2}\right) - \hat{\theta}_n\right) \quad ; \quad \hat{\theta}_n - \left(\hat{G}^{-1}\left(\frac{\alpha}{2}\right) - \hat{\theta}_n\right)\right]$$
$$I_h = \left[2\hat{\theta}_n - \hat{G}^{-1}\left(1 - \frac{\alpha}{2}\right) \quad ; \quad 2\hat{\theta}_n - \hat{G}^{-1}\left(\frac{\alpha}{2}\right)\right].$$

It is clear from this derivation that the term 'Hybrid' stems from the fact that this interval is a mixture (or hybrid) of the bootstrap-$t$ and basic percentile method of bootstrap confidence interval estimation. This method uses a similar standardization technique used in the bootstrap-$t$ interval (the parameter is subtracted, but it is not divided by the standard error) and it also uses the simple percentile ideas found in the basic percentile interval.

A simple Monte-Carlo algorithm for approximating this interval can now be stated.

**Approximating the hybrid percentile confidence interval using the bootstrap**

1. Calculate the statistic $\hat{\theta}_n = \hat{\theta}_n(X_1, X_2, \ldots, X_n)$ from the observed sample $X_1, X_2, \ldots, X_n$.
2. Generate $X_1^*, X_2^*, \ldots, X_n^*$ from the EDF, $F_n$, i.e., generate $X_1^*, X_2^*, \ldots, X_n^*$ by sampling with replacement from $X_1, X_2, \ldots, X_n$.
3. Calculate $\hat{\theta}_n^* = \hat{\theta}_n(X_1^*, X_2^*, \ldots, X_n^*)$.
4. Repeat steps 2 and 3 $B$ times obtaining $\hat{\theta}^*_{n,1}, \hat{\theta}^*_{n,2} \ldots \hat{\theta}^*_{n,B}$.
5. Obtain the order statistics $\hat{\theta}^*_{n,(1)} \leq \hat{\theta}^*_{n,(2)} \leq \cdots \leq \hat{\theta}^*_{n,(B)}$.
6. The interval is then:
$$I_{h,B} = \left[2\hat{\theta}_n - \hat{\theta}^*_{n,(r)}; 2\hat{\theta}_n - \hat{\theta}^*_{n,(s)}\right] ,$$
where
$$r = \left\lfloor (B+1)\cdot\left(1 - \frac{\alpha}{2}\right)\right\rfloor \quad \text{and} \quad s = \left\lfloor (B+1)\cdot\left(\frac{\alpha}{2}\right)\right\rfloor .$$

### 2.4.6 Bootstrap hypothesis testing

It is also possible to employ the bootstrap to test hypotheses regarding population parameters. However, some care should be taken to apply the bootstrap in these situations since the bootstrap can fail when the method is applied naively (see Sakov (1998)).

Define the hypothesis statement for the set of parameters $\boldsymbol{\tau}$ of the distribution $F_{\boldsymbol{\mathcal{T}}}$ in general as:

$$H_0 : \boldsymbol{\tau} \in \mathfrak{T}_0 \qquad \text{vs.} \qquad H_A : \boldsymbol{\tau} \in \mathfrak{T}_A,$$

where $\mathfrak{T}_0$ and $\mathfrak{T}_A$ are two disjoint subsets of some parameter space $\mathfrak{T} = \mathfrak{T}_0 \cup \mathfrak{T}_A$. The distribution function that satisfies the specified null hypothesis is denoted by $F_{\boldsymbol{\mathcal{T}}_0}$.

The bootstrap procedure for testing hypotheses of this form involves a resampling scheme that attempts to estimate the distribution $F_{\boldsymbol{\mathcal{T}}_0}$, and not $F_{\boldsymbol{\mathcal{T}}}$. In other words, the resampling scheme should try and enforce the condition stated in the null hypothesis in the bootstrap world.

Naturally, the relationship between hypotheses of this form and confidence intervals are well known, and one might believe that a separate technique for dealing with hypothesis tests is redundant if there already exist techniques for the construction of confidence intervals. However, this type of thinking is contested in Shao and Tu (1995), where they state that the construction of confidence intervals, as a proxy for a hypothesis test, is "impossible in some cases", and also that the hypothesis test technique is better since "they usually take account of the special nature of the hypothesis." In addition to these reasons, hypothesis testing techniques are attractive because they allow us to calculate $p$-values and critical values for the tests; something which is not possible if one follows the confidence interval route. The virtues of enforcing $H_0$ in the bootstrap world rule are discussed in, among others, Efron and Tibshirani (1993), Davison and Hinkley (1997), Hall and Wilson (1991), Young (1986), Beran (1988), Fisher and Hall (1990), Westfall and Young (1993), Martin (2007) and Allison (2008).

To clarify, suppose that the relevant statistic for testing some hypothesis is $T_n(\boldsymbol{X}_n)$ and the observed value of the statistic is $t_n = T_n(\mathbf{x}_n)$ then a $p$-value, in the real world, would be defined as

$$p = \mathrm{P}(T_n(\boldsymbol{X}_n) \geq t_n | H_0). \tag{2.12}$$

The application of the bootstrap to estimate this $p$-value then simply involves applying the plug-in principle to (2.12) so that the bootstrap $p$-value is given by

$$p_{boot} = \mathrm{P}^*(T_n(\boldsymbol{W}_n^{0*}) \geq t_n), \tag{2.13}$$

where $\boldsymbol{W}_n^0 = \{W_1^0, W_2^0, \ldots, W_n^0\}$ is the transformed sample data modified to reflect the condition stipulated under the null hypothesis, $F_{n,0}$ is the EDF of $\boldsymbol{W}_n^0$, and $\boldsymbol{W}_n^{0*} = \{W_1^{0*}, W_2^{0*}, \ldots, W_n^{0*}\}$ is the bootstrap sample created by independently sampling from $F_{n,0}$.

The Monte Carlo algorithm for approximating $p_{boot}$ is then as follows:

**Approximating bootstrap $p$-values:**

1. Given data $X_1, X_2, \ldots, X_n$. Calculate the statistic $T_n = T_n(X_1, X_2, \ldots, X_n)$.
2. Apply any necessary modifications to $X_1, X_2, \ldots, X_n$ so that the distributional properties in the bootstrap world comply with the stated null hypothesis, $H_0$. We now have $W_1^0, W_2^0, \ldots, W_n^0$.
3. Sample with replacement from $W_1^0, W_2^0, \ldots, W_n^0$ to get $W_1^{0*}, W_2^{0*}, \ldots, W_n^{0*}$ .
4. Calculate the statistics $T_n^* = T_n(W_1^{0*}, W_2^{0*}, \ldots, W_n^{0*})$.
5. Repeat step 4 $B$ times to obtain $T_{n,1}^*, T_{n,2}^*, \ldots, T_{n,B}^*$.
6. Approximate the quantity $p_{boot}$ with:

$$p_{boot,B} = \frac{1}{B}\sum_{i=1}^{B} I(T_{n,i}^* \geq t_n),$$

where $I(\cdot)$ is the indicator function.

**Remark:**

One can also easily determine a bootstrap critical value associated with a nominal significance level $\alpha$, by ordering the bootstrap statistics obtained in step 5, denoting them by $T_{n,(1)}^* \leq T_{n,(2)}^* \leq \cdots \leq T_{n,(B)}^*$, and then define a right one-sided critical value as $T_{n,(\lfloor(1-\alpha)B\rfloor)}^*$.

Bootstrap hypothesis testing will be discussed in greater detail in Chapter 6.

# Chapter 3

# Bootstrap consistency

## 3.1 Introduction

In the cases discussed up to this point we assumed that the bootstrap will return correct results in the sense that the results can be used as estimators for unknown quantities in the real world. Unfortunately, this is not always the case since there are a number of situations where the bootstrap can fail to produce the desired results; these situations are called the *non-regular cases* by Shao and Tu (1995). One of the causes of this type of inconsistency is that, when the plug-in principle is applied, the bootstrap world quantities do not reflect the properties of their real world counterparts. The numerous remedies for these situations often depend on the type of failure which occurs and usually involve some sort of modification of the bootstrap sampling scheme or the bootstrap statistic so that the modified versions mimic the real world quantities better.

These non-regular cases give one pause to think before applying the bootstrap, and not simply treat it as an 'apply-and-forget' type of method. Some thought must be given to the data, the statistic, and the nature of the problem before applying it. While the situations where the bootstrap fails are fairly rare in practice, diagnostics for determining whether a potential failure can occur are still limited (Beran 1997, Bickel and Sakov 1999).

## 3.2 Consistency

Non-regular cases lead to bootstrap failure because the consistency of the bootstrap estimators breaks down in these cases. The concept of consistency, as given by Shao and Tu (1995), can be stated as follows:

First, let $X_1, X_2, \ldots, X_n$ be $d$-dimensional random observations from some unknown probability distribution $F$, and $T_n = T_n(X_1, X_2, \ldots, X_n \; ; \; F)$ be a $s$-dimensional statistic constructed from $X_1, X_2, \ldots, X_n$ and $F$. Define the true sampling distribution of $T_n$ as

$$G_n(x) = \mathrm{P}(T_n \leq x | F).$$

Denote the estimated distribution function by $\hat{F}$. Let $T_n^* = T_n(X_1^*, X_2^*, \ldots, X_n^* \; ; \; \hat{F})$ denote the bootstrap version of $T_n$, where $X_1^*, X_2^*, \ldots, X_n^*$ is a bootstrap sample obtained from $X_1, X_2, \ldots, X_n$. We now estimate $G_n$ with

$$\hat{G}_n(x) = \mathrm{P}(T_n^* \leq x | \hat{F}) = \mathrm{P}^*(T_n^* \leq x).$$

Finally, the definition of consistency can be given by:
*Let $\rho$ be a metric on $\mathscr{F}_{\mathbb{R}^s} = \{$ all distributions on $\mathbb{R}^s\}$.*

- ⋆ *$\hat{G}_n$ is weakly consistent if $\rho(\hat{G}_n, G_n) \xrightarrow{p} 0$ as $n \to \infty$.*
- ⋆ *$\hat{G}_n$ is strongly consistent if $\rho(\hat{G}_n, G_n) \xrightarrow{a.s.} 0$ as $n \to \infty$.*

A popular choice of the metric $\rho$ is $\rho_\infty$, the distance generated by the sup-norm, e.g., $\rho_\infty(F, H) = \sup_{x\in\mathbb{R}} |F(x) - H(x)|$.

This definition of consistency will be used in the description of the non-regular cases in order to show when $\hat{G}_n$ is a consistent estimator of the true sampling distribution $G_n$.

### 3.2.1 Consistency and the bootstrap

Different situations will be mentioned where the bootstrap produces consistent results, but, to emphasize these situations, counter-situations (or non-regular cases) will also be mentioned. Where possible, examples of the consistency (or inconsistency) will be provided. Some remedies for the consistency of the non-regular cases will also be mentioned, but in general these remedies will be deferred to the next chapter where a modification to the bootstrap is used.

Non-regular cases, which lead to bootstrap inconsistency, occur for a number of reasons. Shao and Tu (1995) broadly define three conditions that can help identify non-regular cases:

- ⋆ The bootstrap is sensitive to the 'heavy tailed' behaviour of the population distribution $F$, for example if $F$ has infinite variance. Bootstrap consistency in these cases requires stronger moment conditions than those typically required for standard theoretical limiting distributions.
- ⋆ The lack of certain smoothness conditions of the statistics used in the bootstrap can influence whether or not the bootstrap will be consistent.
- ⋆ The behaviour of the bootstrap estimator sometimes depends on the method used to obtain the bootstrap sample data.

## 3.3 $U$-statistics and bootstrap consistency

### 3.3.1 $U$-statistics

Many statistics fall within a general class of statistics which can be expressed as a generalization of the sample mean; these statistics are called $U$-statistics (Serfling 1980). These $U$-statistics have been shown to have a number of desirable properties including asymptotic normality and the fact that minimum variance unbiased estimators of the central moments can be expressed as $U$-statistics (see Hoeffding (1948)).

These $U$-statistics are defined (see for example Serfling (1980) and Janssen (1997)) as follows: Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables from some unknown distribution function $F$. Define some measurable function $h$, symmetric in its $p$ arguments, which maps $\mathbb{R}^p$ onto $\mathbb{R}$, i.e.,

$$h : \mathbb{R}^p \to \mathbb{R}.$$

Note that the function $h$ with $p$ arguments is often referred to as the *kernel* of degree $p$ of the $U$-statistic. Now define a subset of the ordered indices $\{1, 2, \ldots, n\}$ to be

$$C_{np} = \{(i_1, i_2, \ldots, i_p) \in \mathbb{N}^p : 1 \leq i_1 < i_2 < \cdots < i_p \leq n\}.$$

The $U$-statistic is then defined as the 'average' of the values $h_{C_{np}} = h(X_{i_1}, X_{i_2}, \ldots, X_{i_p})$, where $(i_1, i_2, \ldots, i_p) \in C_{np}$, i.e., the $U$-statistic is defined as

$$U_n = \binom{n}{p}^{-1} \sum_{C_{np}} h(X_{i_1}, X_{i_2}, \ldots, X_{i_p}).$$

The notation $\sum_{C_{np}}$ refers to the summation over the $\binom{n}{p}$ combinations of the $p$ distinct ordered elements $\{i_1, i_2, \ldots, i_p\}$ from the set of indices $\{1, 2, \ldots, n\}$.

The $U$-statistic is then an estimator for the parameter $\theta = \theta(F)$, which is a regular functional which maps, $\mathcal{L}_0$, a subset of $\mathcal{L}$ (i.e., the set of all univariate distributions), onto $\mathbb{R}$, i.e.,

$$\theta : \mathcal{L}_0 \rightarrow \mathbb{R},$$

and where

$$\begin{aligned} \theta(F) &= \mathrm{E}(h(X_1, X_2, \ldots, X_p)) \\ &= \underset{\mathbb{R}^p}{\int \cdots \int} h(x_1, x_2, \ldots, x_p) dF(x_1) dF(x_2) \ldots dF(x_p). \end{aligned}$$

We assume that $\mathrm{E}(|h(X_1, X_2, \ldots, X_p)|)$ is finite.

### Examples of $U$-statistics

Some examples of statistics that can be expressed as $U$-statistics include the $k^{\text{th}}$ sample moment and the sample variance.

**The $k^{\text{th}}$ sample moment** : The $k^{\text{th}}$ moment, $\mu_k$, is a regular functional of $F$, since

$$\mu_k = \theta(F) = \int_{\mathbb{R}} x^k dF(x).$$

The $U$-statistic created with a kernel of degree, one defined as $h(x) = x^k$, is then given by

$$\begin{aligned} U_n &= \binom{n}{1}^{-1} \sum_{C_{n1} = \{1,2,\ldots,n\}} h(X_i) \\ &= \frac{1}{n} \sum_{i=1}^{n} X_i^k, \end{aligned}$$

which is the ordinary $k^{\text{th}}$ sample moment.

**The variance** : The variance, $\sigma^2$, as with the $k^{\text{th}}$ moment described above, is a regular functional of $F$, as can be seen by the following expression,

$$\begin{aligned} \sigma^2 = \theta(F) &= \int_{\mathbb{R}} (x - \mu)^2 dF(x) \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} \frac{(x - y)^2}{2} dF(x) dF(y). \end{aligned}$$

In this case a kernel of degree two, defined as $h(x, y) = (x - y)^2/2$, will be used. As a result, the $U$-statistic estimator is given by

$$\begin{aligned} U_n &= \binom{n}{2}^{-1} \sum_{C_{n2}=\{1\leq i<j\leq n\}} h(X_i, X_j) \\ &= \frac{2}{n(n-1)} \sum_{1\leq i<j\leq n} \frac{(X_i - X_j)^2}{2} \\ &= \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 = S^2_{n-1}, \end{aligned}$$

which is the unbiased estimate of variance.

### 3.3.2 Hoeffding decompositions

In this section we will restrict ourselves to $U$-statistics with kernels of, at most, degree two. An important bootstrap failure is one that occurs with *degenerate* or *pure* $U$-statistics. In order to define these degenerate $U$-statistics it is easiest to first consider the Hoeffding decomposition of the $U$-statistic (see for example Lee (1990), Dehling and Mikosch (1994), Janssen (1997)).

The Hoeffding decomposition defines a new statistic $\hat{U}_n$ as a sequence of other functions such that it has the same asymptotic characteristics as $U_n$.

This decomposition of the $U$-statistic $U_n$ is defined as

$$\hat{U}_n = \sum_{i=1}^{n} \mathrm{E}(U_n | X_i) - (n-1)\theta(F),$$

or equivalently,

$$\hat{U}_n - \theta(F) = \frac{2}{n} \sum_{i=1}^{n} g_F(X_i),$$

where

$$g_F(x) = \int_{\mathbb{R}} h(x, y) dF(y) - \theta(F).$$

We can then define a degenerate $U$-statistic as follows:
The $U$-statistic $U_n$ is a degenerate (or pure) $U$-statistic if $g_F(x) = 0$ for all $x$. If we define $\sigma_g^2 := \mathrm{Var}(g_F(X_i))$, then an equivalent definition is to say that a $U$-statistic is degenerate (or pure) if $\sigma_g^2 = 0$ (Lee 1990, Janssen 1997).

### 3.3.3 Asymptotics of degenerate and non-degenerate $U$-statistics

The asymptotic distributions of degenerate and non-degenerate $U$-statistics have been discussed at length in Serfling (1980). If we limit our discussion to results pertaining to $U$-statistics with kernels of degree two, then these statistics have the following limiting distributions:

⋆ **Non-degenerate $U$-statistics:** If $\sigma_g^2 = \mathrm{Var}(g_F(X_1)) > 0$ and $\mathrm{E}(h^2(X_1, X_2)) < \infty$, then

$$\frac{\sqrt{n}(U_n - \theta(F))}{2\sigma_g} \xrightarrow{\mathcal{D}} Z \sim \mathrm{N}(0, 1),$$

i.e., the standardized statistic has a limiting standard normal distribution.

⋆ **Degenerate $U$-statistics:** If $\sigma_g^2 = \mathrm{Var}(g_F(X_1)) = 0$ and $\mathrm{E}(h^2(X_1, X_2)) < \infty$, then

$$n(U_n - \theta(F)) \xrightarrow{\mathcal{D}} Y,$$

where $Y = \sum_{i=1}^{\infty} \lambda_i(\chi_{1,i}^2 - 1)$, $\chi_{1,i}^2$, $i = 1, 2, \ldots,$ are i.i.d. chi-squared random variables with one degree of freedom each, and $\lambda_1, \lambda_2, \ldots,$ are the eigenvalues defined in, for example, Serfling (1980), pp. 193 – 194. In other words, the degenerate $U$-statistic does not have an asymptotic normal distribution, but rather an asymptotic distribution related to the infinite sum of independent chi-squared random variables.

### 3.3.4 Bootstrapping $U$-statistics

In this section we will look at some successes and failures of the bootstrap applied to $U$-statistics, but we must first establish some new notation before we proceed. Let the distribution of the quantity $\sqrt{n}(U_n - \theta(F))$ be given by

$$F_{U_n}(x) = \mathrm{P}(\sqrt{n}(U_n - \theta(F)) \leq x).$$

The bootstrap estimate of this quantity will then be given by

$$F_{U_n^*}(x) = \mathrm{P}^*(\sqrt{n}(U_n^* - \theta(F_n)) \leq x), \tag{3.1}$$

where

$$U_n^* = \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} h(X_i^*, X_j^*),$$

and

$$\begin{aligned} \theta(F_n) &= \mathrm{E}^*(U_n^*) \\ &= \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} h(X_i, X_j). \end{aligned}$$

Note that since the terms $\theta(F_n)$ and $U_n$ are asymptotically equivalent, one can replace the term $(U_n^* - \theta(F_n))$ in equation (3.1) with $(U_n^* - U_n)$.

#### Bootstrapping non-degenerate $U$-statistics

It has been shown in Bickel and Freedman (1981) and Shao and Tu (1995) that, under the same conditions required for the asymptotic normality of the $U$-statistics in the real world, the bootstrap 'fails', i.e., it produces an inconsistent result. However, it was shown (also in Bickel and Freedman (1981) and Shao and Tu (1995)) that the introduction of an additional assumption rescues the bootstrap in this scenario, viz.

$$\mathrm{E}(|h(X_1, X_1)|) < \infty.$$

The consistency of the bootstrap for non-degenerate $U$-statistics can then be stated using the following theorem stated in Shao and Tu (1995):

**Theorem**

*Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables and let $U_n$ be a $U$-statistic with kernel of degree two, as defined previously. Let $F_{U_n}(x) = \mathrm{P}(\sqrt{n}(U_n - \theta(F)) \leq x)$, and let $F_{U_n^*}(x) = \mathrm{P}^*(\sqrt{n}(U_n^* - \theta(F_n)) \leq x)$. If $\mathrm{E}(h(X_1, X_2)^2) < \infty$, $\mathrm{E}(|h(X_1, X_1)|) < \infty$ and $\sigma_g^2 = \mathrm{Var}(g_F(X_i)) > 0$, then*

$$\mathrm{P}\left(\sup_{x \in \mathbb{R}} \left|F_{U_n^*}(x) - F_{U_n}(x)\right| \underset{\infty}{\overset{n}{\to}} 0\right) = 1$$

*or*

$$\rho_\infty(F_{U_n^*}, F_{U_n}) \xrightarrow{a.s.} 0.$$

**Bootstrapping degenerate $U$-statistics**

Blind application of the plug-in principle to degenerate $U$-statistics is the downfall of the consistency of the bootstrap in this case. This is because the property which defines a degenerate $U$-statistic, i.e., $g_F(x) = 0$ for all $x$, is not mimicked in the bootstrap world when the plug-in principle is applied naively (see Arcones and Giné (1992), Janssen (1997) and, for an alternative perspective, see Mammen (1992)).

In order to overcome this inconsistency it is necessary to force this property in the bootstrap world by redefining the $U$-statistic kernel function. From Arcones and Giné (1992) and Janssen (1997) and for $U$-statistics with kernels of degree two, we redefine the kernel in the bootstrap world as

$$\bar{h}_{F_n}(x,y) = h(x,y) - g_{F_n}(x) - g_{F_n}(y) - \theta(F_n).$$

The bootstrap version of the $U$-statistic then becomes

$$U_n^*(\bar{h}_{F_n}) = \binom{n}{2}^{-1} \sum_{1\le i<j\le n} \bar{h}_{F_n}(X_i^*, X_j^*).$$

If this corrected statistic is used in conjunction with the additional assumption used for bootstrapping non-degenerate $U$-statistics, i.e., $\mathrm{E}(|h(X_1,X_1)|) < \infty$, then the inconsistency is remedied. The consistency of the bootstrap for degenerate $U$-statistics can then be expressed using the following theorem (Arcones and Giné 1992, Dehling and Mikosch 1994).

**Theorem**

*Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables and let $U_n$ be a $U$-statistic with kernel of degree two, as defined previously. Let $F_{U_n,n}(x) = \mathrm{P}(n(U_n - \theta(F)) \le x)$, and let $F_{U_n^*,n}(x) = \mathrm{P}^*(nU_n^*(\bar{h}_{F_n}) \le x)$. If $\mathrm{E}(h(X_1,X_2)^2) < \infty$, $\mathrm{E}(|h(X_1,X_1)|) < \infty$ and $\sigma_g^2 = \mathrm{Var}(g_F(X_i)) = 0$, then*

$$\mathrm{P}\left(\sup_{x\in\mathbb{R}} \left|F_{U_n^*,n}(x) - F_{U_n,n}(x)\right| \overset{n}{\underset{\infty}{\to}} 0\right) = 1$$

*or*

$$\rho_\infty(F_{U_n^*,n}, F_{U_n,n}) \overset{a.s.}{\to} 0.$$

## 3.4 Super-efficient estimators and bootstrap consistency

It has been found that the bootstrap is inconsistent when applied to super-efficient estimators (see for example Beran (1997), Polansky (1999)). In this section we will discuss the specific case where the bootstrap fails when it is applied to Hodges' super-efficient estimator.

### 3.4.1 Super-efficient estimators

Super-efficient estimators are estimators whose asymptotic efficiency exceeds 1. The asymptotic efficiency of an estimator, $\hat{\theta}_n$, which estimates the parameter $\theta$, is defined next.
Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables from some distribution function $F$ characterized by a real valued parameter $\theta$, then the asymptotic efficiency is

$$\phi(\theta) = \frac{1/I(\theta)}{v(\theta)},$$

where $I(\theta)$ is the *Fisher Information* of the statistic defined as

$$I(\theta) = -\mathrm{E}(L''(\theta, X_1)),$$

$v(\theta)/n$ is the asymptotic variance of $\hat{\theta}_n$ and $L(\cdot)$ is the likelihood function.

Typically all estimators have the property that $\phi(\theta) \leq 1$. However, it is possible to construct an estimator such that $\phi(\theta) > 1$ (Kotze, Johnson and Read 1983).

### Hodges' super-efficient estimator

A commonly used example of a super-efficient estimator for the population mean, $\mu$, is Hodges' super-efficient estimator defined as follows:
Let $X_1, X_2, \ldots, X_n$ be i.i.d. $\mathrm{N}(\mu, 1)$ random variables, then

$$\hat{\theta}_n = \begin{cases} \bar{X}_n, & |\bar{X}_n| > n^{-\frac{1}{4}} \\ \alpha\bar{X}_n, & |\bar{X}_n| \leq n^{-\frac{1}{4}} \end{cases} \qquad 0 < \alpha < 1,$$

where $\bar{X}_n = n^{-1}\sum_{i=1}^{n} X_i$. The statistic $\hat{\theta}_n$ is called *Hodges' super-efficient estimator.*

When $\mu \neq 0$ then the asymptotic efficiency of this estimator is $\phi(\mu) = 1$. However, when $\mu = 0$ the asymptotic efficiency exceeds 1, i.e., $\phi(\mu) > 1$.

Beran (1997) notes that the distribution function $H(x)$, defined as

$$H(x) = \mathrm{P}\left(\sqrt{n}(\hat{\theta}_n - \mu) \leq x\right),$$

converges weakly to $\Phi(x)$ if $\mu \neq 0$ and converges weakly to $\Phi(x/\alpha)$ if $\mu = 0$, where $\Phi(\cdot)$ is the standard normal distribution function. In other words,

$$n^{\frac{1}{2}}(\hat{\theta}_n - \mu) \xrightarrow{\mathcal{D}} \begin{cases} Z \sim \mathrm{N}(0,1), & \mu \neq 0 \\ Y \sim \mathrm{N}(0,\alpha^2), & \mu = 0 \end{cases}.$$

### 3.4.2 Bootstrapping super-efficient estimators

When considering Hodges' super-efficient estimator, the bootstrap estimate of the distribution $H(x)$ is $H_n(x)$, defined as

$$H_n(x) = \mathrm{P}^*\left(n^{\frac{1}{2}}\left(\hat{\theta}_n^* - \bar{X}_n\right) \leq x\right).$$

When $\mu \neq 0$, then the limiting distribution of $H_n$ and $H$ are the same, i.e., $H_n$ converges weakly in probability to $\mathrm{N}(0, 1)$. Unfortunately, when $\mu = 0$ one finds that it converges "in distribution, as a random element of the space of all probability measures on the real line metrized by weak convergence, to the random probability measure $\mathrm{N}((\alpha-1)Z, \alpha^2)$" (Beran 1997), where $Z \sim \mathrm{N}(0, 1)$. Two possible remedies for this inconsistency are discussed in Corollary 2.1 of Beran (1997). One of these methods, called the '$m$-out-of-$n$ bootstrap', will be discussed in greater detail in the next chapter.

## 3.5 Mean in the infinite variance case and bootstrap consistency

Usually when one applies the bootstrap method to the distribution of the sample mean one can expect the result to be consistent, i.e., the bootstrap estimate of the distribution converges to the true distribution as the sample size becomes large. However, there is one very important assumption that must be satisfied for the consistency to hold; the second moment must be finite. Bickel and Freedman (1981) and Singh (1981) both show that if this assumption is not met then the bootstrap fails for this very simple statistic. Formally, Singh (1981) states in his Theorem 1.A that if $X_1, X_2, \ldots, X_n$ are i.i.d. random variables from some unknown distribution $F$ and if $G_n$ is the distribution defined as

$$G_n(x) = \mathrm{P}(\sqrt{n}(\bar{X}_n - \mu) \leq x),$$

and $G_n^*$ is the distribution defined as

$$G_n^*(x) = \mathrm{P}^*(\sqrt{n}(\bar{X}_n^* - \bar{X}_n) \leq x),$$

then we require that $\mathrm{E}(X_i^2) < \infty$ for the following statements to be true:

$$\mathrm{P}\left(\sup_{x\in\mathbb{R}} |G_n^*(x) - G_n(x)| \xrightarrow[\infty]{n} 0\right) = 1,$$

or

$$\rho_\infty(G_n^*, G_n) \xrightarrow{a.s.} 0, \quad \text{as } n \to \infty.$$

Babu (1984) notes that in almost all of the situations where the bootstrap method is applied, the distribution of the population is assumed to have at least a finite second moment. The papers by Athreya (1987) and Knight (1989) each provide different proofs of the inconsistency of the bootstrap in the infinite variance case. Giné and Zinn (1989) and Hall (1990) go further and discuss the necessity and sufficiency of the finite second moment.

**Remark:**

The finite second moment must also hold in cases where the statistic being bootstrapped is a function of the sample mean (Shao and Tu 1995). This will be discussed in the next section.

## 3.6 Functions of the sample mean and bootstrap consistency

Additional bootstrap consistency problems occur when one works with statistics that are functions of the sample mean, i.e., we now have a parameter of the form

$$\theta = g(\mu), \tag{3.2}$$

where $g(\cdot)$ is some known function, and which is estimated by the plug-in estimator

$$\hat{\theta}_n = g(\bar{X}_n), \tag{3.3}$$

where $\bar{X}_n = n^{-1}\sum_{i=1}^n X_i$ with $X_1, X_2, \ldots, X_n$ i.i.d. random variables from an unknown distribution function $F$.

Usually there would be no consistency problems when applying the bootstrap in this situation. However, when the function $g(\cdot)$ exhibits certain undesirable properties, or when certain moment conditions on the $X_i$'s are not met, the bootstrap can fail to produce consistent results.

### 3.6.1 Smooth functions of the sample mean

As in the previous section, which dealt with the mean in the infinite variance case, statistics that are smooth functions of the sample mean also exhibit bootstrap consistency problems when the second moment is not finite. Theorem 3.1 in Shao and Tu (1995) states quite simply that if we have a finite variance *and* the function being considered is a smooth function of the sample mean, then the bootstrap is strongly consistent. Stated formally we have:

**Theorem:**

*If* $\mathrm{E}(|X_i|^2) < \infty$, *the statistic* $\hat{\theta}_n$ *is defined as in equation (3.3), the function* $g(\cdot)$ *is continuously differentiable at* $\mu = \mathrm{E}(X_i)$, *with* $g'(\mu) \neq 0$, $G_n$ *is the distribution function defined as* $G_n(x) = \mathrm{P}(\sqrt{n}(\hat{\theta}_n - \theta) \leq x)$ *and* $G_n^*$ *is the distribution function defined as* $G_n^*(x) = \mathrm{P}^*(\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n) \leq x)$, *then*

$$\mathrm{P}\left(\sup_{x\in\mathbb{R}} |G_n^*(x) - G_n(x)| \xrightarrow[\infty]{n} 0\right) = 1,$$

*or*

$$\rho_\infty(G_n^*, G_n) \xrightarrow{a.s.} 0.$$

### 3.6.2 Non-smooth functions of the sample mean

When the underlying distribution has a finite variance, but the function $g$ is non-smooth, then the bootstrap can still fail. This is most easily seen by considering the following example which is also presented in Shao and Tu (1995):

Let $\theta$ and $\hat{\theta}_n$ be defined as in (3.2) and (3.3), and suppose that the function $g$ is continuously second order differentiable in a neighbourhood of $\mu$. Suppose further that the derivative of $g$ in the point $\mu$ is equal to zero, i.e., $g'(\mu) = \frac{d}{dx}g(x)\big|_{x=\mu} = 0$, and that the second derivative of $g$ in the point $\mu$ is non-zero, i.e., $g''(\mu) = \frac{d^2}{dx^2}g(x)\Big|_{x=\mu} \neq 0$. This situation produces an inconsistency when the traditional bootstrap is applied because the bootstrap statistic does not mimic the behaviour of the sample statistic. This can be seen by applying Taylor expansions to the statistic $\hat{\theta}_n - \theta$. The expansion is then expressed as follows:

$$\begin{aligned} \hat{\theta}_n - \theta &= g'(\mu)(\bar{X}_n - \mu) + \tfrac{1}{2}g''(\mu)(\bar{X}_n - \mu)^2 + o_p(n^{-1}) \\ &= \tfrac{1}{2}g''(\mu)(\bar{X}_n - \mu)^2 + o_p(n^{-1}). \end{aligned}$$

Note that the first term in the expansion falls away due to the fact that $g'(\mu) = 0$.

From the Central Limit Theorem we can see the following:

$$n(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{D}} \tfrac{1}{2}g''(\mu)Z_\sigma^2, \tag{3.4}$$

where $Z_\sigma$ is a normally distributed random variable with mean 0 and variance $\sigma^2$.

Let $X_1^*, X_2^*, \ldots, X_n^*$ be an i.i.d. sample drawn from the empirical distribution function of the sample data $X_1, X_2, \ldots, X_n$. Let $\hat{\theta}_n^* = g(\bar{X}_n^*)$ where $\bar{X}_n^*$ is the mean of $X_1^*, X_2^*, \ldots, X_n^*$. Applying the traditional bootstrap to this statistic and performing a Taylor expansion we find that:

$$\hat{\theta}_n^* - \hat{\theta}_n = g'(\bar{X}_n)(\bar{X}_n^* - \bar{X}_n) + \tfrac{1}{2}g''(\bar{X}_n)(\bar{X}_n^* - \bar{X}_n)^2 + o_{p^*}(n^{-1}), \tag{3.5}$$

where $o_{p^*}$ is defined in Appendix C. From Theorem 3.1 of Shao and Tu (1995) we find that

$$\frac{n}{2}g''(\bar{X}_n)(\bar{X}_n^* - \bar{X}_n)^2 \xrightarrow{\mathcal{D}^*} \tfrac{1}{2}g''(\mu)Z_\sigma^2, \quad a.s.,$$

where the notation $\xrightarrow{\mathcal{D}^*}$ is defined in Appendix C. Unfortunately, the first term in the Taylor expansion does not converge to zero (this is because it is not necessarily true that $g'(\bar{X}_n) = 0$ if $g'(\mu) = 0$). We see this fact from a Taylor expansion and applying the Central Limit Theorem:

$$\sqrt{n}g'(\bar{X}_n) = \sqrt{n}g''(\mu)(\bar{X}_n - \mu) + o_p(1) \xrightarrow{\mathcal{D}} g''(\mu)Z_\sigma \quad a.s.$$

Furthermore, Theorem 3.1 of Shao and Tu (1995) implies that $\sqrt{n}\left(\bar{X}_n^* - \bar{X}_n\right) \xrightarrow{\mathcal{D}^*} Z_\sigma$. Hence,

$$ng'(\bar{X}_n)\left(\bar{X}_n^* - \bar{X}_n\right) = \sqrt{n}g'(\bar{X}_n)\sqrt{n}\left(\bar{X}_n^* - \bar{X}_n\right) \xrightarrow{\mathcal{D}^*} g''(\mu)Z_\sigma^2 \neq 0 \quad a.s.$$

This shows the inconsistency of traditional bootstrap, since the distribution of $n$ times the quantity in equation (3.5) does not converge to the same quantity as the distribution of the quantity in equation (3.4).

The article by Shao (1994) not only includes the above example, but also includes examples where the function $g(\cdot)$ is non-differentiable. The bootstrap is also inconsistent in those cases.

## 3.7 Extremes and bootstrap consistency

One of the earliest discussed examples of bootstrap inconsistency involved attempts to bootstrap sample extremes such as the sample minimum or maximum. (see Bickel and Freedman (1981), Swanepoel (1986), Deheuvels, Mason and Shorack (1993), Shao and Tu (1995) Bickel et al. (1997), Politis et al. (1999), and Chernick (2008) ). Limiting this discussion to the sample maximum we will show why the bootstrap fails to produce a consistent result.

Let $X_1, X_2, \ldots, X_n$ be independent uniformly distributed random variables with support $[0, \theta]$. Choose $\hat{\theta}_n = X_{(n)} = \max(X_1, X_2, \ldots, X_n)$.

It is well known that following statistic has a limiting distribution given by:

$$M_n = \frac{n(\theta - X_{(n)})}{\theta} \xrightarrow{\mathcal{D}} Z,$$

where $Z$ is a standard exponential random variable.

Unfortunately, the bootstrap version of $M_n$, viz.

$$M_n^* = \frac{n(X_{(n)} - X_{(n)}^*)}{X_{(n)}},$$

where $X_{(n)}^* = \max(X_1^*, X_2^*, \ldots, X_n^*)$, does not have a weak limit (see for example Bickel and Freedman (1981)). The reason for this can be seen from the following derivation:

$$\begin{aligned} \mathrm{P}^*(M_n^* = 0) &= \mathrm{P}^*(X_{(n)}^* = X_{(n)}) \\ &= 1 - \mathrm{P}^*(X_{(n)}^* \leq X_{(n-1)}) \\ &= 1 - (F_n(X_{(n-1)}))^n \\ &= 1 - \left(1 - \frac{1}{n}\right)^n \\ &\to 1 - e^{-1}. \end{aligned}$$

## 3.8 Other bootstrap inconsistency problems

The literature is positively littered with examples where the bootstrap is inconsistent and the contents of this chapter really only scratches the surface of the research which has gone into the topic. To illustrate, the following list, compiled by considering some of the situations listed in Andrews (2000) and through an extensive scouring of the literature, is a small sample of the examples which have been investigated by other authors, but were excluded from this thesis:

- ⋆ Estimators of the eigenvalues of a covariance matrix whose eigenvalues are not distinct (Beran and Srivastava 1985).
- ⋆ Nonparametric kernel estimators of the mode of a smooth unimodal density when the smoothing parameter for both the bootstrap and real world are optimally chosen (Romano 1988).
- ⋆ Nondifferentiable functions of the EDF (Dümbgen 1993).
- ⋆ Extrema for unbounded distributions (Fukuchi 1994, Deheuvels et al. 1993).
- ⋆ The distribution of the squared sample mean, $\bar{X}_n^2$, when the population mean is zero, i.e., $\mu = 0$ (Datta 1995, Dodd and Korn 2007).

⋆ Approximating the null distribution of Cramér - von Mises goodness-of-fit tests with doubly censored data (Bickel and Ren 1996).

⋆ A sample quantile when the density has a jump (Huang, Sen and Shao 1996).

⋆ Confidence intervals for endpoints of a distribution function (Athreya and Fukuchi 1997).

⋆ When a parameter is on the boundary of a parameter space (Andrews 2000).

**Comment:** It is difficult to see any sort of common ground linking these failures. Intuitively it feels as if there should be a general cause of the failure of the bootstrap, and while many of the above-mentioned articles do provide largely generalized discussions of the particular bootstrap failure being considered, each one seems to be isolated from the others, that is, each one seems to have its own set of reasons for failing and remedies for correcting them.

A possible (but highly ambitious and optimistic) avenue for future research might be to try and unify these situations and determine a simple set of rules which could be used to indicate bootstrap failure in *any* situation.

# Chapter 4

# The $m$-out-of-$n$ bootstrap

## 4.1 Introduction

A modification to the traditional bootstrap proposed by, among others, Bickel and Freedman (1981), Bretagnolle (1983), and Swanepoel (1986), has been shown to remedy many of the inconsistency problems associated with the non-regular cases discussed in Chapter 3. The modification typically involves sampling fewer than $n$ observations from $X_1, X_2, \ldots, X_n$. The notation that will be employed to denote this smaller resampling size is $m$, where $m \leq n$. Other restrictions on $m$ will be discussed shortly. This procedure has been labeled the *modified bootstrap* (Swanepoel 1986), the *rescaled* bootstrap (Dümbgen 1993, Andrews 2000), the $m/n$ bootstrap (Bickel et al. 1997, Chung and Lee 2001) and the *$m$-out-of-$n$ bootstrap* (see, for example, Bickel and Sakov (2008)). However, for the purposes of this text only the terms $m$-out-of-$n$ and MOON bootstrap will be adopted (the abbreviation 'MOON' is an acronym for '$\underline{m}$-$\underline{o}$ut-$\underline{o}$f-$\underline{n}$'). In addition to this we will also distinguish between a *naive* application of the $m$-out-of-$n$ bootstrap and a *corrected* version of the $m$-out-of-$n$ bootstrap, which will be referred to as the CMOON bootstrap.

In order to achieve consistency when applying the $m$-out-of-$n$ bootstrap one must note that the parameter $m$ is a function of the original sample size $n$. Therefore, a more accurate way of representing this quantity might be to use $m_n$ rather than simply $m$. The subscript $n$ will be introduced where necessary to emphasize the dependency on $n$.

The purpose of the MOON bootstrap is twofold (Bickel and Sakov 2002):

⋆ Typically, when the traditional bootstrap is not consistent, then the MOON bootstrap is consistent under minimal conditions (Bickel and Freedman 1981, Bickel et al. 1997, Politis and Romano 1994, Politis et al. 1999).

  When the $m$-out-of-$n$ bootstrap is applied in order to obtain a weakly consistent estimator, then we require the following restriction on the relationship between $m$ and $n$ :

$$m \to \infty \text{ and } \frac{m}{n} \to 0 \text{ as } n \to \infty,$$

  or simply $m = o(n)$.

  On the other hand, if one needs strong consistency, then the relationship needs to change to

$$m \to \infty \text{ and } \frac{m \log\log n}{n} \to 0 \text{ as } n \to \infty,$$

  or $m = o\left(\frac{n}{\log\log(n)}\right)$ (Shao and Tu 1995).

- ⋆ When the traditional bootstrap is consistent, then the MOON bootstrap is used to attain equivalent behaviour, but with second or higher order accuracy, with reduced computational time (Bickel and Yahav 1988, Bickel et al. 1997, Beran 1997, Sakov 1998).

### 4.1.1 Different types of $m$-out-of-$n$ bootstrap

In Bickel et al. (1997) they discuss three different ways of sampling fewer than $n$ observations. The bootstrap schemes discussed there are listed below.

- ⋆ The $n/n$ bootstrap, which we refer to as the 'traditional bootstrap' done with replacement. That is, we randomly sample $n$ observations with replacement from the $n$ observations in the sample.

- ⋆ The $m/n$ bootstrap, which we refer to as the '$m$-out-of-$n$ bootstrap' done with replacement. That is, we randomly sample $m$ observations with replacement from the $n$ observations in the sample.

- ⋆ The $\binom{n}{m}$, which is known as the '$m$-out-of-$n$ bootstrap' done *without* replacement. That is, we randomly sample without replacement $m$ observations from the $n$ observations in the sample. This type of bootstrap is more commonly referred to as *subsampling*. Subsampling will be discussed in more detail in the next subsection.

- ⋆ Sample Splitting. In this case we create samples of size $m$ by simply splitting the original sample into $k$ blocks, each of size $m$ (i.e., $n = mk$).

For the purposes of this chapter, and for future reference, we will be primarily focused on the case where samples of size $m$ are drawn independently (with replacement) from the sample, i.e., the $m/n$ or $m$-out-of-$n$ bootstrap. The other cases will be largely disregarded as they are not useful in this discussion. However, a short description of one of the methods (the $\binom{n}{m}$ bootstrap or subsampling) will be presented in the next subsection so that a contrast can be drawn with the $m$-out-of-$n$ bootstrap method which will form the bulk of this chapter.

### 4.1.2 Subsampling

Before continuing with the $m$-out-of-$n$ bootstrap we will briefly look at the bootstrap method referred to as the $\binom{n}{m}$ bootstrap or subsampling. This technique was introduced by Wu (1990) and subsequently championed by Dimitris N. Politis and Joseph P. Romano in a number of articles (Politis and Romano 1994, Politis, Romano and Wolf 2001) and also in a book written on the subject (Politis et al. 1999). The basics of subsampling for the i.i.d. case, found in Politis et al. (1999), are summarized here.

Let $X_1, X_2, \ldots, X_n$ be i.i.d. from some distribution $F$. Let $\theta$ be the parameter of interest, and let $\hat{\theta}_n = \hat{\theta}_n(X_1, X_2, \ldots, X_n)$ be an estimator for $\theta$. Let the sampling distribution function of the quantity $\tau_n(\hat{\theta}_n - \theta)$, where $\tau_n$ is some normalizing value (e.g., $\tau_n = \sqrt{n}$), be given by

$$J_n(x, F) = \mathrm{P}_F\left(\tau_n(\hat{\theta}_n - \theta) \leq x\right).$$

**Assumption:**
*Assume that there exists a nondegenerate distribution function $J(x, F)$ such that $J_n(x, F)$ converges weakly to $J(x, F)$ as $n \to \infty$, uniformly in $x$.*

Let $\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_{m_n}$ denote a subset of size $m_n$ drawn *without replacement* from $X_1, X_2, \ldots, X_n$. There are $\binom{n}{m_n}$ ways of drawing these samples. Note that in this case the sample size $m_n$ is dependent on $n$ (as indicated by the subscript $n$). For simplicity we will drop the subscript and denote the sample size as $m$. Let $\tilde{\theta}_m = \hat{\theta}_m(\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_m)$ denote the statistic calculated from the sample $\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_m$. We will assume that $m$ and $n$ are related as follows:

$$\frac{m}{n} \to 0 \qquad \text{as } n \to \infty.$$

Define the following estimator for $J_n(x, F)$ as

$$L_m(x) = \binom{n}{m}^{-1} \sum_{1 \leq i_1 < i_2 < \cdots < i_m \leq n} \mathrm{I}\left(\tau_m \left(\hat{\theta}_m(\tilde{X}_{i_1}, \tilde{X}_{i_2}, \ldots, \tilde{X}_{i_m}) - \hat{\theta}_n\right) \leq x\right).$$

This estimator has good asymptotic properties as defined by Theorem 2.2.1 in Politis et al. (1999). If $\tau_m/\tau_n \to 0$ , $m \to \infty$ and $m/n \to 0$ as $n \to \infty$, then the main points of interest in that theorem are:

- ⋆ If $x$ is a continuity point of the limiting distribution $J(\cdot\,, F)$, then $L_m(x) \xrightarrow{p} J(x, F)$.
- ⋆ If $J(\cdot\,, F)$ is continuous, then

$$\sup_{x \in \mathbb{R}} |L_m(x) - J_n(x, F)| \xrightarrow{p} 0.$$

If we adopt the additional restriction that $\tau_m(\hat{\theta}_n - \theta) \xrightarrow{a.s.} 0$ and that $\sum_n \exp\left(-d(n/m)\right) < \infty$ for every $d > 0$, then:

- ⋆ If $x$ is a continuity point of the limiting distribution $J(\cdot\,, F)$, then $L_m(x) \xrightarrow{a.s.} J(x, F)$.
- ⋆ If $J(\cdot\,, F)$ is continuous, then

$$\sup_{x \in \mathbb{R}} |L_m(x) - J_n(x, F)| \xrightarrow{a.s.} 0.$$

Clearly the subsampling technique has some good properties, but one of the major criticisms of the technique is that it requires the rather restrictive assumption that the quantity $J_n(x, F)$ must have a weak limit. Also, like the traditional bootstrap, it suffers from 'nonregular' situations where it is inconsistent (Politis et al. 1999, Example 2.3.2).

## 4.2 Does the $m$-out-of-$n$ bootstrap work?

This section will discuss, by referring to the situations mentioned in Chapter 3, how the $m$-out-of-$n$ bootstrap can remedy the inconsistencies experienced by the traditional bootstrap in the non-regular cases.

### 4.2.1 The $m$-out-of-$n$ bootstrap and $U$-statistics

The bootstrap applied to non-degenerate and degenerate $U$-statistics (discussed in the previous chapter) were both shown to be consistent (Bickel and Freedman 1981, Arcones and Giné 1992, Janssen 1997). However, the application to the degenerate $U$-statistic did require one to modify the kernel of the $U$-statistic in order to reflect the 'degeneracy' of the $U$-statistic in the bootstrap world.

Bretagnolle (1983) investigated the option of using the $m$-out-of-$n$ bootstrap to gain consistency of the bootstrap without having to resort to modifications of the kernel. Bretagnolle states that if the bootstrap is applied to the naive bootstrap estimate of the degenerate $U$-statistic, then consistency can be achieved in probability if the bootstrap sample size, $m$, has the property $m/n \to 0$ as $n \to \infty$, and almost sure consistency can be achieved if it has the property $m(\log(n))^b/n \to 0$ as $n \to \infty$ and for some $b > 1$. However, these results are obtained under some rather strong moment conditions on the kernel of the $U$-statistic (see for example Arcones and Giné (1992)).

### 4.2.2 The $m$-out-of-$n$ bootstrap and super-efficient estimators

Beran (1997) shows that if we choose the resample size, $m$, such that $m/n \to 0$ as $m \to \infty$ and $n \to \infty$, then the bootstrap is consistent for Hodges' super-efficient estimator.

### 4.2.3 The $m$-out-of-$n$ bootstrap and functions of the sample mean

#### The sample mean in the infinite variance case

In the case where the function of the sample mean is $g(x) = x$, then the condition for consistency of the bootstrap is simply that the second moment must be finite. The inconsistency that occurs when the second moment is not finite can be remedied by choosing the resample size, $m$, to be smaller than the original sample's size, $n$. This is discussed in Arcones (1991), Arcones and Giné (1989) and Wu, Carlstein and Cambanis (1989).

#### Non-smooth functions of the sample mean in the finite variance case

We will consider the example presented in Section 3.6.2 (also found in Shao and Tu (1995)) to show that the $m$-out-of-$n$ bootstrap remedies the inconsistency or failure of the traditional bootstrap found there.

Let $\hat{\theta}_n = g(\bar{X}_n)$, $\hat{\theta}_n^* = \hat{\theta}_n(X_1^*, X_2^*, \ldots, X_n^*) = g(\bar{X}_n^*)$, $\hat{\theta}_m^* = \hat{\theta}_m(X_1^*, X_2^*, \ldots, X_m^*) = g(\bar{X}_m^*)$ and $\bar{X}_m^* = 1/m \sum_{i=1}^m X_i^*$. Recall that in this example $g'(\mu) = \frac{d}{dx} g(x)\big|_{x=\mu} = 0$, and that the second derivative of $g$ in the point $\mu$ is non-zero. Also, let $\mathrm{E}(X_i^2) < \infty$, $i = 1, 2, \ldots, n$. Then,

$$\hat{\theta}_m^* - \hat{\theta}_n = g'(\bar{X}_n)\left(\bar{X}_m^* - \bar{X}_n\right) + \tfrac{1}{2} g''(\bar{X}_n)\left(\bar{X}_m^* - \bar{X}_n\right)^2 + o_{p^*}(m^{-1}).$$

By Theorem 2.1 in Bickel and Freedman (1981), for almost all sequences $X_1, X_2, \ldots$,

$$\frac{m}{2} g''(\bar{X}_n)\left(\bar{X}_m^* - \bar{X}_n\right)^2 \xrightarrow{\mathcal{D}^*} \tfrac{1}{2} g''(\mu) Z_\sigma^2, \quad a.s.,$$

as $m \to \infty$ and $n \to \infty$.

It can be shown that if $m \log\log n/n \to 0$ then the term $g'(\bar{X}_n)\left(\bar{X}_m^* - \bar{X}_n\right)$ is $o_{p^*}(m^{-1})$, i.e.,

$$\hat{\theta}_m^* - \hat{\theta}_n = \tfrac{1}{2} g''(\bar{X}_n)\left(\bar{X}_m^* - \bar{X}_n\right)^2 + o_{p^*}(m^{-1}).$$

Therefore, $\mathrm{P}^*(m(\hat{\theta}_m^* - \hat{\theta}_n) \le x)$ is strongly consistent if $m = m_n \to \infty$, and $m_n \log\log n/n \to 0$, since $g'(\bar{X}_n) = O(\sqrt{\log\log n/n})$ $a.s.$

It can also be shown that it is weakly consistent if $m = m_n \to \infty$ and $m_n/n \to 0$, since $g'(\bar{X}_n) = O_p(n^{-1/2})$.

### 4.2.4 The $m$-out-of-$n$ bootstrap and extremes

In Section 3.7, the bootstrap was shown to be inconsistent for the sample maximum. It has been shown in the case when the distribution of the data, $F$, is uniform on $[0, \theta]$ that the application of the $m$-out-of-$n$ bootstrap can rectify this inconsistency. Swanepoel (1986) showed that if $m = o\left(n^{(\epsilon+1)/2}/\sqrt{\log(n)}\right)$, with $0 < \epsilon < 1$, then

$$\frac{m(X_{(n)} - X^*_{(m)})}{X_{(n)}} \xrightarrow{\mathcal{D}^*} Z, \quad a.s., \tag{4.1}$$

where $Z$ is a standard exponential random variable and $X^*_{(m)}$ is defined as the maximum of $X^*_1, X^*_2, \ldots, X^*_m$, which are i.i.d. from the EDF, $F_n$.

Deheuvels et al. (1993) improved on this result by showing that:

- ⋆ If $m/n \to 0$ then the convergence in equation (4.1) holds in probability.
- ⋆ If $m \log\log(n)/n \to 0$ then the convergence in equation (4.1) holds almost surely.

## 4.3 The naive $m$-out-of-$n$ bootstrap versus the corrected $m$-out-of-$n$ bootstrap

In the following section the importance of first investigating the statistic being used before applying the $m$-out-of-$n$ bootstrap is discussed, in particular we need to examine the normalization constant which causes the statistic to have a *non-degenerate* limiting distribution. This discussion is motivated using simple examples involving the sample mean; examples where the tradtional bootstrap is consistent. The naive application of the $m$-out-of-$n$ bootstrap to statistics which do not have a non-degenerate limiting distribution can produce results which do not agree with the traditional bootstrap.

*We propose rewriting a statistic which does not have a non-degenerate limiting distribution in terms of a statistic that does have a non-degenerate limiting distribution, and then applying the m-out-of-n bootstrap plug-in principle to the portion of the statistic that has the non-degenerate limiting distribution while replacing parameters by their estimators. This procedure will be referred to as the 'corrected' m-out-of-n bootstrap or CMOON bootstrap.*

To illustrate these ideas, consider the following example. Suppose that $X_1, X_2, \ldots, X_n$ are i.i.d. random variables. We start by looking at the statistic $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. The variance of this statistic can be estimated using the traditional bootstrap through the following expression:

$$\text{Var}^*(\bar{X}^*_n),$$

where $\bar{X}^*_n = \frac{1}{n}\sum_{i=1}^n X^*_i$ and $X^*_1, X^*_2, \ldots, X^*_n$ is a random sample of size $n$ drawn with replacement from $X_1, X_2, \ldots, X_n$.

The MOON bootstrap estimate of this quantity is then:

$$\text{Var}^*(\bar{X}^*_m),$$

where $\bar{X}^*_m = \frac{1}{m}\sum_{i=1}^m X^*_i$ and $X^*_1, X^*_2, \ldots, X^*_m$ is a random sample of size $m$ drawn with replacement from $X_1, X_2, \ldots, X_n$.

The MOON bootstrap estimate simplifies to:

$$\begin{aligned}\text{Var}^*(\bar{X}^*_m) &= \text{Var}^*\left(\frac{1}{m}\sum_{i=1}^m X^*_i\right)\\ &= \frac{1}{m^2}\text{Var}^*\left(\sum_{i=1}^m X^*_i\right)\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{m}\operatorname{Var}^*\left(X_1^*\right) \\
&= \frac{1}{m}\left[\mathrm{E}^*\left(X_1^{*2}\right) - \mathrm{E}^*\left(X_1^*\right)^2\right] \\
&= \frac{1}{m}\left[\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2\right] \\
&= \frac{S_n^2}{m},
\end{aligned} \tag{4.2}$$

where $S_n^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2$.

Now, consider rewriting the statistic, $\bar{X}_n$ as:

$$\bar{X}_n = \frac{\sigma}{\sqrt{n}}\left[\frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{\sigma}\right] + \mu.$$

This form of the statistic is preferable because it is now expressed in terms of a statistic with a non-degenerate limiting distribution. In fact, $\bar{X}_n$ has been expressed in terms of an asymptotic *pivotal* statistic. When it is possible, it is always better to express a statistic in a pivotal form because, when applying the bootstrap, any good choice of an estimator of $F$ will usually result in bootstrap consistency. Cases where this is not the case are discussed in Chapter 3.

The variance of $\bar{X}_n$ can now be expressed as:

$$\operatorname{Var}(\bar{X}_n) = \frac{\sigma^2}{n}\operatorname{Var}\left[\frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{\sigma}\right],$$

and we can now estimate the variance of $\bar{X}_n$ using the CMOON bootstrap principle as follows:

$$\begin{aligned}
\widehat{\operatorname{Var}(\bar{X}_n)} &= \frac{S_n^2}{n}\operatorname{Var}^*\left(\frac{\sqrt{m}\left(\bar{X}_m^* - \bar{X}_n\right)}{S_n}\right) \\
&= \frac{S_n^2}{n}\frac{m}{S_n^2}\operatorname{Var}^*\left(\bar{X}_m^* - \bar{X}_n\right) \\
&= \frac{m}{n}\operatorname{Var}^*\left(\bar{X}_m^* - \bar{X}_n\right) \\
&= \frac{m}{n}\operatorname{Var}^*\left(\bar{X}_m^*\right) \\
&= \frac{S_n^2}{n}, \qquad \text{from equation (4.2)},
\end{aligned} \tag{4.3}$$

which is the same as the traditional ideal bootstrap estimate of the variance of the mean.

From equation (4.3) it is clear that the naive MOON bootstrap estimator, (4.2), multiplied by a simple correction factor, $(\frac{m}{n})$, can be used to obtain the same result as the traditional $n$-out-of-$n$ bootstrap.

**Rewriting a statistic in terms of a statistic with a non-degenerate limiting distribution: The corrected $m$-out-of-$n$ bootstrap form**

In the above example we saw that the sample mean, $\bar{X}_n$, was rewritten in such a way that it contained a statistic with a non-degenerate limiting distribution. In general, this method of rewriting simple statistics so that they contain a statistic with a non-degenerate limiting distribution is only possible if we know the normalization constant beforehand. In our discussion we will assume that the normalization constant is known for the statistic being studied. We will now propose a general method for rewriting these simple statistics in terms of quantities with non-degenerate limiting distributions:

*Let $\hat{\theta}_n$ denote some estimator for the parameter $\theta$. Assume that $n^\alpha(\hat{\theta}_n - \theta)$ has a non-degenerate limiting distribution, where the normalizing constant $n^\alpha$ is known with $\alpha > 0$. Then, in order to apply the m-out-of-n bootstrap correctly we will rewrite the statistic $\hat{\theta}_n$ as*

$$\hat{\theta}_n = \frac{1}{n^\alpha}\left[n^\alpha(\hat{\theta}_n - \theta)\right] + \theta. \tag{4.4}$$

*When this technique is coupled with the m-out-of-n bootstrap we will refer to the result as the 'corrected' m-out-of-n bootstrap, or the CMOON bootstrap.*

**Remark:**
It should be noted that this CMOON technique is only required when the statistic being studied does *not* have a non-degenerate limiting distribution. When the statistic has a non-degenerate limiting distribution the CMOON bootstrap and the naive $m$-out-of-$n$ bootstrap are equivalent.

### 4.3.1 The 'naive' bootstrap applied to standard error estimation

Application of the $m$-out-of-$n$ bootstrap in the estimation of standard error may seem simple since one only has to calculate the statistic from the bootstrap samples of size $m$, and determine the standard error using the usual techniques. However, as illustrated with the sample mean in the previous section, it is not always as simple as this.

Consider the standard error of an estimator $\hat{\theta}_n$:

$$\mathrm{SE}(\hat{\theta}_n) = \sqrt{\mathrm{Var}\left(\hat{\theta}_n\right)}.$$

A 'naive' application of the $m$-out-of-$n$ bootstrap is:

$$\widehat{SE_{moon}}(\hat{\theta}_n) = \sqrt{\mathrm{Var}^*(\hat{\theta}^*_m)}. \tag{4.5}$$

This implementation of the $m$-out-of-$n$ bootstrap is clearly flawed because the bootstrap has been applied to a quantity which might not have a non-degenerate limiting distribution. A more reliable way is to use the corrected method described next.

### 4.3.2 The 'corrected' $m$-out-of-$n$ bootstrap applied to standard error estimation

In order for one to correctly apply the $m$-out-of-$n$ bootstrap it is necessary to make an assumption about the limiting distribution of the statistic being studied.

We assume that the statistic $n^\alpha(\hat{\theta}_n - \theta)$ has a non-degenerate limiting distribution, i.e.,

$$n^\alpha(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{D}} Z, \tag{4.6}$$

where $Z$ is a non-degenerate random variable for some $\alpha > 0$.

Consider rewriting the standard error of the statistic $\hat{\theta}_n$ in terms of the form given in (4.6) as follows:

$$\begin{aligned}\mathrm{SE}(\hat{\theta}_n) &= \sqrt{\mathrm{Var}(\hat{\theta}_n)} \\ &= \sqrt{\frac{1}{n^{2\alpha}}\,\mathrm{Var}\left(n^\alpha(\hat{\theta}_n - \theta)\right)}, \text{ from (4.4).}\end{aligned}$$

Applying the $m$-out-of-$n$ plug-in principle, we find the following:

$$\begin{aligned}
\widehat{SE_{cmoon}}\left(\hat{\theta}_n\right) &= \sqrt{\frac{1}{n^{2\alpha}}\operatorname{Var}^*\left(m^\alpha\left(\hat{\theta}^*_m - \hat{\theta}_n\right)\right)} \\
&= \sqrt{\frac{1}{n^{2\alpha}}\operatorname{Var}^*\left(m^\alpha\hat{\theta}^*_m\right)} \\
&= \sqrt{\frac{m^{2\alpha}}{n^{2\alpha}}\operatorname{Var}^*\left(\hat{\theta}^*_m\right)} \\
&= \left(\frac{m}{n}\right)^\alpha\sqrt{\operatorname{Var}^*\left(\hat{\theta}^*_m\right)} \\
&= \left(\frac{m}{n}\right)^\alpha \widehat{SE_{moon}}\left(\hat{\theta}_n\right), \qquad \text{from equation (4.5).} \qquad (4.7)
\end{aligned}$$

Thus, we find that the standard error of the 'corrected $m$-out-of-$n$ bootstrap' technique reduces to the standard error of the 'naive $m$-out-of-$n$ bootstrap' multiplied by the factor $(m/n)^\alpha$.

**Some remarks:**

We use the more general notation of $\alpha$ and not the more common exponent of $\frac{1}{2}$ because there are some cases where it is not equal to $\frac{1}{2}$. For example,

1. For so-called 'degenerate $U$-statistics' we have that $\alpha = 1$. For an example see Bickel and Freedman (1981).
2. For extreme order statistics (Swanepoel 1986), we typically have that $\alpha$ is equal to 1.
3. For the non-smooth statistic presented in Section 3.6.2 (and also Shao and Tu (1995)) we once again have that $\alpha = 1$.

**Example:**

We will now consider the case where $\hat{\theta}_n = \bar{X}_n$. That is,

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{\mathcal{D}} Z \sim \mathrm{N}(0, \sigma^2).$$

In this case we have $\alpha = 0.5$. This is a relatively simple example, but it helps illustrate the correct method of using the $m$-out-of-$n$ bootstrap to estimate standard error. We will now compare the theoretical results using the 'naive' $m$-out-of-$n$ bootstrap, and the corrected $m$-out-of-$n$ bootstrap when estimating the standard error of the sample mean.

As shown in (4.3), the 'naive $m$-out-of-$n$ bootstrap' estimate of standard error of $\bar{X}_n$ is equal to $\sqrt{\frac{S_n^2}{m}}$. The correct implementation of the $m$-out-of-$n$ bootstrap will remedy this error. It is easy to see that if one chooses $\alpha = \frac{1}{2}$ in equation (4.7) and applies the equation on the statistic $\hat{\theta}_n = \bar{X}_n$, then the result is equal to $\frac{S_n}{\sqrt{n}}$, which is the same as the traditional $n$-out-of-$n$ bootstrap estimate of the standard error of the sample mean.

**Remarks:**

1. We can now also provide an expression for a 'corrected $m$-out-of-$n$ bootstrap' estimator for the mean squared error of an estimator, namely $\mathrm{MSE}(\hat{\theta}_n) = (\mathrm{SE}(\hat{\theta}_n))^2 + (\mathrm{Bias}(\hat{\theta}_n))^2$.
2. We focused in the above example on the sample mean because, in an overwhelming number of cases, $\hat{\theta}_n$ can be expressed asymptotically as an average of i.i.d. random variables.

3. Naturally, it would not be reasonable to expect that this correction will work this 'perfectly' for all statistics (if it did, then there would be absolutely no need to determine the size of $m$: $m$ does not affect the result). In these cases one would expect that $m$ will only play a role in higher order terms.

4. In applying the bootstrap procedure and drawing samples $X_1^*, X_2^*, \ldots, X_m^*$, it means that we can choose $m$ relatively 'small' – this translates into a saving in computer time. When performing bootstrap simulations we will make use of $m \in M \subset \{1, 2, \ldots, n\}$, and we will choose $M$ to be a relatively small subset of $\{1, 2, \ldots, n\}$.

5. This is a possible inspiration for finding a data-based estimator for $m$.

### 4.3.3 A limited simulation study comparing the corrected and naive $m$-out-of-$n$ bootstrap

In this section the results of a limited simulation study are presented. This study compares the performance of the traditional $n$-out-of-$n$ bootstrap, the naive $m$-out-of-$n$ bootstrap, and the corrected $m$-out-of-$n$ bootstrap when estimating the standard error of the mean and the median. The results are compared to the true standard errors of the mean and median.

It should be noted that the statistics used here form part of the 'regular' cases, i.e., when the traditional bootstrap is consistent. The point of this simulation is simply to illustrate the effectiveness of applying the correct method of $m$-out-of-$n$ bootstrapping in cases where the traditional bootstrap is typically applied.

The distribution and parameter configuration of the study is as follows:

- ⋆ The number of Monte-Carlo iterations is set to $MC = 1000$,
- ⋆ the number of bootstrap replications is set to $B = 1000$,
- ⋆ the original sample size is taken to be equal to $n = 101$,
- ⋆ the bootstrap resample size is taken to be equal to $m = 30$,
- ⋆ the distributions used are
  - Exponential with parameter $\lambda = 0.5$,
  - Contaminated normal with parameters $\mu_1 = 0$, $\mu_2 = 0.5$, $\sigma_1 = 2.5$, $\sigma_2 = 3.5$ and contamination probability equal to $p = 0.5$,
- ⋆ the statistics used are
  - the sample mean, $\bar{X}_n$ and
  - the sample median, $\text{Med}(X_1, X_2, \ldots, X_n)$.

In the results of this Monte-Carlo study we provide the true standard error (SE) of the sample mean and median (denoted by $\text{SE}(\bar{X}_n)$ and $\text{SE}(\text{Med}(X_1, X_2, \ldots, X_n))$ respectively). The standard errors of the mean and median are estimated using a bootstrap simulation, these estimates are denoted by $\text{SE}^*_{mean} = \text{SE}^*(\bar{X}^*_n)$ and $\text{SE}^*_{med} = \text{SE}^*(\text{Med}(X_1^*, X_2^*, \ldots, X_n^*))$ respectively. The bootstrap estimation procedure is repeated $MC = 1000$ times and the Monte-Carlo expected values of $\text{SE}^*_{mean}$ and $\text{SE}^*_{med}$ are obtained (these expected values are shown in the table and are denoted by $\text{E}_{MC,SEmean} = \text{E}(\text{SE}^*_{mean})$ and $\text{E}_{MC,SEmed} = \text{E}(\text{SE}^*_{med})$ respectively). In addition to these results, the standard errors of these standard error estimates are also calculated are denoted by $\text{SE}_{MC,SEmean} = \text{SE}(\text{SE}^*_{mean})$ and $\text{SE}_{MC,SEmed} = \text{SE}(\text{SE}^*_{med})$ respectively. Finally, the standard errors of the Monte-Carlo expected values, denoted by $SE(E_{MC,SEmean})$ and $SE(E_{MC,SEmed})$ respectively, are determined.

These three calculations are repeated using the traditional bootstrap, the naive $m$-out-of-$n$ bootstrap and the corrected $m$-out-of-$n$ bootstrap.

| Distribution used: | $Exp\,(\lambda = \frac{1}{2})$ |
|---:|---|
| **The Mean** | |
| Theoretical value: $SE(X_n)=$ | 0.19901 |
| **Traditional bootstrap** | |
| $E_{MC,SEmean} = E(SE^*_{mean}) =$ | 0.19658 |
| $SE_{MC,SEmean} = SE(SE^*_{mean}) =$ | 0.02729 |
| $SE(E_{MC,SEmean}) =$ | 0.00086 |
| **Naive $m$-out-of-$n$ bootstrap** | |
| $E_{MC,SEmean} = E(SE^*_{mean}) =$ | 0.36081 |
| $SE_{MC,SEmean} = SE(SE^*_{mean}) =$ | 0.05070 |
| $SE(E_{MC,SEmean}) =$ | 0.00160 |
| **Corrected $m$-out-of-$n$ bootstrap** | |
| $E_{MC,SEmean} = E(SE^*_{mean}) =$ | 0.19664 |
| $SE_{MC,SEmean} = SE(SE^*_{mean}) =$ | 0.02763 |
| $SE(E_{MC,SEmean}) =$ | 0.00087 |

| The Median | |
|---:|---|
| Theoretical value: $SE(Med(X_1, X_2, \ldots, X_n))=$ | 0.19949 |
| **Traditional bootstrap** | |
| $E_{MC,SEmed} = E(SE^*_{med}) =$ | 0.20567 |
| $SE_{MC,SEmed} = SE(SE^*_{med}) =$ | 0.04847 |
| $SE(E_{MC,SEmed}) =$ | 0.00153 |
| **Naive $m$-out-of-$n$ bootstrap** | |
| $E_{MC,SEmed} = E(SE^*_{med}) =$ | 0.36656 |
| $SE_{MC,SEmed} = SE(SE^*_{med}) =$ | 0.06463 |
| $SE(E_{MC,SEmed}) =$ | 0.00204 |
| **Corrected $m$-out-of-$n$ bootstrap** | |
| $E_{MC,SEmed} = E(SE^*_{med}) =$ | 0.19978 |
| $SE_{MC,SEmed} = SE(SE^*_{med}) =$ | 0.03522 |
| $SE(E_{MC,SEmed}) =$ | 0.00111 |

**Table 4.1:** *Standard errors of the traditional, naive m-out-of-n and corrected m-out-of-n bootstrap for the sample mean and sample median: Exponential distribution with* $\lambda = 0.5$.

| Distribution used: | $ConNorm$ $(\mu_1 = 0, \sigma_1^2 = 6.25, \mu_2 = \frac{1}{2}, \sigma_2^2 = 12.25, p = \frac{1}{2})$ |
|---:|---|
| **The Mean** | |
| Theoretical value: $\mathrm{SE}(\bar{X}_n)=$ | 0.30365 |
| **Traditional bootstrap** | |
| $\mathrm{E}_{MC,SEmean} = \mathrm{E}(\mathrm{SE}^*_{mean}) =$ | 0.30127 |
| $\mathrm{SE}_{MC,SEmean} = \mathrm{SE}(\mathrm{SE}^*_{mean}) =$ | 0.02406 |
| $\mathrm{SE}(\mathrm{E}_{MC,SEmean}) =$ | 0.00076 |
| **Naive $m$-out-of-$n$ bootstrap** | |
| $\mathrm{E}_{MC,SEmean} = \mathrm{E}(\mathrm{SE}^*_{mean}) =$ | 0.55340 |
| $\mathrm{SE}_{MC,SEmean} = \mathrm{SE}(\mathrm{SE}^*_{mean}) =$ | 0.04341 |
| $\mathrm{SE}(\mathrm{E}_{MC,SEmean}) =$ | 0.00137 |
| **Corrected $m$-out-of-$n$ bootstrap** | |
| $\mathrm{E}_{MC,SEmean} = \mathrm{E}(\mathrm{SE}^*_{mean}) =$ | 0.30160 |
| $\mathrm{SE}_{MC,SEmean} = \mathrm{SE}(\mathrm{SE}^*_{mean}) =$ | 0.02366 |
| $\mathrm{SE}(\mathrm{E}_{MC,SEmean}) =$ | 0.00075 |

| **The Median** | |
|---:|---|
| Theoretical value: $\mathrm{SE}(\mathrm{Med}(X_1, X_2, \ldots, X_n))=$ | 0.36447 |
| **Traditional bootstrap** | |
| $\mathrm{E}_{MC,SEmed} = \mathrm{E}(\mathrm{SE}^*_{med}) =$ | 0.37401 |
| $\mathrm{SE}_{MC,SEmed} = \mathrm{SE}(\mathrm{SE}^*_{med}) =$ | 0.08169 |
| $\mathrm{SE}(\mathrm{E}_{MC,SEmed}) =$ | 0.00258 |
| **Naive $m$-out-of-$n$ bootstrap** | |
| $\mathrm{E}_{MC,SEmed} = \mathrm{E}(\mathrm{SE}^*_{med}) =$ | 0.66450 |
| $\mathrm{SE}_{MC,SEmed} = \mathrm{SE}(\mathrm{SE}^*_{med}) =$ | 0.10057 |
| $\mathrm{SE}(\mathrm{E}_{MC,SEmed}) =$ | 0.00318 |
| **Corrected $m$-out-of-$n$ bootstrap** | |
| $\mathrm{E}_{MC,SEmed} = \mathrm{E}(\mathrm{SE}^*_{med}) =$ | 0.36216 |
| $\mathrm{SE}_{MC,SEmed} = \mathrm{SE}(\mathrm{SE}^*_{med}) =$ | 0.05481 |
| $\mathrm{SE}(\mathrm{E}_{MC,SEmed}) =$ | 0.00173 |

**Table 4.2:** *Standard errors of the traditional, naive $m$-out-of-$n$ and corrected $m$-out-of-$n$ bootstrap for the sample mean and sample median: Contaminated normal distribution with* $(\mu_1 = 0, \sigma_1^2 = 6.25, \mu_2 = \frac{1}{2}, \sigma_2^2 = 12.25, p = \frac{1}{2})$

We see from these simulation results that the naive $m$-out-of-$n$ bootstrap estimate of the standard error of the mean and median deviates greatly from the correct standard error. The traditional estimate and corrected $m$-out-of-$n$ bootstrap estimates, on the other hand are very close to the correct answer.

Surprisingly, for the corrected $m$-out-of-$n$ bootstrap estimated standard error of the sample median, we see that the Monte-Carlo standard error of this result is smaller than the standard error of the traditional bootstrap standard error estimate. However, this result is not seen when we consider the standard error of the sample mean. This is a promising result, but we will need to investigate further to see if this is a significant result or just a fluke of the Monte-Carlo procedure.

## 4.4 How to choose $m$

We will now briefly look at some of the techniques that have been suggested in the literature to select the value $m$ in the $m$-out-of-$n$ bootstrap.

### 4.4.1 A suggestion by Swanepoel (1986)

Swanepoel (1986) proved that the $m$-out-of-$n$ bootstrap will correct one of the non-regular cases outlined in Bickel and Freedman (1981). Additionally, this paper presents a Monte-Carlo study whereby values of $m$ are selected by calculating the coverage probability of a confidence interval for $\mu$ based on the $m$-out-of-$n$ bootstrap. The value of $m$ that produced a confidence interval with coverage probability closest to the specified nominal confidence level was deemed the 'best'.

The results of the Monte-Carlo study produced the following simple *rule-of-thumb*: $m = 2n/3$.

### 4.4.2 A suggestion by Götze & Račkauskas (2001)

Define $T_n = T_n(X_1, X_2, \ldots, X_n; F)$, a sequence of statistics, possibly dependent on an unknown distribution $F$. Let $L_n(F,x) = \mathrm{P}(T_n \leq x)$ and $L_m(F_n, x) = \mathrm{P}^*(T_m(X_1^*, X_2^*, \ldots, X_m^*; F_n) \leq x)$.

Götze and Račkauskas (2001) studied a rule first proposed by Bickel, Gotzë and van Zwet (personal communication) whereby the value of $m$ is selected by minimizing the 'distance' between the quantities $L_m(F_n, x)$ and $L_{\lfloor \frac{m}{2} \rfloor}(F_n, x)$, where $\lfloor x \rfloor$ is the largest integer less than or equal to $x$.

The motivation for this procedure is that the authors showed that the random distance between the quantities

$$L_m(F_n, x) \text{ and } L_{\lfloor \frac{m}{2} \rfloor}(F_n, x),$$

is *stochastically equivalent* to the random distance between

$$L_n(F, x) \text{ and } L_m(F_n, x).$$

#### A suggestion by Sakov (1998)

This method was also considered in a PhD thesis by Sakov (1998) (supervised by Peter J. Bickel). Sakov (1998), using a sequence of possible $m$ values $m_j = \lfloor q^j n \rfloor$, $j = 0, 1, 2, \ldots$, $0 < q < 1$, defines the choice of $m$ as follows:

$$\hat{m} = \arg\min_{m_j} \sup_x \left| L_{m_j}(F_n, x) - L_{m_{j+1}}(F_n, x) \right|,$$

where $L_m(F_n, x)$ is defined as above.

### 4.4.3 A suggestion by Chung & Lee (2001)

Chung and Lee (2001) worked on the choice of $m$ for the construction of bootstrap percentile confidence bounds. Their work is based on a combination of two of the bootstrap percentile confidence bounds, namely the hybrid and backwards bootstrap percentile confidence bounds discussed in Section 2.4.5. Their confidence upper bound is also based on $m$-out-of-$n$ bootstrap concepts and is denoted by

$$\begin{aligned} I(m, \delta, \alpha) &= (1 - \delta m^{1/2} n^{-1/2})\hat{\theta}_n + \delta m^{1/2} n^{-1/2} \hat{u}_{\{1+\delta(2\alpha-1)\}/2, m} \\ &= \begin{cases} \hat{\theta}_n - \left(\frac{m}{n}\right)^{1/2} (\hat{u}_{1-\alpha,m} - \hat{\theta}_n) & \text{if } \delta = -1 \\ \hat{\theta}_n + \left(\frac{m}{n}\right)^{1/2} (\hat{u}_{\alpha,m} - \hat{\theta}_n) & \text{if } \delta = 1, \end{cases} \end{aligned} \tag{4.8}$$

where $m$ is the resample size used in the bootstrap procedure to calculate $I(m, \delta, \alpha)$ and is defined as $m = \gamma n$ where $0 < \gamma \leq 1$. Here $\delta$ is a parameter which determines whether the bound calculated is a hybrid or backwards percentile bound, $\alpha$ is the specified nominal coverage probability and the quantile $\hat{u}_{\alpha,m}$ is defined as

$$\mathrm{P}^*(\hat{\theta}^*_m \leq \hat{u}_{\alpha,m}) = \alpha.$$

Through the use of the smooth function model, the authors provide the following expansion of the coverage probability:

$$\mathrm{P}(\theta \leq I(m, \delta, \alpha)) = \alpha + n^{-1/2}\left\{\left(\frac{m}{n}\right)^{-1/2} \delta p_1(z(\alpha)) - q_1(z(\alpha))\right\}\phi(z(\alpha)) + O(n^{-1}),$$

where $z(\alpha) = \Phi^{-1}(\alpha)$, $\Phi(\cdot)$ is the standard normal distribution function, and $\phi(\cdot)$ is the standard normal density function. The polynomials $p_1$ and $q_1$ are obtained from the Edgeworth expansions of the distribution of the standardized and studentized version of $\hat{\theta}_n$ respectively. Their estimated forms, $\hat{p}_1$ and $\hat{q}_1$, appear in the article, but will not be presented here. To reduce the coverage error for this percentile confidence bound to $O(n^{-1})$ they select $(m/n)^{-1/2}\,\delta = q_1(z(\alpha))/p_1(z(\alpha))$, or equivalently $\gamma = (\delta p_1(z(\alpha))/q_1(z(\alpha)))^2$, where $\gamma = m/n$. This choice of $\gamma$ leads directly to the optimal choice of $m$.

They show (through a lengthy proof) that if the following estimates are used

$$\hat{m}_1 = \left\lfloor n\left\{\frac{\hat{p}_1(z(\alpha))}{\hat{q}_1(z(\alpha))}\right\}^2\right\rfloor \qquad \text{and} \qquad \hat{\delta}_1 = -\,\mathrm{sign}\left(\frac{\hat{p}_1(z(\alpha))}{\hat{q}_1(z(\alpha))}\right),$$

then the confidence bound $I(\hat{m}_1, \hat{\delta}_1, \alpha)$ has coverage probability error $O(n^{-1})$.

In practice if one does not have knowledge of the explicit forms of the polynomials $\hat{p}_1$ and $\hat{q}_1$, then one can still approximate the choice of $m$ and $\delta$ using a double bootstrap algorithm presented below. The algorithm approximates $\hat{m}_1$ and $\hat{\delta}_1$ and these approximations are denoted by $\hat{m}_2$ and $\hat{\delta}_2$ respectively. Chung and Lee (2001) state that these approximations are asymptotically equivalent to the estimates $\hat{m}_1$ and $\hat{\delta}_1$, and that the intervals constructed using these values will have the same coverage probability error as the original estimates, i.e., the interval $I(\hat{m}_2, \hat{\delta}_2, \alpha)$ also has coverage probability error $O(n^{-1})$.

The algorithm below employs a double bootstrap procedure and the following double bootstrap version of the expanded coverage probability

$$\mathrm{P}^*(\hat{\theta}_n \leq I^*(k, 1, \alpha)) = \alpha - m^{-\frac{1}{2}}\left\{\gamma^{-\frac{1}{2}}\hat{p}_1(z(\alpha)) + \hat{q}_1(z(\alpha))\right\}\phi(z(\alpha)) + O_p(m^{-1}). \tag{4.9}$$

where $\gamma = k/m$, $I^*(k, 1, \alpha) = \hat{\theta}^*_m + (k/m)^{1/2}(\hat{u}^*_{\alpha,k} - \hat{\theta}^*_m)$ and $\hat{u}^*_{\alpha,k}$ is defined as the value that satisfies the following expression

$$\mathrm{P}^{**}(\hat{\theta}^{**}_k \leq \hat{u}^*_{\alpha,k}) = \alpha.$$

In equation (4.9) they choose $\delta = 1$ because the resulting estimates are asymptotically equivalent to the case where $\delta = -1$.

**Approximating the optimal $m$ for percentile confidence bounds using the method of Chung and Lee (2001)**

1. Select a fixed $m_0 < n$. Also select $t$ possible values for $\gamma$: $\gamma_1, \gamma_2, \ldots, \gamma_t$ where $0 < \gamma_i \leq 1$. Next define double bootstrap resample sizes $k_1, k_2, \ldots, k_t$ as $k_i = \gamma_i m_0$.

2. For each $\gamma_i$ calculate $\tilde{\alpha}_i$, where

$$\tilde{\alpha}_i := \mathrm{P}^* \left( \hat{\theta}_n \leq I^*(k_i, \delta, \alpha) \right). \tag{4.10}$$

   The steps for obtaining the values $\tilde{\alpha}_1, \tilde{\alpha}_2, \ldots, \tilde{\alpha}_t$ are:

   a. Given $X_1, X_2, \ldots, X_n$, calculate $\hat{\theta}_n = \hat{\theta}_n(X_1, X_2, \ldots, X_n)$.
   b. Sample with replacement from $X_1, X_2, \ldots, X_n$ to get $X_1^*, X_2^*, \ldots, X_{m_0}^*$. Calculate $\hat{\theta}_{m_0}^* = \hat{\theta}_{m_0}(X_1^*, X_2^*, \ldots, X_{m_0}^*)$.
   c. Sample with replacement from $X_1^*, X_2^*, \ldots, X_{m_0}^*$ to get $X_1^{**}, X_2^{**}, \ldots, X_{k_i}^{**}$. Calculate $\hat{\theta}_{k_i}^{**} = \hat{\theta}_{k_i}(X_1^{**}, X_2^{**}, \ldots, X_{k_i}^{**})$.
   d. Repeat step 2c $R$ times to get $\hat{\theta}_{k_i,1}^{**}, \hat{\theta}_{k_i,2}^{**}, \ldots, \hat{\theta}_{k_i,R}^{**}$. Order these replications to get $\hat{\theta}_{k_i,(1)}^{**} \leq \hat{\theta}_{k_i,(2)}^{**} \leq \cdots \leq \hat{\theta}_{k_i,(R)}^{**}$.
   e. Approximate the value $\hat{u}_{\alpha,k_i}^*$ with the value $\hat{\theta}_{k_i,(\lfloor R\cdot\alpha \rfloor)}^{**}$.
   f. Repeat 2b to 2e $B$ times to obtain the following quantities:

$$\begin{gathered} \hat{\theta}_{m_0,1}^*, \hat{\theta}_{m_0,2}^*, \ldots, \hat{\theta}_{m_0,B}^* \\ \hat{u}_{\alpha,k_i,1}^*, \hat{u}_{\alpha,k_i,2}^*, \ldots, \hat{u}_{\alpha,k_i,B}^* \\ I_1^*, I_2^*, \ldots, I_B^* \end{gathered}$$

   where $I_b^* = \hat{\theta}_{m_0,b}^* + (k_i/m_0)^{1/2}(\hat{u}_{\alpha,k_i,b}^* - \hat{\theta}_{m_0,b}^*)$.
   g. Finally, approximate $\alpha$ with

$$\tilde{\alpha}_i = \frac{1}{B} \sum_{b=1}^{B} \mathrm{I}(\hat{\theta}_n \leq I_b^*)$$

   where $\mathrm{I}(\cdot)$ is the indicator function.

3. Calculate step 2 for all $t$ chosen values of $\gamma_i$. We now have the following values: $\gamma_1, \gamma_2, \ldots, \gamma_t$ and $\tilde{\alpha}_1, \tilde{\alpha}_2, \ldots, \tilde{\alpha}_t$.

4. We now need to find that value $\gamma_i$ that makes $\tilde{\alpha}_i \approx \alpha$. We can argue from the expansion given in equation (4.9) and the definition of $\tilde{\alpha}_i$ given in (4.10) that the quantities $(\alpha - \tilde{\alpha}_i)$ and $\gamma_i^{-1/2}$ have an approximate linear relationship, i.e.,

$$\alpha - \tilde{\alpha}_i = C + K\gamma_i^{-1/2},$$

   where $C$ and $K$ are parameters which will need to be estimated.

Using the Ordinary Least Squares equations we can estimate $K$ and $C$ by

$$\hat{K} = \frac{\sum_{i=1}^{t}(\alpha - \tilde{\alpha}_i)\gamma_i^{-1/2} - \frac{1}{t}\sum_{i=1}^{t}(\alpha - \tilde{\alpha}_i)\sum_{i=1}^{t}\gamma_i^{-1/2}}{\sum_{i=1}^{t}\gamma_i^{-1} - \frac{1}{t}\left(\sum_{i=1}^{t}\gamma_i^{-1/2}\right)^2},$$

and

$$\hat{C} = \frac{1}{t}\sum_{i=1}^{t}(\alpha - \tilde{\alpha}_i) - \hat{K}\frac{1}{t}\sum_{i=1}^{t}\gamma_i^{-1/2}.$$

The optimal value of $\gamma_i$, denoted $\hat{\gamma}$, is obtained when $\alpha - \tilde{\alpha}_i \approx 0$. Using the above linear least squares approximation we find that

$$\hat{\gamma} = \left(\frac{\hat{K}}{\hat{C}}\right)^2.$$

Therefore,

$$\hat{m}_2 = \left\lfloor \left(\frac{\hat{K}}{\hat{C}}\right)^2 n \right\rfloor = \lfloor \hat{\gamma} n \rfloor \qquad \text{and} \qquad \hat{\delta}_2 = -\operatorname{sign}\left(\frac{\hat{K}}{\hat{C}}\right).$$

### 4.4.4 A suggestion by Samworth (2003)

In Samworth (2003), a class of data-based choices of $m$ are discussed for Hodges' super-efficient estimator (given in Section 3.4). This class is given by

$$\hat{m} = \begin{cases} An^{\alpha}, & \text{if } |\bar{X}_n| \leq Bn^{-\beta} \\ n, & \text{if } |\bar{X}_n| > Bn^{-\beta}, \end{cases}$$

where $A > 0$, $B > 0$, $0 < \alpha < 1$ and $0 < \beta < \frac{1}{2}$. A drawback of this procedure is that $A$, $B$, $\alpha$ and $\beta$ are all parameters which need to be estimated, but no estimators are proposed in the paper. Putter and van Zwet (1996) showed, in their Corollary 1.1, that bootstrap estimators based on $\hat{m}$ are consistent.

### 4.4.5 A suggestion by Cheung & Lee (2005)

Cheung and Lee (2005) attempted to determine the best $m$ when the $m$-out-of-$n$ bootstrap is applied to the estimation of the variance of a sample quantile. The method employs an exact expression for the bootstrap estimate of standard error of the sample quantiles (Hall and Martin 1988). The algorithm used here is fairly convoluted, and uses an $l$-out-of-$m$ bootstrap sampling scheme nested within an $m$-out-of-$n$ bootstrap sampling scheme to determine the value of $\hat{m}$. The basic idea employed in the method is to obtain sample sizes that minimise estimated mean squared error (MSE) values. Fortunately the exact expression for the variance simplifies the algorithm somewhat and only one 'level' of sampling is necessary to perform a double bootstrap. The choice of $m$ is based on the following form: $m = cn^{\gamma}$, $c > 0$ and $\gamma > 0$.

**Approximating $m$ for variance estimation of the $p^{\text{th}}$ sample quantiles using the method of Cheung and Lee (2005)**

1. Given $X_1, X_2, \ldots, X_n$, fix a set of resample sizes $m_1 < m_2 < \cdots < m_S < n$ where $S \geq 2$.

2. For a chosen value $s$, use the explicit expression for the variance of the $p^{\text{th}}$ $m$-out-of-$n$ bootstrap sample quantile and calculate

$$\sigma_s^{*2} = \left(\frac{n}{m_s}\right) \hat{\sigma}_{m_s}^2$$

   where

$$\hat{\sigma}_{m_s}^2 = \left(\frac{m_s}{n}\right) \sum_{j=1}^{n} (X_{(j)} - X_{(r)})^2 w_{m_s,j},$$

$$w_{m_s,j} = k \binom{m_s}{k} \int_{\frac{j-1}{n}}^{\frac{j}{n}} x^{k-1}(1-x)^{m_s-k} dx,$$

   $k = \lfloor m_s p \rfloor + 1$ and $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ are the order statistics of the original sample.

3. Sample with replacement from $X_1, X_2, \ldots, X_n$ using sample size $m_s$ to obtain the sample $X_1^*, X_2^*, \ldots, X_{m_s}^*$. Sorting this sample produces

$$X_{(1)}^* \leq X_{(2)}^* \leq \cdots \leq X_{(m_s)}^*.$$

4. Select a value $l$ from the set $\{1, 2, \ldots, m_s\}$. Calculate the $l$-out-of-$m_s$ double bootstrap variance $\sigma_{s,l}^{**2}$ using the following explicit expression (there is no need to generate a double bootstrap sample):

$$\sigma_{s,l}^{**2} = \left(\frac{l}{m_s}\right) \sum_{j=1}^{m_s} \left(X_{(j)}^* - X_{(r_s)}^*\right)^2 k_l \binom{l}{k_l} \int_{\frac{j-1}{m_s}}^{\frac{j}{m_s}} x^{k_l-1}(1-x)^{l-k_l} dx,$$

   where $k_l = \lfloor lp \rfloor + 1$ and $r_s = \lfloor m_s p \rfloor + 1$.

5. Repeat steps 3 to 4 $B$ times to obtain

$$\sigma_{s,l,1}^{**2}, \sigma_{s,l,2}^{**2}, \ldots, \sigma_{s,l,B}^{**2},$$

   and calculate

$$\text{MSE}_s(l) = \frac{1}{B} \sum_{b=1}^{B} (\sigma_{s,l,b}^{**2} - \sigma_s^{*2})^2.$$

6. Repeat steps 4 to 5 for each possible value of $l \in \{1, 2, \ldots, m_s\}$, i.e., obtain

$$\text{MSE}_s(1), \text{MSE}_s(2), \ldots, \text{MSE}_s(m_s).$$

Find the value $l_s$ that minimizes $\text{MSE}_s(l)$ for $l \in \{1, 2, \dots, m_s\}$, i.e.,

$$l_s = \arg \min_{l \in \{1,2,\dots,m_s\}} (\text{MSE}_s(l)) .$$

7. Repeat steps 2 to 6 for all choices of $m_s$, $s = 1, 2, \dots, S$, i.e., we now have

$$m_1, m_2, \dots, m_S \quad \text{and} \quad l_1, l_2, \dots, l_S.$$

8. Using the pairs $(m_1, l_1), (m_2, l_2), \dots, (m_S, l_S)$ and the approximate relationship

$$l_s \approx c m_s^{\gamma},$$

to approximate the parameters $c$ and $\gamma$ using Ordinary Least Squares techniques. We start by noting that

$$y_s' \approx a' + \gamma x_s',$$

where $y_s' = \log(l_s)$, $a' = \log(c)$ and $x_s' = \log(m_s)$. The estimates of $a'$ and $\gamma$ are then given by

$$\hat{\gamma} = \frac{\sum_{s=1}^{S} x_s' y_s' - \frac{1}{S} \sum_{s=1}^{S} x_s' \sum_{s=1}^{S} y_s'}{\sum_{s=1}^{S} x_s'^2 - \frac{1}{S} \left( \sum_{s=1}^{S} x_s' \right)^2},$$

and

$$\hat{a}' = \frac{1}{S} \sum_{s=1}^{S} y_s' - \hat{\gamma} \frac{1}{S} \sum_{s=1}^{S} x_s'.$$

The estimate of $c$ can then be written as $\hat{c} = e^{\hat{a}'}$.

9. Finally, the optimal value of $m$, denoted $\hat{m}$, is given by

$$\hat{m} = \hat{c} n^{\hat{\gamma}}.$$

**Remark:**
In the following chapters we will develop our own data-based methods of choosing the bootstrap sample size $m$ for two new problems, viz. point estimation of a parameter, and bootstrap hypothesis testing.

# Chapter 5

# A nonparametric point estimation technique using the $m$-out-of-$n$ bootstrap

## 5.1 Introduction

The bootstrap has been used in a wide variety of areas of estimation, from estimating standard errors and bias of statistics, to estimating confidence bounds for parameters. In this chapter we will investigate a bootstrap method which can be used to improve an existing point estimate by a modification of the statistic and by using the corrected $m$-out-of-$n$ bootstrap methods already discussed.

The methods which will be used have come to be known as bootstrap aggregating (or BAGGing) and bootstrap robust aggregating (or BRAGGing). The name BAGGing was coined in the field of machine learning by Breiman (1994), but the concept applied to statistical point estimates is slightly older, having first appeared in Swanepoel (1988) and Swanepoel (1990) (these papers referred to the technique as *an approximating functional approach*). In recent years this topic has been the recipient of renewed interest with articles concerning BAGGing being published by Bühlmann (2002), Buja and Stuetzle (2006) and Croux, Joossens and Lemmens (2007), and articles concerning BRAGGing being published by Bühlmann (2003) and Berrendero (2007).

In this discussion we will primarily focus on the work done concerning BRAGGing which appears in Swanepoel (1988) and Berrendero (2007). In particular we will take the view held by Swanepoel (1988) which enables a more general approach. We will look at the naive BRAGGing estimate, but we will also investigate three other statistics based on *corrected* $m$-out-of-$n$ bootstrap concepts. The choice of which of these four statistics is 'best' will be determined based on a set of pilot Monte-Carlo simulations. Ultimately, since these estimation techniques are based on $m$-out-of-$n$ bootstrap ideas, we will be interested in a data-based choice of the resample size $m$. This chapter will conclude with a proposed data-based choice of the resample size in this situation.

## 5.2 BAGGing and BRAGGing

The definition of a BAGGing and BRAGGing point estimator of a parameter will now be discussed. Let $\theta$ be a parameter of interest which can be expressed as some functional $t$ of an unknown distribution $F$ as

$$\theta = t(F).$$

Suppose also that $t(F)$ can be approximated by a sequence of functionals $t_m(F)$, i.e.,

$$t_m(F) \approx t(F),$$

with the approximation becoming increasingly accurate as $m \to \infty$. The proposed estimator for $\theta$ makes use of this *smoothed* functional sequence approximation and the empirical distribution function $F_n$ to create the plug-in expression

$$\tilde{\theta}_{n,m} = t_m(F_n).$$

As discussed in Swanepoel (1990) and Berrendero (2007), two possibilities for this estimator are

- ⋆ $\tilde{\theta}_{n,m,bag} = t_m(F_n) = \mathrm{E}^*(\hat{\theta}^*_m)$, where $\hat{\theta}^*_m = \hat{\theta}_m(X^*_1, X^*_2, \ldots, X^*_m)$ and $\hat{\theta}_n = \hat{\theta}_n(X_1, X_2, \ldots, X_n) = t(F_n)$. The estimator $\tilde{\theta}_{n,m,bag}$ is known as the BAGGing estimator in the literature.
- ⋆ $\tilde{\theta}_{n,m,brag} = t_m(F_n) = \mathrm{Med}^*(\hat{\theta}^*_m)$. This estimator is known as the BRAGGing estimator in the literature because of the use of the median which is considered to be more robust.

Both techniques are defined in such a way that one can easily implement a simple Monte-Carlo algorithm to approximate their values. For example, the Monte-Carlo algorithm for approximating the BAGGing estimator would be as follows:

**BAGGing algorithm:**

1. Sample with replacement from $X_1, X_2, \ldots, X_n$ to obtain a bootstrap sample of size $m$, $X^*_1, X^*_2, \ldots, X^*_m$.
2. Calculate the bootstrap statistic $\hat{\theta}^*_m = \hat{\theta}_m(X^*_1, X^*_2, \ldots, X^*_m)$.
3. Repeat the previous two steps $B$ times to obtain the bootstrap replications $\hat{\theta}^*_m(1), \hat{\theta}^*_m(2), \ldots, \hat{\theta}^*_m(B)$.
4. Calculate:

$$\tilde{\theta}_{n,m,bag,B} = \frac{1}{B}\sum_{b=1}^{B} \hat{\theta}^*_m(b).$$

The algorithm for the BRAGGing estimator follows in a similar fashion except that the last step is replaced with the calculation of the sample median of the statistics.

**BRAGGing algorithm:**

4. Calculate:

$$\tilde{\theta}_{n,m,brag,B} = \mathrm{Med}(\hat{\theta}^*_m(1), \ldots, \hat{\theta}^*_m(B)).$$

Our primary focus in this thesis will lie with the BRAGGing estimator, since it has been shown in the literature to improve an estimation procedure (e.g., see Bühlmann (2003)).

## 5.3 Variants of the BRAGGing estimator

The BRAGGing estimate discussed in the previous two sections was based on the bootstrap world's version of the median of the bootstrap statistic $\hat{\theta}^*_m$. If one employs the *corrected* $m$-out-of-$n$

bootstrap ideas discussed in Section 4.3 then it is possible to derive three new versions of this BRAGGing estimator.

We will now show how these estimators are defined and later we will show (through the use of a basic Monte-Carlo study) which of these estimators are preferable in certain situations.

- ⋆ **The original version of the estimator:** To distinguish between the new BRAGGing estimators and the original one we will adopt a new notation for these estimators. Let the original BRAGGing estimator be renamed as

$$\tilde{\theta}_{brag,1} := \tilde{\theta}_{n,m,brag}.$$

- ⋆ **The first new version of the estimator:** The first new version of the BRAGGing estimator, denoted by $\tilde{\theta}_{brag,2}$, is derived by first noting that the estimator $\hat{\theta}_n$ can be written using the corrected $m$-out-of-$n$ form given in (4.4) (we select $\alpha = 0.5$ without loss of generality)

$$\hat{\theta}_n = \frac{1}{\sqrt{n}}\sqrt{n}(\hat{\theta}_n - \theta) + \theta.$$

Applying the corrected $m$-out-of-$n$ bootstrap (CMOON) to the median of the above expression we get the following estimator

$$\begin{aligned}\tilde{\theta}_{brag,2} &:= \frac{1}{\sqrt{n}}\,\text{Med}^*\left(\sqrt{m}\left(\hat{\theta}^*_m - \hat{\theta}_n\right)\right) + \hat{\theta}_n \\ &= \sqrt{\frac{m}{n}}\,\text{Med}^*\left(\hat{\theta}^*_m\right) + \left(1 - \sqrt{\frac{m}{n}}\right)\hat{\theta}_n \\ &= \sqrt{\frac{m}{n}}\tilde{\theta}_{brag,1} + \left(1 - \sqrt{\frac{m}{n}}\right)\hat{\theta}_n\end{aligned}$$

Note that this CMOON estimator is a convex combination between the original BRAGGing estimator, $\tilde{\theta}_{brag,1}$, and the estimator $\hat{\theta}_n$.

- ⋆ **The second new version of the estimator:** The second new version of the BRAGGing estimator, denoted by $\tilde{\theta}_{brag,3}$, is similar to the previous one, except that it is based on the assumption that the distribution of $\sqrt{n}(\hat{\theta}_n - \theta)$ is symmetric. This means that the derivation deviates from the previous derivation as follows: We first note that, due to symmetry, we can write

$$\text{Med}(\hat{\theta}_n) = \frac{1}{\sqrt{n}}\,\text{Med}\left(\sqrt{n}(\hat{\theta}_n - \theta)\right) + \theta = \frac{1}{\sqrt{n}}\,\text{Med}\left(-\sqrt{n}(\hat{\theta}_n - \theta)\right) + \theta.$$

Therefore, the CMOON bootstrap estimator of the above quantity is

$$\begin{aligned}\tilde{\theta}_{brag,3} &:= \frac{1}{\sqrt{n}}\,\text{Med}^*\left(-\sqrt{m}\left(\hat{\theta}^*_m - \hat{\theta}_n\right)\right) + \hat{\theta}_n \\ &= -\sqrt{\frac{m}{n}}\tilde{\theta}_{brag,1} + \left(1 + \sqrt{\frac{m}{n}}\right)\hat{\theta}_n.\end{aligned}$$

- ⋆ **The third new version of the estimator:** The third and final version of the estimator, denoted by $\tilde{\theta}_{brag,4}$, is simply an amalgamation of the previous two. It is written as

$$\tilde{\theta}_{brag,4} := \delta\sqrt{\frac{m}{n}}\tilde{\theta}_{brag,1} + \left(1 - \delta\sqrt{\frac{m}{n}}\right)\hat{\theta}_n$$

$$= \begin{cases} \tilde{\theta}_{brag,2}, & \text{if } \delta = 1 \\ \tilde{\theta}_{brag,3}, & \text{if } \delta = -1. \end{cases}$$

The choice of $\delta$ will be discussed in the next section.

**Remark:**

- ⋆ Note that the estimators $\tilde{\theta}_{brag,2}$ and $\tilde{\theta}_{brag,3}$ appear to be stochastically equivalent, and that it may not seem necessary to consider both of them. However, there two estimators are used in the construction of the estimator $\tilde{\theta}_{brag,4}$, and will thus also be employed.
- ⋆ The estimator $\tilde{\theta}_{brag,4}$ is highly reminiscent of the percentile confidence bound discussed in Chung and Lee (2001). In fact, $\tilde{\theta}_{brag,4}$ is equivalent to the upper 0.5-level of the $m$-out-of-$n$ bootstrap percentile confidence bound given in (4.8).

### 5.3.1 Some considerations for $\tilde{\theta}_{brag,4}$

For the practical application of the last statistic the value of $\delta$ needs to be approximated, i.e., one needs to determine whether or not the distribution of $\sqrt{n}(\hat{\theta}_n - \theta)$ is symmetric or not before deciding on a value for $\delta$.

An elementary choice of $\delta$ involves applying a hypothesis test for skewness using the bootstrap replications $\hat{\theta}^*_m(1), \dots, \hat{\theta}^*_m(B)$ as the observed set of sample data. One possible test for testing for skewness is the $\sqrt{b_1}$ test described next.

#### The $\sqrt{b_1}$ test for skewness

The following is a test for testing the hypothesis

$$H_0 : F \text{ is symmetric vs. } H_A : F \text{ not symmetric.}$$

The $\sqrt{b_1}$ test for symmetry, as described in Ngatchou-Wandji (2006), is based on the sample version of the skewness coefficient $W_n$ defined as

$$W_n = \frac{\hat{\mu}_3}{\hat{\mu}_2^{3/2}},$$

where $\hat{\mu}_k$ is the $k^{\text{th}}$ sample central moment defined as

$$\hat{\mu}_k = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^k. \tag{5.1}$$

If we define $\eta^2$ as

$$\eta^2 = \frac{\mu_6 - 6\mu_2\mu_4 + 9\mu_2^3}{\mu_2^3},$$

and if the sample $X_1, X_2, \dots, X_n$ was generated from a symmetric distribution (i.e., under $H_0$) with the property that $\mathrm{E}(X_i^6) < \infty$, then

$$\frac{\sqrt{n}W_n}{\eta} \xrightarrow{\mathcal{D}} Z, \text{ as } n \to \infty,$$

where $Z$ is a $\mathrm{N}(0,1)$ random variable. Estimate $\eta^2$ by its sample version

$$\hat{\eta}^2 = \frac{\hat{\mu}_6 - 6\hat{\mu}_2\hat{\mu}_4 + 9\hat{\mu}_2^3}{\hat{\mu}_2^3},$$

then, by using Kolmogorov's Strong Law of Large Numbers and Slutsky's Theorem, we obtain that

$$SQ_n := \frac{\sqrt{n}W_n}{\hat{\eta}} \xrightarrow{\mathcal{D}} Z, \text{ as } n \to \infty \text{ under } H_0,$$

where $Z$ is a $\mathrm{N}(0,1)$ random variable.

### Application of the $\sqrt{b_1}$ test to determine $\delta$

We will now briefly describe the algorithm that was used to determine the value of $\delta$ in the Monte-Carlo simulations which will be presented in the next section. To determine whether the value of $\delta$ in the statistic $\tilde{\theta}_{brag,4}$ should be equal to 1 or $-1$ we need to test the skewness of the distribution of $\sqrt{n}(\hat{\theta}_n - \theta)$ using a data-dependent approach. Since we do not know the true distribution of this statistic, we will make use of the bootstrap replications $\hat{\theta}^*_m(1), \ldots, \hat{\theta}^*_m(B)$, obtained through a simple $m$-out-of-$n$ bootstrap Monte-Carlo simulation.

The replications $\hat{\theta}^*_m(1), \ldots, \hat{\theta}^*_m(B)$ are then treated as the data points $X_1, X_2, \ldots, X_n$ and used in the calculation of the test statistic $SQ_n$ (as described above). If the absolute value of $SQ_n$ is greater than a specified standard normal quantile, say $z\left(1-\frac{\alpha}{2}\right)$, where $z(x) = \Phi^{-1}(x)$, then we will reject the null hypothesis of symmetry (concluding skewness) and set $\delta = 1$. If the absolute value of $SQ_n$ is less than $z\left(1-\frac{\alpha}{2}\right)$ then we will conclude symmetry and set $\delta = -1$.

The algorithm is then:

**Determining $\delta$**

1. Sample with replacement from $X_1, X_2, \ldots, X_n$ to obtain a bootstrap sample of size $m$, $X^*_1, X^*_2, \ldots, X^*_m$.
2. Calculate the bootstrap statistic $\hat{\theta}^*_m = \hat{\theta}_m(X^*_1, X^*_2, \ldots, X^*_m)$.
3. Repeat the previous two steps $B$ times to obtain the bootstrap replications $\hat{\theta}^*_m(1), \hat{\theta}^*_m(2), \ldots, \hat{\theta}^*_m(B)$.
4. Calculate:
$$\tilde{SQ}_B = \frac{\sqrt{B}\tilde{W}_B}{\tilde{\eta}_B}$$
where
$$\tilde{\eta}^2_B = \frac{\tilde{\mu}_6 - 6\tilde{\mu}_2\tilde{\mu}_4 + 9\tilde{\mu}^3_2}{\tilde{\mu}^3_2},$$
$$\tilde{W}_B = \frac{\tilde{\mu}_3}{\tilde{\mu}_2^{3/2}},$$
$$\tilde{\mu}_k = \frac{1}{B}\sum_{b=1}^{B}\left(\hat{\theta}^*_m(b) - \hat{\theta}^*_m(\cdot)\right)^k.$$
and
$$\hat{\theta}^*_m(\cdot) = \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}^*_m(b).$$
5. If $\left|\tilde{SQ}_B\right| > z\left(1-\frac{\alpha}{2}\right)$ then $\delta = 1$, else, if $\left|\tilde{SQ}_B\right| \leq z\left(1-\frac{\alpha}{2}\right)$ then $\delta = -1$.

## 5.4 Monte-Carlo study: Determining the "best" BRAGGing estimate

In this section we present a pilot Monte-Carlo study which attempts to determine which of the four methods is best when applied to the estimation of the population mean $\mu$. The Monte-Carlo method will be based on the repeated calculation of the following statistic:

$$\zeta(\tilde{\theta}_{brag}) = \frac{\text{MSE}(\bar{X}_n)}{\text{MSE}(\tilde{\theta}_{brag})},$$

where $\tilde{\theta}_{brag}$ is any one of the estimators $\tilde{\theta}_{brag,1}$, $\tilde{\theta}_{brag,2}$, $\tilde{\theta}_{brag,3}$ or $\tilde{\theta}_{brag,4}$. The purpose of calculating the statistic $\zeta(\tilde{\theta}_{brag})$ is to compare the relative performance of the BRAGGing estimator to the traditional sample mean for the estimation of the population mean. Large values of $\zeta(\tilde{\theta}_{brag})$ will indicate that the BRAGGing estimator outperforms the sample mean in the sense that its standard error or the bias (or both) are smaller than the corresponding quantities of the sample mean.

The configurations used in this Monte-Carlo study for the BRAGGing estimators $\tilde{\theta}_{brag,1}$, $\tilde{\theta}_{brag,2}$ and $\tilde{\theta}_{brag,3}$ are:

- ⋆ Sample sizes: $n = 10, 30, 50, 100, 300$.
- ⋆ Resample sizes (for the calculation of $\tilde{\theta}_{brag,1}$): $m = \frac{n}{3}, \frac{2n}{3}, n$. In this example we will use fixed choices for the resample sizes.
- ⋆ Monte Carlo number of iterations is 100 000.
- ⋆ Bootstrap number of iterations for the calculation of the BRAGGing estimator is 5 000.
- ⋆ Distributions:
  - Normal with $\mu = 0$ and $\sigma = 1$,
  - Contaminated normal with $\mu_1 = \mu_2 = 0$, $\sigma_1 = 1$, $\sigma_2 = 6$ and $p = 0.05$,
  - Contaminated normal with $\mu_1 = \mu_2 = 0$, $\sigma_1 = 1$, $\sigma_2 = 6$ and $p = 0.1$,
  - Double exponential with $\mu = 0$ and $\sigma = \sqrt{2}$,
  - F with 8 and 5 degrees of freedom,

The configuration for the calculation of $\tilde{\theta}_{brag,4}$ is the same as above except that the Monte-Carlo number of iterations was taken to be equal to 30 000.

**The output for the statistics $\zeta(\tilde{\theta}_{brag,1})$ , $\zeta(\tilde{\theta}_{brag,2})$, $\zeta(\tilde{\theta}_{brag,3})$ and $\zeta(\tilde{\theta}_{brag,4})$**

| | |
|---|---|
| **Distribution used** | **Normal** |
| **Parameters used** | $\mu = 0$ and $\sigma = 1$ |

| | $n = 10$ | | | $n = 30$ | | |
|---|---|---|---|---|---|---|
| | $m = \frac{n}{3}$ | $m = \frac{2n}{3}$ | $m = n$ | $m = \frac{n}{3}$ | $m = \frac{2n}{3}$ | $m = n$ |
| $\zeta(\tilde{\theta}_{brag,1})$ | 0.9775 | 0.9962 | 0.9971 | 1.0010 | 0.9883 | 0.9902 |
| $\zeta(\tilde{\theta}_{brag,2})$ | 1.0006 | 0.9936 | 0.9977 | 1.0002 | 1.0046 | 1.0007 |
| $\zeta(\tilde{\theta}_{brag,3})$ | 0.9904 | 0.9895 | 0.9943 | 0.9890 | 1.0091 | 1.0012 |
| $\zeta(\tilde{\theta}_{brag,4})$ | 1.0112 | 0.9852 | 1.0026 | 0.9969 | 1.0000 | 0.9972 |
| | $n = 50$ | | | $n = 100$ | | |
| | $m = \frac{n}{3}$ | $m = \frac{2n}{3}$ | $m = n$ | $m = \frac{n}{3}$ | $m = \frac{2n}{3}$ | $m = n$ |
| $\zeta(\tilde{\theta}_{brag,1})$ | 0.9985 | 1.0007 | 1.0045 | 1.0034 | 1.0033 | 0.9983 |
| $\zeta(\tilde{\theta}_{brag,2})$ | 1.0034 | 0.9923 | 1.0041 | 0.9980 | 1.0060 | 1.0002 |
| $\zeta(\tilde{\theta}_{brag,3})$ | 0.9942 | 1.0029 | 0.9967 | 0.9991 | 0.9877 | 0.9972 |
| $\zeta(\tilde{\theta}_{brag,4})$ | 1.0061 | 0.9976 | 1.0035 | 0.9877 | 0.9938 | 0.9967 |

**Table 5.1:** *The ratio of the MSE of the sample mean $\bar{X}_n$ to the MSE of each of the four estimators, $\tilde{\theta}_{brag,1}$, $\tilde{\theta}_{brag,2}$, $\tilde{\theta}_{brag,3}$ and $\tilde{\theta}_{brag,4}$.*

| | |
|---|---|
| **Distribution used** | **Contaminated Normal** |
| **Parameters used** | $\mu_1 = \mu_2 = 0$, $\sigma_1 = 1$, $\sigma_2 = 6$, $p = 0.05$ |

| | $n = 10$ | | | $n = 30$ | | |
|---|---|---|---|---|---|---|
| | $m = \frac{n}{3}$ | $m = \frac{2n}{3}$ | $m = n$ | $m = \frac{n}{3}$ | $m = \frac{2n}{3}$ | $m = n$ |
| $\zeta(\tilde{\theta}_{brag,1})$ | 1.9355 | 1.4320 | 1.0887 | 1.4722 | 1.1777 | 1.1133 |
| $\zeta(\tilde{\theta}_{brag,2})$ | 1.4967 | 1.3334 | 1.0905 | 1.2707 | 1.1460 | 1.1048 |
| $\zeta(\tilde{\theta}_{brag,3})$ | 0.6690 | 0.7181 | 0.9162 | 0.7766 | 0.8633 | 0.9109 |
| $\zeta(\tilde{\theta}_{brag,4})$ | 1.4829 | 1.3569 | 1.0827 | 1.2821 | 1.1381 | 1.1114 |
| | $n = 50$ | | | $n = 100$ | | |
| | $m = \frac{n}{3}$ | $m = \frac{2n}{3}$ | $m = n$ | $m = \frac{n}{3}$ | $m = \frac{2n}{3}$ | $m = n$ |
| $\zeta(\tilde{\theta}_{brag,1})$ | 1.3513 | 1.1386 | 1.0954 | 1.1850 | 1.0804 | 1.0520 |
| $\zeta(\tilde{\theta}_{brag,2})$ | 1.1931 | 1.1175 | 1.0804 | 1.0978 | 1.0697 | 1.0582 |
| $\zeta(\tilde{\theta}_{brag,3})$ | 0.9948 | 0.9976 | 0.9984 | 0.9124 | 0.9381 | 0.9555 |
| $\zeta(\tilde{\theta}_{brag,4})$ | 1.1887 | 1.1126 | 1.0851 | 1.1049 | 1.0684 | 1.0517 |

**Table 5.2:** *The ratio of the MSE of the sample mean $\bar{X}_n$ to the MSE of each of the four estimators, $\tilde{\theta}_{brag,1}$, $\tilde{\theta}_{brag,2}$, $\tilde{\theta}_{brag,3}$ and $\tilde{\theta}_{brag,4}$.*

| Distribution used | Contaminated Normal |
|---|---|
| Parameters used | $\mu_1 = \mu_2 = 0$, $\sigma_1 = 1$, $\sigma_2 = 6$, $p = 0.1$ |

| | $n = 10$ | | | $n = 30$ | | |
|---|---|---|---|---|---|---|
| | $m = \frac{n}{3}$ | $m = \frac{2n}{3}$ | $m = n$ | $m = \frac{n}{3}$ | $m = \frac{2n}{3}$ | $m = n$ |
| $\zeta(\tilde{\theta}_{brag,1})$ | 2.1884 | 1.4566 | 1.1093 | 1.4558 | 1.1736 | 1.0949 |
| $\zeta(\tilde{\theta}_{brag,2})$ | 1.5767 | 1.3441 | 1.1076 | 1.2405 | 1.1324 | 1.1094 |
| $\zeta(\tilde{\theta}_{brag,3})$ | 0.6449 | 0.7291 | 0.9002 | 0.7915 | 0.8866 | 0.9015 |
| $\zeta(\tilde{\theta}_{brag,4})$ | 1.5799 | 1.3298 | 1.0881 | 1.2398 | 1.1429 | 1.1052 |
| | $n = 50$ | | | $n = 100$ | | |
| | $m = \frac{n}{3}$ | $m = \frac{2n}{3}$ | $m = n$ | | | |
| $\zeta(\tilde{\theta}_{brag,1})$ | 1.3011 | 1.1199 | 1.0846 | 1.1420 | 1.0720 | 1.0531 |
| $\zeta(\tilde{\theta}_{brag,2})$ | 1.1660 | 1.1020 | 1.0761 | 1.0809 | 1.0538 | 1.0296 |
| $\zeta(\tilde{\theta}_{brag,3})$ | 0.8528 | 0.9095 | 0.9179 | 0.9259 | 0.9484 | 0.9572 |
| $\zeta(\tilde{\theta}_{brag,4})$ | 1.1682 | 1.0923 | 1.0802 | 1.0881 | 1.0526 | 1.0413 |

**Table 5.3:** *The ratio of the MSE of the sample mean $\bar{X}_n$ to the MSE of each of the four estimators, $\tilde{\theta}_{brag,1}$, $\tilde{\theta}_{brag,2}$, $\tilde{\theta}_{brag,3}$ and $\tilde{\theta}_{brag,4}$.*

| Distribution used | Double Exponential |
|---|---|
| Parameters used | $\mu = 0$ and $\sigma = \sqrt{2}$ |

| | $n = 10$ | | | $n = 30$ | | |
|---|---|---|---|---|---|---|
| | $m = \frac{n}{3}$ | $m = \frac{2n}{3}$ | $m = n$ | $m = \frac{n}{3}$ | $m = \frac{2n}{3}$ | $m = n$ |
| $\zeta(\tilde{\theta}_{brag,1})$ | 1.1836 | 1.0753 | 1.0396 | 1.0849 | 1.0380 | 1.0140 |
| $\zeta(\tilde{\theta}_{brag,2})$ | 1.1177 | 1.0671 | 1.0436 | 1.0498 | 1.0336 | 1.0188 |
| $\zeta(\tilde{\theta}_{brag,3})$ | 0.8839 | 0.9377 | 0.9531 | 0.9538 | 0.9632 | 0.9789 |
| $\zeta(\tilde{\theta}_{brag,4})$ | 1.1116 | 1.0647 | 1.0283 | 1.0368 | 1.0190 | 1.0261 |
| | $n = 50$ | | | $n = 100$ | | |
| | $m = \frac{n}{3}$ | $m = \frac{2n}{3}$ | $m = n$ | $m = \frac{n}{3}$ | $m = \frac{2n}{3}$ | $m = n$ |
| $\zeta(\tilde{\theta}_{brag,1})$ | 1.0504 | 1.0249 | 1.0206 | 1.0215 | 1.0147 | 1.0041 |
| $\zeta(\tilde{\theta}_{brag,2})$ | 1.0337 | 1.0218 | 1.0119 | 1.0150 | 1.0133 | 1.0064 |
| $\zeta(\tilde{\theta}_{brag,3})$ | 0.9648 | 0.9822 | 0.9781 | 0.9807 | 0.9865 | 0.9869 |
| $\zeta(\tilde{\theta}_{brag,4})$ | 1.0312 | 1.0169 | 1.0121 | 1.0084 | 1.0072 | 1.0078 |

**Table 5.4:** *The ratio of the MSE of the sample mean $\bar{X}_n$ to the MSE of each of the four estimators, $\tilde{\theta}_{brag,1}$, $\tilde{\theta}_{brag,2}$, $\tilde{\theta}_{brag,3}$ and $\tilde{\theta}_{brag,4}$.*

| **Distribution used** | **F** |
|---|---|
| **Parameters used** | $df_1 = 8$ and $df_2 = 5$ |

| | $n = 10$ | | | $n = 30$ | | |
|---|---|---|---|---|---|---|
| | $m = \frac{n}{3}$ | $m = \frac{2n}{3}$ | $m = n$ | $m = \frac{n}{3}$ | $m = \frac{2n}{3}$ | $m = n$ |
| $\zeta(\tilde{\theta}_{brag,1})$ | 2.5621 | 2.2811 | 1.1545 | 2.0670 | 1.6059 | 1.2541 |
| $\zeta(\tilde{\theta}_{brag,2})$ | 2.1631 | 2.1250 | 1.2562 | 1.8241 | 1.5357 | 1.2303 |
| $\zeta(\tilde{\theta}_{brag,3})$ | 0.6781 | 0.6597 | 1.0921 | 0.6995 | 0.7959 | 1.0674 |
| $\zeta(\tilde{\theta}_{brag,4})$ | 2.1227 | 2.1074 | 1.1154 | 1.8361 | 1.5175 | 1.3181 |
| | $n = 50$ | | | $n = 100$ | | |
| | $m = \frac{n}{3}$ | $m = \frac{2n}{3}$ | $m = n$ | $m = \frac{n}{3}$ | $m = \frac{2n}{3}$ | $m = n$ |
| $\zeta(\tilde{\theta}_{brag,1})$ | 1.9424 | 1.6461 | 1.2617 | 1.7562 | 1.5196 | 1.2694 |
| $\zeta(\tilde{\theta}_{brag,2})$ | 1.7325 | 1.5837 | 1.2569 | 1.5968 | 1.4853 | 1.2566 |
| $\zeta(\tilde{\theta}_{brag,3})$ | 0.7266 | 0.8219 | 1.0782 | 0.7890 | 0.8411 | 1.0677 |
| $\zeta(\tilde{\theta}_{brag,4})$ | 1.7145 | 1.5217 | 1.2106 | 1.5874 | 1.4831 | 1.2592 |

**Table 5.5:** *The ratio of the MSE of the sample mean $\bar{X}_n$ to the MSE of each of the four estimators, $\tilde{\theta}_{brag,1}$, $\tilde{\theta}_{brag,2}$, $\tilde{\theta}_{brag,3}$ and $\tilde{\theta}_{brag,4}$.*

### 5.4.1 Remarks on the output

Looking at the output from the tables separately we find the following:

⋆ **Table 5.1**: The output here is as we expected it to be; all the values are close to 1. This is because, in the case where we generate from a normal distribution, the sample mean is an admissible estimator for the population mean, i.e., any estimator we propose will perform at most as well as the sample mean.

⋆ **Table 5.2**: Introducing heavier tails into the standard normal distribution using the contaminated normal distribution, we find that all the estimators (except $\tilde{\theta}_{brag,3}$) suddenly perform much better for smaller resample sizes. In most cases the traditional $n$-out-of-$n$ bootstrap estimator produces values close to 1, while the $m$-out-of-$n$ estimators produce values greater than 1 (up to 90% greater in the case of $\tilde{\theta}_{brag,1}$ and sample size $n = 10$).

⋆ **Table 5.3**: Here we see the same result as in Table 5.2, but to a greater degree because of the increased contamination.

⋆ **Table 5.4**: We see here that the performance of the BRAGGing estimator is very good for small to moderate samples and small resample sizes (we get improvements of up to 10% in the case of $\tilde{\theta}_{brag,1}$, $\tilde{\theta}_{brag,2}$ and $\tilde{\theta}_{brag,4}$).

⋆ **Table 5.5**: The distribution used here (the $F(8,5)$ distribution) is the most skewed distribution considered in this study, but it is also the one which shows the greatest improvement of the BRAGGing estimator over the sample mean. For the statistics $\tilde{\theta}_{brag,1}$, $\tilde{\theta}_{brag,2}$ and $\tilde{\theta}_{brag,4}$ we routinely obtain values which improve on the sample mean's values by more than 20%

(in some cases it improves by more than 150%!). A probable, and logical reason why the estimator $\tilde{\theta}_{brag,3}$ performs so poorly in this case can be possibly ascribed to the fact that, while the original statistic $\sqrt{n}\left(\hat{\theta}_n - \theta\right)$ was assumed to have a symmetric distribution, we are not assured that the distribution of the $m$-out-of-$n$ bootstrap statistic, i.e., $\sqrt{m}\left(\hat{\theta}^*_m - \hat{\theta}_n\right)$, is also symmetric. Typically it will not be symmetric and so the calculations are performed under false assumptions.

$\star$ We see in all the tables that, as $n$ becomes larger, the statistic $\zeta$ converges to 1.

Based on this preliminary run it would seem that the estimators $\tilde{\theta}_{brag,1}$, $\tilde{\theta}_{brag,2}$ and (possibly) $\tilde{\theta}_{brag,4}$ will have the desirable property that their MSEs are smaller than the MSE of the sample mean. The poor behaviour of $\tilde{\theta}_{brag,3}$ can possibly be ascribed to the fact that, while we assumed in the construction of $\tilde{\theta}_{brag,3}$ that $\sqrt{n}(\hat{\theta}_n - \theta)$ has a symmetric distribution, we are not assured that the distribution of the $m$-out-of-$n$ bootstrap statistic, $\sqrt{m}(\hat{\theta}^*_m - \hat{\theta}_n)$, is also symmetric (typically it will not be symmetric). Note that, due to the complexity of the calculation of $\tilde{\theta}_{brag,4}$ and its almost negligible benefits, we will continue our investigation of these estimators by focusing only on $\tilde{\theta}_{brag,1}$ and $\tilde{\theta}_{brag,2}$.

## 5.5 The choice of $m$

In this section we will consider the various ways of selecting an optimal value of $m$ when estimating the population mean using the BRAGGing technique. The discussion will follow from Cornish-Fisher expansions discussed in detail in Hall (1992) and Chung and Lee (2001). The basic derivation of these expansions are summarized in Appendix A.

We will begin the discussion by first finding an expansion for a general BRAGGing estimator that falls in a class of statistics determined by the smooth function model (as discussed in Hall (1992)), and we will then go on to derive an expression for the BRAGGing mean estimator. This expansion will facilitate the development of a general rule for selecting $m$ which will finally lead to a data-dependent rule for selecting $m$ in the case of the BRAGGing mean estimator.

### 5.5.1 Cornish-Fisher expansion

A Cornish-Fisher expansion will be employed to determine an asymptotic expression for the general BRAGGing statistic that satisfies the smooth function model discussed in Appendix A.3. If all of terms in the expansion are known then the optimal value of $m$ can easily be derived.

#### General statistic

Considering the smooth function model, let $\theta$ be defined as some parameter which is a function of a $d$-dimensional mean $\boldsymbol{\mu} = \mathrm{E}(\boldsymbol{X})$ and where $\boldsymbol{X}$ is a $d$-dimensional column vector whose $i^{\text{th}}$ component is denoted by $X^{(i)}$. In other words, we have that $\theta = g(\boldsymbol{\mu})$, where $g$ is defined such that $g : \mathbb{R}^d \to \mathbb{R}$. An estimator for $\theta$ is then the simple plug-in estimator given by $\hat{\theta}_n = g(\bar{\boldsymbol{X}}_n)$, where $\bar{\boldsymbol{X}}_n$ is defined as

$$\bar{\boldsymbol{X}}_n = \left(\frac{1}{n}\sum_{j=1}^{n} X_j^{(1)}, \frac{1}{n}\sum_{j=1}^{n} X_j^{(2)}, \ldots, \frac{1}{n}\sum_{j=1}^{n} X_j^{(d)}\right)^T. \tag{5.2}$$

Define the standardized version of $\hat{\theta}_n$ as the statistic $T_n$ in the following way:

$$T_n := \frac{\sqrt{n}\left(g(\bar{\boldsymbol{X}}_n) - g(\boldsymbol{\mu})\right)}{h(\boldsymbol{\mu})},$$

where $\{h(\boldsymbol{\mu})\}^2$ is the asymptotic variance of $\sqrt{n}g(\bar{\boldsymbol{X}}_n)$. We assume therefore that $T_n$ satisfies the assumptions of the smooth function model as discussed in Appendix A.3. The $m$-out-of-$n$ bootstrap version of this statistic is then

$$T_m^* = \frac{\sqrt{m}\left(g(\bar{\boldsymbol{X}}_m^*) - g(\bar{\boldsymbol{X}}_n)\right)}{h(\bar{\boldsymbol{X}}_n)}.$$

Let the median of the bootstrap distribution of the statistic $g(\bar{\boldsymbol{X}}_m^*)$ be denoted by $\mathrm{Med}^*(g(\bar{\boldsymbol{X}}_m^*)) = \tilde{\theta}_{brag,1}$. Using the Cornish-Fisher expansion we can obtain an expansion for $\tilde{\theta}_{brag,1}$. We proceed by first noting that

$$\begin{aligned} \mathrm{P}^*(g(\bar{\boldsymbol{X}}_m^*) \leq \tilde{\theta}_{brag,1}) &\approx \frac{1}{2}, \text{ i.e.,} \\ \mathrm{P}^*\left(T_m^* \leq \frac{\sqrt{m}\left(\tilde{\theta}_{brag,1} - g(\bar{\boldsymbol{X}}_n)\right)}{h(\bar{\boldsymbol{X}}_n)}\right) &\approx \frac{1}{2}. \end{aligned}$$

Now, the Cornish-Fisher expansion of the $\alpha^{\text{th}}$ quantile of the bootstrap distribution of $T_m^*$, denoted by $v(\alpha)$, is given by

$$v(\alpha) = z(\alpha) + m^{-1/2}\hat{p}_1^{cf}\left(z(\alpha)\right) + m^{-1}\hat{p}_2^{cf}\left(z(\alpha)\right) + O_p(m^{-3/2}).$$

However, since we are interested in calculating the median we choose $\alpha = 0.5$ and obtain the following expression:

$$\begin{aligned} \sqrt{m}\left(\tilde{\theta}_{brag,1} - g(\bar{\boldsymbol{X}}_n)\right)/h(\bar{\boldsymbol{X}}_n) &= m^{-1/2}\hat{p}_1^{cf}(0) + m^{-1}\hat{p}_2^{cf}(0) + O_p(m^{-3/2}) \\ &= -m^{-1/2}\hat{p}_1(0) + m^{-1}\left\{\hat{p}_1(0)\,\frac{d}{dx}\hat{p}_1(x)\bigg|_{x=0} - \hat{p}_2(0)\right\} + O_p(m^{-3/2}) \\ &= -m^{-1/2}\hat{p}_1(0) + m^{-1}\hat{p}_1(0)\,\frac{d}{dx}\hat{p}_1(x)\bigg|_{x=0} - m^{-1}\hat{p}_2(0) + O_p(m^{-3/2}). \end{aligned}$$

Note the following (Hall 1992):

⋆ $\hat{p}_1^{cf}(x) = -\hat{p}_1(x)$, so that

$$\hat{p}_1^{cf}(0) = -\hat{p}_1(0),$$

⋆ $\hat{p}_2^{cf}(x) = \hat{p}_1(x)\left[\frac{d}{dx}\hat{p}_1(x)\right] - \frac{1}{2}x\{\hat{p}_1(x)\}^2 - \hat{p}_2(x)$, so that

$$\hat{p}_2^{cf}(0) = \hat{p}_1(0)\left[\frac{d}{dx}\hat{p}_1(x)\bigg|_{x=0}\right] - \hat{p}_2(0),$$

⋆ $\hat{p}_1(x) = -(\hat{k}_{1,2} + \frac{1}{6}\hat{k}_{3,1}(x^2-1))$ so that

$$\hat{p}_1(0) = -\hat{k}_{1,2} + \frac{1}{6}\hat{k}_{3,1},$$

⋆ $\frac{d}{dx}\hat{p}_1(x) = -\frac{1}{3}\hat{k}_{3,1}x$ so that

$$\frac{d}{dx}\hat{p}_1(x)\bigg|_{x=0} = 0,$$

⋆ $\hat{p}_2(x) = -x\left[\frac{1}{2}(\hat{k}_{2,2}+\hat{k}_{1,2}^2) + \frac{1}{24}(\hat{k}_{4,1} + 4\hat{k}_{1,2}\hat{k}_{3,1})(x^2-3) + \frac{1}{72}\hat{k}_{3,1}^2(x^4-10x+15)\right]$ so that

$$\hat{p}_2(0) = 0.$$

It should also be noted that the $\hat{k}_{i,j}$ terms are defined in Appendix A.3 in equation (1.15).

The Cornish-Fisher expansion then becomes

$$\sqrt{m}\left(\tilde{\theta}_{brag,1} - g(\bar{\boldsymbol{X}}_n)\right)/h(\bar{\boldsymbol{X}}_n) = m^{-1/2}\left[\hat{k}_{1,2} - \frac{1}{6}\hat{k}_{3,1}\right] + O_p(m^{-3/2}).$$

Solving for $\tilde{\theta}_{brag,1}$ we get

$$\tilde{\theta}_{brag,1} = g(\bar{\boldsymbol{X}}_n) + m^{-1}h(\bar{\boldsymbol{X}}_n)\left[\hat{k}_{1,2} - \frac{1}{6}\hat{k}_{3,1}\right] + O_p(m^{-2}). \tag{5.3}$$

### 5.5.2 A general rule for selecting an optimal $m$ using the smooth function model

We begin the development of a general rule for selecting an optimal $m$ by first defining an approximation to the quantity $\tilde{\theta}_{brag,1}$. This approximation is simply based on the leading terms of the Cornish-Fisher expansion. Therefore, we will approximate $\tilde{\theta}_{brag,1}$ with $\tilde{\theta}^A_{brag,1}$ which is defined as (i.e., dropping the $O_p(m^{-2})$ term in (5.3)),

$$\tilde{\theta}^A_{brag,1} := g(\bar{\boldsymbol{X}}_n) + m^{-1}h(\bar{\boldsymbol{X}}_n)\left[\hat{k}_{1,2} - \frac{1}{6}\hat{k}_{3,1}\right]. \tag{5.4}$$

The rule for selecting $m$ based on $\tilde{\theta}^A_{brag,1}$ will involve finding the $m$ value that minimizes the MSE, i.e., we will minimize $\mathrm{MSE}(\tilde{\theta}^A_{brag,1})$ over the possible $m$ values.

#### The optimal choice of $m$ when estimating $\theta$ using $\tilde{\theta}^A_{brag,1}$

We will now obtain the MSE of $\tilde{\theta}^A_{brag,1}$ using the definition given in (5.4).

$$\begin{aligned}
\mathrm{MSE}(\tilde{\theta}^A_{brag,1}) &= \mathrm{E}\left\{\left(\tilde{\theta}^A_{brag,1} - g(\boldsymbol{\mu})\right)^2\right\} \\
&= \mathrm{E}\left\{\left(g(\bar{\boldsymbol{X}}_n) + m^{-1}h(\bar{\boldsymbol{X}}_n)\left[\hat{k}_{1,2} - \frac{1}{6}\hat{k}_{3,1}\right] - g(\boldsymbol{\mu})\right)^2\right\} \\
&= \mathrm{E}\left\{\left(g(\bar{\boldsymbol{X}}_n) - g(\boldsymbol{\mu})\right)^2\right\} + 2m^{-1}\,\mathrm{E}\left\{\left(g(\bar{\boldsymbol{X}}_n) - g(\boldsymbol{\mu})\right)h(\bar{\boldsymbol{X}}_n)\left[\hat{k}_{1,2} - \frac{1}{6}\hat{k}_{3,1}\right]\right\} \\
&\quad + m^{-2}\,\mathrm{E}\left\{h^2(\bar{\boldsymbol{X}}_n)\left[\hat{k}_{1,2} - \frac{1}{6}\hat{k}_{3,1}\right]^2\right\} =: q(m).
\end{aligned} \tag{5.5}$$

The first derivative of $q(m)$ is

$$\frac{dq(m)}{dm} = -2m^{-2}\,\mathrm{E}\left\{\left(g(\bar{\boldsymbol{X}}_n) - g(\boldsymbol{\mu})\right)h(\bar{\boldsymbol{X}}_n)\left[\hat{k}_{1,2} - \frac{1}{6}\hat{k}_{3,1}\right]\right\} - 2m^{-3}\,\mathrm{E}\left\{h^2(\bar{\boldsymbol{X}}_n)\left[\hat{k}_{1,2} - \frac{1}{6}\hat{k}_{3,1}\right]^2\right\}.$$

Setting $\frac{dq(m)}{dm}$ to zero and solving for $m$ we get:

$$m_1 = \frac{\mathrm{E}\left\{h^2(\bar{\boldsymbol{X}}_n)\left[\frac{1}{6}\hat{k}_{3,1} - \hat{k}_{1,2}\right]^2\right\}}{\mathrm{E}\left\{\left(g(\bar{\boldsymbol{X}}_n) - g(\boldsymbol{\mu})\right)h(\bar{\boldsymbol{X}}_n)\left[\frac{1}{6}\hat{k}_{3,1} - \hat{k}_{1,2}\right]\right\}}. \tag{5.6}$$

We will estimate $m_1$ by making use of this expression for $m_1$ and by

⋆ linear approximations or

⋆ bootstrap methods.

These techniques will be discussed in the next section.

**Remark:**
A problem for future research is to determine the coefficient of the $m^{-2}$ term in (5.3) and proceed as above to obtain a new estimator, i.e., $\tilde{\theta}^A_{brag,1}$ + [coefficient]$\cdot m^{-2}$, ultimately leading to a more accurate version of the corresponding optimal bootstrap resample size.

### The optimal choice of $m$ when estimating $\theta$ using $\tilde{\theta}^A_{brag,2}$

If we define $\tilde{\theta}^A_{brag,2}$ as

$$\tilde{\theta}^A_{brag,2} = \sqrt{\frac{m}{n}}\tilde{\theta}^A_{brag,1} + \left(1 - \sqrt{\frac{m}{n}}\right) g(\bar{\boldsymbol{X}}_n),$$

then we can obtain a similar general expression for the optimal choice of $m$ when using the estimator $\tilde{\theta}^A_{brag,2}$ to estimate $\theta$. The derivation is analogous to the previous one since it also relies on the expression obtained in (5.4).

We begin by defining the MSE of $\tilde{\theta}^A_{brag,2}$ in the general case:

$$\begin{aligned}
\text{MSE}(\tilde{\theta}^A_{brag,2}) &= \text{E}\left\{\left(\tilde{\theta}^A_{brag,2} - g(\boldsymbol{\mu})\right)^2\right\} \\
&= \text{E}\left\{\left(\sqrt{\frac{m}{n}}\tilde{\theta}^A_{brag,1} + \left(1 - \sqrt{\frac{m}{n}}\right) g(\bar{\boldsymbol{X}}_n) - g(\boldsymbol{\mu})\right)^2\right\} \\
&= \text{E}\left\{\left(\sqrt{\frac{m}{n}}\left\{g(\bar{\boldsymbol{X}}_n) + m^{-1}h(\bar{\boldsymbol{X}}_n)\left[\hat{k}_{1,2} - \frac{1}{6}\hat{k}_{3,1}\right]\right\} + \left(1 - \sqrt{\frac{m}{n}}\right) g(\bar{\boldsymbol{X}}_n) - g(\boldsymbol{\mu})\right)^2\right\} \\
&= \text{E}\left\{\left(g(\bar{\boldsymbol{X}}_n) - g(\boldsymbol{\mu})\right)^2\right\} + 2m^{-\frac{1}{2}}n^{-\frac{1}{2}}\,\text{E}\left\{\left(g(\bar{\boldsymbol{X}}_n) - g(\boldsymbol{\mu})\right) h(\bar{\boldsymbol{X}}_n)\left[\hat{k}_{1,2} - \frac{1}{6}\hat{k}_{3,1}\right]\right\} \\
&\quad + m^{-1}n^{-1}\,\text{E}\left\{h^2(\bar{\boldsymbol{X}}_n)\left[\hat{k}_{1,2} - \frac{1}{6}\hat{k}_{3,1}\right]^2\right\} =: r(m).
\end{aligned}$$

The first derivative of $r(m)$ is

$$\frac{dr(m)}{dm} = -m^{-\frac{3}{2}}n^{-\frac{1}{2}}\,\text{E}\left\{\left(g(\bar{\boldsymbol{X}}_n) - g(\boldsymbol{\mu})\right) h(\bar{\boldsymbol{X}}_n)\left[\hat{k}_{1,2} - \frac{1}{6}\hat{k}_{3,1}\right]\right\} - m^{-2}n^{-1}\,\text{E}\left\{h^2(\bar{\boldsymbol{X}}_n)\left[\hat{k}_{1,2} - \frac{1}{6}\hat{k}_{3,1}\right]^2\right\}.$$

Setting $\frac{dr(m)}{dm}$ to zero and solving for $m$ we get:

$$\begin{aligned}
m_2 &= \frac{1}{n}\left[\frac{\text{E}\left\{h^2(\bar{\boldsymbol{X}}_n)\left[\frac{1}{6}\hat{k}_{3,1} - \hat{k}_{1,2}\right]^2\right\}}{\text{E}\left\{\left(g(\bar{\boldsymbol{X}}_n) - g(\boldsymbol{\mu})\right) h(\bar{\boldsymbol{X}}_n)\left[\frac{1}{6}\hat{k}_{3,1} - \hat{k}_{1,2}\right]\right\}}\right]^2 \\
&= \frac{1}{n}\left[m_1\right]^2. \qquad (5.7)
\end{aligned}$$

**Remark:**
Further discussion will focus only on the derivation of an estimator for $m_1$, since, once this value is determined, an estimator for $m_2$ can be obtained trivially through the relationship given in (5.7).

**The optimal choice of $m$ when estimating the population mean $\mu$ using $\tilde{\theta}^A_{brag,1}$**

Let $X_1, X_2, \ldots, X_n$ be i.i.d. observations and let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Set $\mathbf{x} = (x^{(1)}, x^{(2)})^T$, where $x^{(i)}$ is the $i^{\text{th}}$ component of $\mathbf{x}$. Then define $g(\mathbf{x}) = x^{(1)}$ and $h^2(\mathbf{x}) = x^{(2)} - \left(x^{(1)}\right)^2$. Choosing $x^{(1)} = x$ and $x^{(2)} = x^2$ we find $g(\bar{\boldsymbol{X}}_n) = \bar{X}_n$, and $h^2(\bar{\boldsymbol{X}}_n) = \frac{1}{n}\sum_{i=1}^n X_i^2 - \left(\frac{1}{n}\sum_{i=1}^n X_i\right)^2 = \hat{\mu}_2$, where $\bar{\boldsymbol{X}}_n$ is defined as in equation (5.2). Finally, the quantities stated in equation (5.4) are then given by

$$g(\bar{\boldsymbol{X}}_n) = \bar{X}_n, \qquad g(\boldsymbol{\mu}) = \mu, \qquad h^2(\bar{\boldsymbol{X}}_n) = \hat{\mu}_2, \qquad \hat{k}_{3,1} = \hat{\kappa}_3, \qquad \hat{k}_{1,2} = 0.$$

Therefore,

$$\frac{1}{6}\hat{k}_{3,1} - \hat{k}_{1,2} = \frac{1}{6}\hat{\kappa}_3 = \frac{1}{6}\frac{\hat{\mu}_3}{\hat{\mu}_2^{3/2}},$$

where $\hat{\kappa}_3 = \hat{\mu}_3/\hat{\mu}_2^{3/2}$ and $\hat{\mu}_\nu = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^\nu$, so the approximate Cornish-Fisher expansion of the median of the bootstrap sample mean is now given by:

$$\tilde{\theta}^A_{brag,1} = \bar{X}_n - \frac{1}{6}m^{-1}\hat{\mu}_2^{1/2}\hat{\kappa}_3.$$

The approximate optimal choice of $m$ in this case is then obtained by substituting the appropriate values into (5.6). This gives us the expression

$$m_1 = \frac{1}{6}\cdot\frac{\mathrm{E}\left[\hat{\kappa}_3^2\hat{\mu}_2\right]}{\mathrm{E}\left[\left(\bar{X} - \mu\right)\hat{\kappa}_3\hat{\mu}_2^{1/2}\right]}.$$

Now, since we define $Y_i = X_i - \mathrm{E}(X_i)$ we have the following form of $m$:

$$m_1 = \frac{1}{6}\cdot\frac{\mathrm{E}\left[\hat{\mu}_3^2/\hat{\mu}_2^2\right]}{\mathrm{E}\left[\bar{Y}_n\hat{\mu}_3/\hat{\mu}_2\right]}. \tag{5.8}$$

**Remark:**
Asymptotic expansions for the numerator and denominator of (5.8) will be derived in the next section.

## 5.6 Estimators for $m_1$ when estimating $\mu$

Using equation (5.8) as a theoretical starting point we will now attempt to estimate this optimal choice of $m$ using various strategies. The strategies which will be followed are:

1. A bootstrap approximation of (5.8).

2. A Taylor series expansion of the numerator and denominator of (5.8) (individually) and then using a simple sample moment substitution scheme.

3. A Taylor series expansion of the numerator and denominator of (5.8) (individually) followed by the estimation of the products of population moments using a bias correction approach (to order $1/n$ and $1/n^2$).

4. Estimation of the numerator of (5.8) with an unbiased sample moment estimator and then applying a Taylor series expansion to the denominator followed by a bias correction approach for the estimation of population moment product terms (to the order $1/n$ and $1/n^2$).

Techniques 2 to 4 rely on a Taylor series expansion of the numerator and denominator in (5.8) which we will need to show before any further discussion can take place.

### 5.6.1 Taylor series expansions of the numerator and denominator of $m_1$

We begin by noting that the multivariate version of the Taylor series expansion of a function $f$ evaluated in the points $x_1, \ldots, x_d$ around the points $a_1, \ldots, a_d$ is given by

$$f(x_1, \ldots, x_d) = \sum_{n_1=0}^{\infty} \cdots \sum_{n_d=0}^{\infty} \frac{(x_1 - a_1)^{n_1} \cdots (x_d - a_d)^{n_d}}{n_1! \ldots n_d!} \left( \frac{\partial^{n_1 + \cdots + n_d}}{\partial x_1^{n_1} \ldots \partial x_d^{n_d}} f(a_1 \ldots a_d) \right).$$

**Expression for the numerator:**

Using the above multivariable Taylor series expansion, we find that the numerator of (5.8) can be written as the expected value of the function $f(x, y) = (x/y)^2$ evaluated in the points $x = \hat{\mu}_3$ and $y = \hat{\mu}_2$, expanded about the values $\mu_3$ and $\mu_2$. The expansion is provided below:

$$\begin{aligned}
\mathrm{E}\left\{ \left( \frac{\hat{\mu}_3}{\hat{\mu}_2} \right)^2 \right\} &\approx \mathrm{E}\left\{ \left( \frac{\mu_3}{\mu_2} \right)^2 + 2\frac{\mu_3}{\mu_2^2}(\hat{\mu}_3 - \mu_3) - 2\frac{\mu_3^2}{\mu_2^3}(\hat{\mu}_2 - \mu_2) \right. \\
&\quad \left. + \frac{1}{\mu_2^2}(\hat{\mu}_3 - \mu_3)^2 + 3\frac{\mu_3^2}{\mu_2^4}(\hat{\mu}_2 - \mu_2)^2 - 4\frac{\mu_3}{\mu_2^3}(\hat{\mu}_3 - \mu_3)(\hat{\mu}_2 - \mu_2) \right\} \\
&= \left( \frac{\mu_3}{\mu_2} \right)^2 + 2\frac{\mu_3}{\mu_2^2}\mathrm{E}(\hat{\mu}_3 - \mu_3) - 2\frac{\mu_3^2}{\mu_2^3}\mathrm{E}(\hat{\mu}_2 - \mu_2) \\
&\quad + \frac{1}{\mu_2^2}\mathrm{E}\left((\hat{\mu}_3 - \mu_3)^2\right) + 3\frac{\mu_3^2}{\mu_2^4}\mathrm{E}\left((\hat{\mu}_2 - \mu_2)^2\right) - 4\frac{\mu_3}{\mu_2^3}\mathrm{E}\left((\hat{\mu}_3 - \mu_3)(\hat{\mu}_2 - \mu_2)\right) \\
&= \left( \frac{\mu_3}{\mu_2} \right)^2 + 2\frac{\mu_3}{\mu_2^2}\mathrm{E}(\hat{\mu}_3) - 2\frac{\mu_3^2}{\mu_2^3}\mathrm{E}(\hat{\mu}_2) \\
&\quad + \frac{1}{\mu_2^2}\mathrm{E}\left(\hat{\mu}_3^2\right) - 2\frac{\mu_3}{\mu_2^2}\mathrm{E}\left(\hat{\mu}_3\right) + 3\frac{\mu_3^2}{\mu_2^4}\mathrm{E}\left(\hat{\mu}_2^2\right) - 6\frac{\mu_3^2}{\mu_2^3}\mathrm{E}\left(\hat{\mu}_2\right) \\
&\quad - 4\frac{\mu_3}{\mu_2^3}\mathrm{E}\left(\hat{\mu}_3\hat{\mu}_2\right) + 4\frac{\mu_3}{\mu_2^2}\mathrm{E}\left(\hat{\mu}_3\right) + 4\frac{\mu_3^2}{\mu_2^3}\mathrm{E}\left(\hat{\mu}_2\right).
\end{aligned} \tag{5.9}$$

The expected values in this expression can be simplified by making use of the expressions found in Appendix D. The expressions become:

$$\mathrm{E}(\hat{\mu}_2) = \mu_2 + \frac{1}{n}\left\{-\mu_2\right\},$$

$$\mathrm{E}(\hat{\mu}_3) = \mu_3 + \frac{1}{n}\left\{-3\mu_3\right\} + \frac{1}{n^2}\left\{2\mu_3\right\},$$

$$\mathrm{E}(\hat{\mu}_2^2) = \mu_2^2 + \frac{1}{n}\left\{\mu_4 - 3\mu_2^2\right\} + \frac{1}{n^2}\left\{-2\mu_4 + 5\mu_2^2\right\} + \frac{1}{n^3}\left\{\mu_4 - 3\mu_2^2\right\},$$

$$\mathrm{E}(\hat{\mu}_3^2) = \mu_3^2 + \frac{1}{n}\left\{\mu_6 + 9\mu_2^3 - 7\mu_3^2 - 6\mu_2\mu_4\right\} + O\left(\frac{1}{n^2}\right),$$

$$\mathrm{E}(\hat{\mu}_2\hat{\mu}_3) = \mu_2\mu_3 + \frac{1}{n}\left\{\mu_5 - 8\mu_2\mu_3\right\} + O\left(\frac{1}{n^2}\right).$$

Once all of the above expressions have been substituted into equation (5.9) then we obtain the final expression for the expansion of the numerator of (5.8).

$$\mathrm{E}\left\{ \left( \frac{\hat{\mu}_3}{\hat{\mu}_2} \right)^2 \right\} = \left( \frac{\mu_3}{\mu_2} \right)^2 + \frac{1}{n}\left\{ \frac{1}{\mu_2^4}\left[\mu_2^2\mu_6 + 9\mu_2^5 + 8\mu_2^2\mu_3^2 - 6\mu_2^3\mu_4 + 3\mu_3^2\mu_4 - 4\mu_2\mu_3\mu_5\right] \right\} + O(n^{-2}).$$

**Expression for the denominator:**

The expression for the denominator is obtained in a similar fashion to the numerator except that the Taylor expansion is carried out on a function of three variables as opposed to the two variable Taylor expansion used for the numerator.

The denominator of (5.8) (without the multiplier 6) can now be written as a multivariable Taylor series expansion of the expected value of the function $f(x, y, z) = xz/y$ evaluated in the points $x = \bar{Y}_n$, $y = \hat{\mu}_2$ and $z = \hat{\mu}_3$, expanded about the values $\mu_1 = 0$, $\mu_2$ and $\mu_3$. The expansion is given below:

$$\begin{aligned}
\mathrm{E}\left(\bar{Y}_n\frac{\hat{\mu}_3}{\hat{\mu}_2}\right) \approx\ & \mathrm{E}\left\{\frac{\mu_1\mu_3}{\mu_2} + \frac{\mu_1}{\mu_2}(\hat{\mu}_3-\mu_3) - \frac{\mu_1\mu_3}{\mu_2^2}(\hat{\mu}_2-\mu_2) + \frac{\mu_3}{\mu_2}(\bar{Y}_n-\mu_1)\right. \\
& +\frac{\mu_1\mu_3}{\mu_2^3}(\hat{\mu}_2-\mu_2)^2 - \frac{\mu_3}{\mu_2^2}(\bar{Y}_n-\mu_1)(\hat{\mu}_2-\mu_2) + \frac{1}{\mu_2}(\bar{Y}_n-\mu_1)(\hat{\mu}_3-\mu_3) \\
& -\frac{\mu_1}{\mu_2^2}(\hat{\mu}_2-\mu_2)(\hat{\mu}_3-\mu_3) - \frac{\mu_1\mu_3}{\mu_2^4}(\hat{\mu}_2-\mu_2)^3 + \frac{\mu_3}{\mu_2^3}(\bar{Y}_n-\mu_1)(\hat{\mu}_2-\mu_2)^2 \\
& +\frac{\mu_1}{\mu_2^3}(\hat{\mu}_2-\mu_2)^2(\hat{\mu}_3-\mu_3) - \frac{1}{\mu_2^2}(\bar{Y}_n-\mu_1)(\hat{\mu}_2-\mu_2)(\hat{\mu}_3-\mu_3) + \frac{\mu_1\mu_3}{\mu_2^5}(\hat{\mu}_2-\mu_2)^4 \\
& \left.-\frac{\mu_3}{\mu_2^4}(\bar{Y}_n-\mu_1)(\hat{\mu}_2-\mu_2)^3 - \frac{\mu_1}{\mu_2^4}(\hat{\mu}_2-\mu_2)^3(\hat{\mu}_3-\mu_3) + \frac{1}{\mu_2^3}(\bar{Y}_n-\mu_1)(\hat{\mu}_2-\mu_2)^2(\hat{\mu}_3-\mu_3)\right\} \\
=\ & -\frac{\mu_3}{\mu_2^2}\mathrm{E}\left\{\bar{Y}_n(\hat{\mu}_2-\mu_2)\right\} + \frac{1}{\mu_2}\mathrm{E}\left\{\bar{Y}_n(\hat{\mu}_3-\mu_3)\right\} + \frac{\mu_3}{\mu_2^3}\mathrm{E}\left\{\bar{Y}_n(\hat{\mu}_2-\mu_2)^2\right\} \\
& -\frac{1}{\mu_2^2}\mathrm{E}\left\{\bar{Y}_n(\hat{\mu}_2-\mu_2)(\hat{\mu}_3-\mu_3)\right\} - \frac{\mu_3}{\mu_2^4}\mathrm{E}\left\{\bar{Y}_n(\hat{\mu}_2-\mu_2)^3\right\} \\
& +\frac{1}{\mu_2^3}\mathrm{E}\left\{\bar{Y}_n(\hat{\mu}_2-\mu_2)^2(\hat{\mu}_3-\mu_3)\right\}. \qquad (5.10)
\end{aligned}$$

The expected values in the above equation simplify as follows (brief derivations of these terms can be found in Appendix D):

$$\mathrm{E}\left(\bar{Y}_n(\hat{\mu}_2-\mu_2)\right) = \frac{1}{n}\{\mu_3\} + \frac{1}{n^2}\{-\mu_3\},$$

$$\mathrm{E}\left(\bar{Y}_n(\hat{\mu}_3-\mu_3)\right) = \frac{1}{n}\{\mu_4 - 3\mu_2^2\} + \frac{1}{n^2}\{-3\mu_4 + 9\mu_2^2\} + O(n^{-3}),$$

$$\mathrm{E}\left(\bar{Y}_n(\hat{\mu}_2-\mu_2)^2\right) = \frac{1}{n^2}\{2\mu_3\mu_5 + \mu_4^2 - 4\mu_2^2\mu_4 - 8\mu_3^2\mu_2 + 3\mu_2^4\} + O(n^{-3}),$$

$$\mathrm{E}\left(\bar{Y}_n(\hat{\mu}_3-\mu_3)(\hat{\mu}_2-\mu_2)\right) = \frac{1}{n^2}\{\mu_6 - 10\mu_2\mu_4 - 7\mu_3^2 + 15\mu_2^3\} + O(n^{-3}),$$

$$\mathrm{E}\left(\bar{Y}_n(\hat{\mu}_2-\mu_2)^3\right) = \frac{1}{n^2}\{3\mu_3\mu_4 - 3\mu_2^2\mu_3\} + O(n^{-3}),$$

$$\mathrm{E}\left(\bar{Y}_n(\hat{\mu}_2-\mu_2)^2(\hat{\mu}_3-\mu_3)\right) = \frac{1}{n^2}\{2\mu_3\mu_5 + \mu_4^2 - 4\mu_2^2\mu_4 - 8\mu_3^2\mu_2 + 3\mu_2^4\} + O(n^{-3}).$$

Once all of these expressions have been substituted into (5.10) and the terms have been collected, we find that the denominator becomes

$$\mathrm{E}\left(\bar{Y}_n\frac{\hat{\mu}_3}{\hat{\mu}_2}\right) = \frac{1}{n}\left\{\frac{1}{\mu_2^2}\left[\mu_2\mu_4 - 3\mu_2^3 - \mu_3^2\right]\right\}$$

$$+\frac{1}{n^2}\left\{\frac{1}{\mu_2^4}\left[3\mu_2^3\mu_4-\mu_2^2\mu_6-3\mu_2^5+3\mu_2\mu_3\mu_5\right.\right.$$
$$\left.\left.+\mu_2\mu_4^2-3\mu_3^2\mu_4-5\mu_2^2\mu_3^2\right]\right\}+O(n^{-3})$$

### Combining the numerator and denominator

Combining the numerator and the denominator terms in the expression (5.8) we get

$$m_1^A=\frac{n}{6}\cdot\frac{A(\boldsymbol{\mu})+\frac{1}{n}B(\boldsymbol{\mu})}{C(\boldsymbol{\mu})+\frac{1}{n}D(\boldsymbol{\mu})}, \tag{5.11}$$

where $\boldsymbol{\mu}$ is the vector containing the elements $\{\mu_2,\mu_3,\mu_4,\mu_5,\mu_6\}$, and the functions $A(\boldsymbol{\mu})$, $B(\boldsymbol{\mu})$, $C(\boldsymbol{\mu})$ and $D(\boldsymbol{\mu})$ are defined as follows:

$$A(\boldsymbol{\mu})=\frac{\mu_3^2}{\mu_2^2},$$
$$B(\boldsymbol{\mu})=\frac{1}{\mu_2^4}\left[\mu_2^2\mu_6+9\mu_2^5+8\mu_2^2\mu_3^2-6\mu_2^3\mu_4+3\mu_3^2\mu_4-4\mu_2\mu_3\mu_5\right],$$
$$C(\boldsymbol{\mu})=\frac{1}{\mu_2^2}\left[\mu_2\mu_4-3\mu_2^3-\mu_3^2\right],$$
$$D(\boldsymbol{\mu})=\frac{1}{\mu_2^4}\left[3\mu_2^3\mu_4-\mu_2^2\mu_6-3\mu_2^5+3\mu_2\mu_3\mu_5+\mu_2\mu_4^2-3\mu_3^2\mu_4-5\mu_2^2\mu_3^2\right].$$

A variation of the quantity in (5.11) is one where the terms have been simplified such that the terms $1/\mu_2^2$ and $1/\mu_2^4$ no longer appear, i.e.,

$$m_1^A=\frac{n}{6}\cdot\frac{\tilde{A}(\tilde{\boldsymbol{\mu}})+\frac{1}{n}\tilde{B}(\tilde{\boldsymbol{\mu}})}{\tilde{C}(\tilde{\boldsymbol{\mu}})+\frac{1}{n}\tilde{D}(\tilde{\boldsymbol{\mu}})}, \tag{5.12}$$

where

$$\tilde{A}(\tilde{\boldsymbol{\mu}})=\mu_2^2\mu_3^2,$$
$$\tilde{B}(\tilde{\boldsymbol{\mu}})=\mu_2^2\mu_6+9\mu_2^5+8\mu_2^2\mu_3^2-6\mu_2^3\mu_4+3\mu_3^2\mu_4-4\mu_2\mu_3\mu_5,$$
$$\tilde{C}(\tilde{\boldsymbol{\mu}})=\mu_2^3\mu_4-3\mu_2^5-\mu_2^2\mu_3^2,$$
$$\tilde{D}(\tilde{\boldsymbol{\mu}})=3\mu_2^3\mu_4-\mu_2^2\mu_6-3\mu_2^5+3\mu_2\mu_3\mu_5+\mu_2\mu_4^2-3\mu_3^2\mu_4-5\mu_2^2\mu_3^2,$$

where $\tilde{\boldsymbol{\mu}}$ is the vector consisting of the products of population moments $\{\mu_2^2\mu_3^2,\ \mu_2^2\mu_6,\ \mu_2^3\mu_4,\ \mu_2^5,\ \mu_3^2\mu_4,\ \mu_2\mu_3\mu_5,\ \mu_2\mu_4^2\}$. Note that (5.11) and (5.12) are identical; these two different forms were created to enable us to derive different estimators for $m_1^A$.

### Remarks:

- ⋆ In the above approximations of $m_1$ we do not convert the result to an integer. This conversion will take place when we develop data-dependent choices of $m_1^A$.
- ⋆ Notice that $m_1^A$ can take on a wide range of values, depending on the complexity of the underlying distribution (i.e., it is dependent on the behaviour of the central moments).

### 5.6.2 Practical estimates

Expressions (5.11) and (5.12) now allow us to formally define the practical estimates for $m$ mentioned at the beginning of this section. We will now present a list of various estimators for the resample size $m$ when estimating the population mean with the statistic $\tilde{\theta}_{brag,1}$. The letters in parentheses indicate the abbreviations which will be used to represent these estimators.

1. **The bootstrap (BS) estimator:** The bootstrap estimate for $m$ will be based on the quantity given in (5.8). This estimator simply employs the bootstrap estimate of the two expected values which appear in that equation. It will be denoted with the subscript **BS** and it is given by:

**BS:**

$$\hat{m}_{1,BS} := \frac{1}{6} \cdot \frac{\mathrm{E}^* \left[\hat{\mu}_3^{*2} / \hat{\mu}_2^{*2}\right]}{\mathrm{E}^* \left[\bar{Y}_n^* \hat{\mu}_3^* / \hat{\mu}_2^*\right]}, \tag{5.13}$$

where the terms $\hat{\mu}_j^*$ are the sample moments based on the resampled bootstrap data.

When we apply the bootstrap to calculate (5.13) we can once again sample fewer than $n$ observations to determine the value. To avoid confusion, we will refer to the resample size used in the bootstrap calculation of the quantity (5.13) as $k$, thus leading to a $k$-out-of-$n$ bootstrap. Different choices of $k$ in this $k$-out-of-$n$ bootstrap lead to different estimators. In particular, we will consider three simple choices of $k$ for the bootstrap calculation of this estimator, i.e., $k = n/3$, $k = 2n/3$ and $k = n$. We will distinguish between the resulting estimators as follows:

- ⋆ **The *BS1* estimator:** When we use the $k$-out-of-$n$ bootstrap with $k = n/3$ in (5.13) then the resulting estimator is denoted $\hat{m}_{1,BS1}$.
- ⋆ **The *BS2* estimator:** When we use the $k$-out-of-$n$ bootstrap with $k = 2n/3$ in (5.13) then the resulting estimator is denoted $\hat{m}_{1,BS2}$.
- ⋆ **The *BS3* estimator:** When we use the traditional $n$-out-of-$n$ bootstrap, i.e., with $k = n$, in (5.13) then the resulting estimator is denoted $\hat{m}_{1,BS3}$.

2. **The naive biased (NB) estimator:** The second estimate of (5.8) involves using the form of $m$ found in (5.11). An estimator is then obtained by simply plugging in the sample estimates of the population parameters. The estimator will be called the 'naive biased' estimator, because it does not use 'good' estimators for the products of population moments. For example, the product $\mu_2^2 \mu_6$ is estimated by the biased estimator $\hat{\mu}_2^2 \hat{\mu}_6$.

This estimator for $m$ will be denoted with the subscript **NB** and it is expressed as follows:

**NB:**

$$\hat{m}_{1,NB} := \frac{n}{6} \cdot \frac{A(\hat{\mu}) + \frac{1}{n} B(\hat{\mu})}{C(\hat{\mu}) + \frac{1}{n} D(\hat{\mu})},$$

where $\hat{\mu}$ is the vector containing the sample moments $\hat{\mu} = (\hat{\mu}_2, \hat{\mu}_3, \hat{\mu}_4, \hat{\mu}_5, \hat{\mu}_6)$, and the functions $A(\hat{\mu})$, $B(\hat{\mu})$, $C(\hat{\mu})$ and $D(\hat{\mu})$ are defined as follows:

$$A(\hat{\mu}) = \frac{\hat{\mu}_3^2}{\hat{\mu}_2^2},$$

$$B(\hat{\mu}) = \frac{1}{\hat{\mu}_2^4} \left[\hat{\mu}_2^2 \hat{\mu}_6 + 9\hat{\mu}_2^5 + 8\hat{\mu}_2^2 \hat{\mu}_3^2 - 6\hat{\mu}_2^3 \hat{\mu}_4 + 3\hat{\mu}_3^2 \hat{\mu}_4 - 4\hat{\mu}_2 \hat{\mu}_3 \hat{\mu}_5\right],$$

$$C(\hat{\mu}) = \frac{1}{\hat{\mu}_2^2}\left[\hat{\mu}_2\hat{\mu}_4 - 3\hat{\mu}_2^3 - \hat{\mu}_3^2\right],$$

$$D(\hat{\mu}) = \frac{1}{\hat{\mu}_2^4}\left[3\hat{\mu}_2^3\hat{\mu}_4 - \hat{\mu}_2^2\hat{\mu}_6 - 3\hat{\mu}_2^5 + 3\hat{\mu}_2\hat{\mu}_3\hat{\mu}_5 + \hat{\mu}_2\hat{\mu}_4^2 - 3\hat{\mu}_3^2\hat{\mu}_4 - 5\hat{\mu}_2^2\hat{\mu}_3^2\right].$$

3. **The bias corrected (BC) estimators:** The third technique makes use of the alternative expression for $m_1^A$ given in (5.12). For the previous estimate we naively substituted sample moments for population moments to obtain an estimator. Unfortunately, the estimators obtained in this way are highly biased for small samples. To counteract this effect we will now estimate the product of population moments by estimators that have been slightly corrected for bias. Three separate estimators will be constructed from (5.12):

   ⋆ **The *BC0* estimator:** This estimator for $m$ contains uncorrected estimators for the population moments (similar to the previous estimator, but without the $1/n$ terms). This estimator is denoted by $\hat{m}_{1,BC0}$.

   ⋆ **The *BC1* estimator:** This estimator for $m$ contains estimators for the products of population moments that are 'corrected' to order $1/n$. This estimator is denoted by $\hat{m}_{1,BC1}$.

   ⋆ **The *BC2* estimator:** The final estimator for $m$ of this type contains estimators for the products of population moments that are 'corrected' to order $1/n^2$. This estimator is denoted by $\hat{m}_{1,BC2}$.

   The accuracy of the estimator will depend on the accuracy of the bias corrected estimates of the population moments.

   We will now present three different sets of bias corrected products of sample moments; the ones used in the estimator that are uncorrected will be denoted using the subscript **BC0**, those that are accurate up to order $1/n$ will be denoted using the subscript **BC1**, and the ones used for the estimator which are accurate up to order $1/n^2$ will be denoted using the subscript **BC2**. The products of population moments that we are required to 'correct' in this estimator are $\mu_2^2\mu_3^2$, $\mu_2^2\mu_6$, $\mu_2^3\mu_4$, $\mu_2^5$, $\mu_3^2\mu_4$, $\mu_2\mu_3\mu_5$, and $\mu_2\mu_4^2$.
   **Remark:** *The method of bias correction used to obtain these estimators consists of three parts:*

   *(i) Obtain the naive plug-in estimator for the product of population moments. For example, if we wish to estimate $\mu_2^2\mu_3^2$ then we obtain $\hat{\mu}_2^2\hat{\mu}_3^2$.*

   *(ii) Determine the expected value of this naive estimator. For example, the expected value of $\hat{\mu}_2^2\hat{\mu}_3^2$ is*

$$\begin{aligned}\mathrm{E}(\hat{\mu}_2^2\hat{\mu}_3^2) &= \mu_2^2\mu_3^2 + \frac{1}{n}\{\mu_2^2\mu_6 + 4\mu_2\mu_3\mu_5 + \mu_3^2\mu_4 - 26\mu_2^2\mu_3^2 - 6\mu_2^3\mu_4 + 9\mu_2^5\}\\ &\quad + \frac{1}{n^2}\{2\mu_2\mu_8 + \mu_4\mu_6 - 23\mu_2^2\mu_6 + 2\mu_3\mu_7 + 2\mu_5^2 - 74\mu_2\mu_3\mu_5\\ &\quad - 31\mu_3^2\mu_4 + 354\mu_2^2\mu_3^2 + 174\mu_2^3\mu_4 - 22\mu_2\mu_4^2 - 180\mu_2^5\} + O(n^{-3}).\end{aligned}$$

   *Obtaining expressions for these expected values is a* very *long and* incredibly *tedious process. Fortunately, all of the relevant expected values have been derived and are listed in Appendix D. The terms used in the derivation of these expressions are also provided.*

*(iii) To correct the naive plug-in estimator up to order, say $1/n^k$, we subtract the plug-in estimate of the $1/n^k$ terms of the expected value from the naive plug-in estimate of the product of population moments. For example, to correct the naive estimator $\hat{\mu}_2^2\hat{\mu}_3^2$ to order $1/n$, we subtract the plug-in estimator of the $1/n$ terms in $\mathrm{E}(\hat{\mu}_2^2\hat{\mu}_3^2)$ from $\hat{\mu}_2^2\hat{\mu}_3^2$. Since we correct up to the order of $1/n$, this estimator is of the **BC1** type, and will be denoted by*

$$\widehat{(\mu_2^2\mu_3^2)}_{BC1} := \hat{\mu}_2^2\hat{\mu}_3^2 - \frac{1}{n}\{\hat{\mu}_2^2\hat{\mu}_6 + 4\hat{\mu}_2\hat{\mu}_3\hat{\mu}_5 + \hat{\mu}_3^2\hat{\mu}_4 - 26\hat{\mu}_2^2\hat{\mu}_3^2 - 6\hat{\mu}_2^3\hat{\mu}_4 + 9\hat{\mu}_2^5\}.$$

**The *BC0* estimators for products of population moments:** The relevant uncorrected estimators are then simply the plug-in estimators as before, i.e.,

⋆ For $\mu_2^2\mu_3^2$, the uncorrected estimator is

$$\widehat{(\mu_2^2\mu_3^2)}_{BC0} := \hat{\mu}_2^2\hat{\mu}_3^2,$$

⋆ for $\mu_2^2\mu_6$, the uncorrected estimator is

$$\widehat{(\mu_2^2\mu_6)}_{BC0} := \hat{\mu}_2^2\hat{\mu}_6,$$

⋆ for $\mu_2^3\mu_4$, the uncorrected estimator is

$$\widehat{(\mu_2^3\mu_4)}_{BC0} := \hat{\mu}_2^3\hat{\mu}_4,$$

⋆ for $\mu_2^5$, the uncorrected estimator is

$$\widehat{(\mu_2^5)}_{BC0} := \hat{\mu}_2^5,$$

⋆ for $\mu_3^2\mu_4$, the uncorrected estimator is

$$\widehat{(\mu_3^2\mu_4)}_{BC0} := \hat{\mu}_3^2\hat{\mu}_4,$$

⋆ for $\mu_2\mu_3\mu_5$, the uncorrected estimator is

$$\widehat{(\mu_2\mu_3\mu_5)}_{BC0} := \hat{\mu}_2\hat{\mu}_3\hat{\mu}_5,$$

⋆ for $\mu_2\mu_4^2$, the uncorrected estimator is

$$\widehat{(\mu_2\mu_4^2)}_{BC0} := \hat{\mu}_2\hat{\mu}_4^2.$$

**The *BC1* estimators for products of population moments:** The relevant corrected estimators up to the order of $1/n$ (as explained above) are then:

⋆ For $\mu_2^2\mu_3^2$, the bias corrected estimator up to the order of $1/n$ is

$$\widehat{(\mu_2^2\mu_3^2)}_{BC1} := \hat{\mu}_2^2\hat{\mu}_3^2 - \frac{1}{n}\{\hat{\mu}_2^2\hat{\mu}_6 + 4\hat{\mu}_2\hat{\mu}_3\hat{\mu}_5 + \hat{\mu}_3^2\hat{\mu}_4 - 26\hat{\mu}_2^2\hat{\mu}_3^2 - 6\hat{\mu}_2^3\hat{\mu}_4 + 9\hat{\mu}_2^5\},$$

⋆ for $\mu_2^2\mu_6$, the bias corrected estimator up to the order of $1/n$ is

$$\widehat{(\mu_2^2\mu_6)}_{BC1} := \hat{\mu}_2^2\hat{\mu}_6 - \frac{1}{n}\{\hat{\mu}_4\hat{\mu}_6 + 2\hat{\mu}_2\hat{\mu}_8 - 11\hat{\mu}_2^2\hat{\mu}_6 - 12\hat{\mu}_2\hat{\mu}_3\hat{\mu}_5 + 15\hat{\mu}_2^3\hat{\mu}_4\},$$

⋆ for $\mu_2^3\mu_4$, the bias corrected estimator up to the order of $1/n$ is

$$\widehat{(\mu_2^3\mu_4)}_{BC1} := \hat{\mu}_2^3\hat{\mu}_4 - \frac{1}{n}\{-13\hat{\mu}_2^3\hat{\mu}_4 + 3\hat{\mu}_2^2\hat{\mu}_6 + 3\hat{\mu}_2\hat{\mu}_4^2 - 12\hat{\mu}_2^2\hat{\mu}_3^2 + 6\hat{\mu}_2^5\},$$

⋆ for $\mu_2^5$, the bias corrected estimator up to the order of $1/n$ is

$$\widehat{(\mu_2^5)}_{BC1} := \hat{\mu}_2^5 - \frac{1}{n}\{-15\hat{\mu}_2^5 + 10\hat{\mu}_2^3\hat{\mu}_4\},$$

⋆ for $\mu_3^2\mu_4$, the bias corrected estimator up to the order of $1/n$ is

$$\widehat{(\mu_3^2\mu_4)}_{BC1} := \hat{\mu}_3^2\hat{\mu}_4 - \frac{1}{n}\{-21\hat{\mu}_3^2\hat{\mu}_4 + 2\hat{\mu}_3\hat{\mu}_7 + \hat{\mu}_4\hat{\mu}_6 + 30\hat{\mu}_2^2\hat{\mu}_3^2 - 18\hat{\mu}_2\hat{\mu}_4^2 - 18\hat{\mu}_2\hat{\mu}_3\hat{\mu}_5 + 9\hat{\mu}_2^3\hat{\mu}_4\},$$

⋆ for $\mu_2\mu_3\mu_5$, the bias corrected estimator up to the order of $1/n$ is

$$\begin{aligned}(\widehat{\mu_2\mu_3\mu_5})_{BC1} \;:=\; & \hat{\mu}_2\hat{\mu}_3\hat{\mu}_5 - \frac{1}{n}\{\hat{\mu}_3\hat{\mu}_7 + \hat{\mu}_2\hat{\mu}_8 - 15\hat{\mu}_2\hat{\mu}_3\hat{\mu}_5 + 15\hat{\mu}_2^3\hat{\mu}_4 + \hat{\mu}_5^2 - 5\hat{\mu}_3^2\hat{\mu}_4 \\ & -5\hat{\mu}_2\hat{\mu}_4^2 + 10\hat{\mu}_2^2\hat{\mu}_3^2 - 3\hat{\mu}_2^2\hat{\mu}_6\},\end{aligned}$$

⋆ for $\mu_2\mu_4^2$, the bias corrected estimator up to the order of $1/n$ is

$$\widehat{(\mu_2\mu_4^2)}_{BC1} := \hat{\mu}_2\hat{\mu}_4^2 - \frac{1}{n}\{-12\hat{\mu}_2\hat{\mu}_4^2 + 2\hat{\mu}_4\hat{\mu}_6 + \hat{\mu}_2\hat{\mu}_8 - 8\hat{\mu}_3^2\hat{\mu}_4 - 8\hat{\mu}_2\hat{\mu}_3\hat{\mu}_5 + 12\hat{\mu}_2^3\hat{\mu}_4 + 16\hat{\mu}_2^2\hat{\mu}_3^2\}.$$

**The $BC2$ estimators for products of population moments:** The corrected estimators up to the order of $1/n^2$ are then

⋆ For $\mu_2^2\mu_3^2$, the bias corrected estimator up to the order of $1/n^2$ is

$$\begin{aligned}\widehat{(\mu_2^2\mu_3^2)}_{BC2} \;:=\; & \hat{\mu}_2^2\hat{\mu}_3^2 - \frac{1}{n}\{\hat{\mu}_2^2\hat{\mu}_6 + 4\hat{\mu}_2\hat{\mu}_3\hat{\mu}_5 + \hat{\mu}_3^2\hat{\mu}_4 - 26\hat{\mu}_2^2\hat{\mu}_3^2 - 6\hat{\mu}_2^3\hat{\mu}_4 + 9\hat{\mu}_2^5\} \\ & -\frac{1}{n^2}\{2\hat{\mu}_2\hat{\mu}_8 + \hat{\mu}_4\hat{\mu}_6 - 23\hat{\mu}_2^2\hat{\mu}_6 + 2\hat{\mu}_3\hat{\mu}_7 + 2\hat{\mu}_5^2 - 74\hat{\mu}_2\hat{\mu}_3\hat{\mu}_5 \\ & -31\hat{\mu}_3^2\hat{\mu}_4 + 354\hat{\mu}_2^2\hat{\mu}_3^2 + 174\hat{\mu}_2^3\hat{\mu}_4 - 22\hat{\mu}_2\hat{\mu}_4^2 - 180\hat{\mu}_2^5\},\end{aligned}$$

⋆ for $\mu_2^3\mu_4$, the bias corrected estimator up to the order of $1/n^2$ is

$$\begin{aligned}\widehat{(\mu_2^3\mu_4)}_{BC2} \;:=\; & \hat{\mu}_2^3\hat{\mu}_4 - \frac{1}{n}\{-13\hat{\mu}_2^3\hat{\mu}_4 + 3\hat{\mu}_2^2\hat{\mu}_6 + 3\hat{\mu}_2\hat{\mu}_4^2 - 12\hat{\mu}_2^2\hat{\mu}_3^2 + 6\hat{\mu}_2^5\} \\ & -\frac{1}{n^2}\{228\hat{\mu}_2^2\hat{\mu}_3^2 + 158\hat{\mu}_2^3\hat{\mu}_4 - 18\hat{\mu}_3^2\hat{\mu}_4 - 30\hat{\mu}_2\hat{\mu}_4^2 - 48\hat{\mu}_2\hat{\mu}_3\hat{\mu}_5 \\ & -30\hat{\mu}_2^2\hat{\mu}_6 + 3\hat{\mu}_2\hat{\mu}_8 + 4\hat{\mu}_4\hat{\mu}_6 - 123\hat{\mu}_2^5\},\end{aligned}$$

⋆ for $\mu_2^5$, the bias corrected estimator up to the order of $1/n^2$ is

$$\widehat{(\mu_2^5)}_{BC2} := \hat{\mu}_2^5 - \frac{1}{n}\{-15\hat{\mu}_2^5 + 10\hat{\mu}_2^3\hat{\mu}_4\} - \frac{1}{n^2}\{115\hat{\mu}_2^5 - 110\hat{\mu}_2^3\hat{\mu}_4 + 15\hat{\mu}_2\hat{\mu}_4^2 + 10\hat{\mu}_2^2\hat{\mu}_6 - 60\hat{\mu}_2^2\hat{\mu}_3^2\}.$$

The three estimators will then be denoted with the subscripts **BC0**, **BC1** and **BC2**, and are given by:

**BC0:**

$$\hat{m}_{1,BC0} := \frac{n}{6} \cdot \frac{\tilde{A}(\hat{\boldsymbol{\mu}}_{BC0})}{\tilde{C}(\hat{\boldsymbol{\mu}}_{BC0})},$$

where

$\tilde{A}(\hat{\boldsymbol{\mu}}_{BC0}) = \widehat{(\mu_2^2\mu_3^2)}_{BC0}$,

$\tilde{C}(\hat{\boldsymbol{\mu}}_{BC0}) = \widehat{(\mu_2^3\mu_4)}_{BC0} - 3\widehat{(\mu_2^5)}_{BC0} - \widehat{(\mu_2^2\mu_3^2)}_{BC0}$,

and where $\hat{\boldsymbol{\mu}}_{BC0}$ is the vector consisting of the uncorrected products of sample moments, $\hat{\boldsymbol{\mu}}_{BC0} = \Big(\widehat{(\mu_2^2\mu_3^2)}_{BC0}, \widehat{(\mu_2^2\mu_6)}_{BC0}, \widehat{(\mu_2^3\mu_4)}_{BC0}, \widehat{(\mu_2^5)}_{BC0}, \widehat{(\mu_3^2\mu_4)}_{BC0}, \widehat{(\mu_2\mu_3\mu_5)}_{BC0}, \widehat{(\mu_2\mu_4^2)}_{BC0}\Big)$.

**BC1:**

$$\hat{m}_{1,BC1} := \frac{n}{6} \cdot \frac{\tilde{A}(\hat{\boldsymbol{\mu}}_{BC1}) + \frac{1}{n}\tilde{B}(\hat{\boldsymbol{\mu}}_{BC0})}{\tilde{C}(\hat{\boldsymbol{\mu}}_{BC1}) + \frac{1}{n}\tilde{D}(\hat{\boldsymbol{\mu}}_{BC0})},$$

where

$\tilde{A}(\hat{\boldsymbol{\mu}}_{BC1}) = \widehat{(\mu_2^2\mu_3^2)}_{BC1}$,

$\tilde{B}(\hat{\boldsymbol{\mu}}_{BC0}) = \widehat{(\mu_2^2\mu_6)}_{BC0} + 9\widehat{(\mu_2^5)}_{BC0} + 8\widehat{(\mu_2^2\mu_3^2)}_{BC0} - 6\widehat{(\mu_2^3\mu_4)}_{BC0} + 3\widehat{(\mu_3^2\mu_4)}_{BC0} - 4\widehat{(\mu_2\mu_3\mu_5)}_{BC0}$,

$\tilde{C}(\hat{\boldsymbol{\mu}}_{BC1}) = \widehat{(\mu_2^3\mu_4)}_{BC1} - 3\widehat{(\mu_2^5)}_{BC1} - \widehat{(\mu_2^2\mu_3^2)}_{BC1}$,

$$\begin{aligned}\tilde{D}(\hat{\boldsymbol{\mu}}_{BC0}) &= 3\widehat{(\mu_2^3\mu_4)}_{BC0} - \widehat{(\mu_2^2\mu_6)}_{BC0} - 3\widehat{(\mu_2^5)}_{BC0} + 3\widehat{(\mu_2\mu_3\mu_5)}_{BC0} + \widehat{(\mu_2\mu_4^2)}_{BC0} \\ &\quad -3\widehat{(\mu_3^2\mu_4)}_{BC0} - 5\widehat{(\mu_2^2\mu_3^2)}_{BC0},\end{aligned}$$

and where $\hat{\boldsymbol{\mu}}_{BC0}$ is the vector consisting of the uncorrected products of sample moments, $\hat{\boldsymbol{\mu}}_{BC0} = \Big(\widehat{(\mu_2^2\mu_3^2)}_{BC0}, \widehat{(\mu_2^2\mu_6)}_{BC0}, \widehat{(\mu_2^3\mu_4)}_{BC0}, \widehat{(\mu_2^5)}_{BC0}, \widehat{(\mu_3^2\mu_4)}_{BC0}, \widehat{(\mu_2\mu_3\mu_5)}_{BC0}, \widehat{(\mu_2\mu_4^2)}_{BC0}\Big)$ and $\hat{\boldsymbol{\mu}}_{BC1}$ is the vector consisting of the bias corrected products of sample moments to the first order, $\hat{\boldsymbol{\mu}}_{BC1} = \Big(\widehat{(\mu_2^2\mu_3^2)}_{BC1}, \widehat{(\mu_2^2\mu_6)}_{BC1}, \widehat{(\mu_2^3\mu_4)}_{BC1}, \widehat{(\mu_2^5)}_{BC1}, \widehat{(\mu_3^2\mu_4)}_{BC1}, \widehat{(\mu_2\mu_3\mu_5)}_{BC1}, \widehat{(\mu_2\mu_4^2)}_{BC1}\Big)$.

**BC2:**

$$\hat{m}_{1,BC2} := \frac{n}{6} \cdot \frac{\tilde{A}(\hat{\hat{\mu}}_{BC2}) + \frac{1}{n}\tilde{B}(\hat{\hat{\mu}}_{BC1})}{\tilde{C}(\hat{\hat{\mu}}_{BC2}) + \frac{1}{n}\tilde{D}(\hat{\hat{\mu}}_{BC1})},$$

where

$$\tilde{A}(\hat{\hat{\mu}}_{BC2}) = \widehat{(\mu_2^2\mu_3^2)}_{BC2},$$

$$\tilde{B}(\hat{\hat{\mu}}_{BC1}) = \widehat{(\mu_2^2\mu_6)}_{BC1} + 9\widehat{(\mu_2^5)}_{BC1} + 8\widehat{(\mu_2^2\mu_3^2)}_{BC1} - 6\widehat{(\mu_2^3\mu_4)}_{BC1} + 3\widehat{(\mu_3^2\mu_4)}_{BC1} - 4\widehat{(\mu_2\mu_3\mu_5)}_{BC1},$$

$$\tilde{C}(\hat{\hat{\mu}}_{BC2}) = \widehat{(\mu_2^3\mu_4)}_{BC2} - 3\widehat{(\mu_2^5)}_{BC2} - \widehat{(\mu_2^2\mu_3^2)}_{BC2},$$

$$\tilde{D}(\hat{\hat{\mu}}_{BC1}) = 3\widehat{(\mu_2^3\mu_4)}_{BC1} - \widehat{(\mu_2^2\mu_6)}_{BC1} - 3\widehat{(\mu_2^5)}_{BC1} + 3\widehat{(\mu_2\mu_3\mu_5)}_{BC1} + \widehat{(\mu_2\mu_4^2)}_{BC1}$$
$$-3\widehat{(\mu_3^2\mu_4)}_{BC1} - 5\widehat{(\mu_2^2\mu_3^2)}_{BC1},$$

and where $\hat{\hat{\mu}}_{BC1}$ is the vector consisting of the bias corrected products of sample moments to the first order, $\hat{\hat{\mu}}_{BC1} = \left(\widehat{(\mu_2^2\mu_3^2)}_{BC1}, \widehat{(\mu_2^2\mu_6)}_{BC1}, \widehat{(\mu_2^3\mu_4)}_{BC1}, \widehat{(\mu_2^5)}_{BC1}, \widehat{(\mu_3^2\mu_4)}_{BC1}, \widehat{(\mu_2\mu_3\mu_5)}_{BC1}, \widehat{(\mu_2\mu_4^2)}_{BC1}\right)$ and $\hat{\hat{\mu}}_{BC2}$ is the vector consisting of the bias corrected products of sample moments to the second order, $\hat{\hat{\mu}}_{BC2} = \left(\widehat{(\mu_2^2\mu_3^2)}_{BC2}, \widehat{(\mu_2^2\mu_6)}_{BC2}, \widehat{(\mu_2^3\mu_4)}_{BC2}, \widehat{(\mu_2^5)}_{BC2}, \widehat{(\mu_3^2\mu_4)}_{BC2}, \widehat{(\mu_2\mu_3\mu_5)}_{BC2}, \widehat{(\mu_2\mu_4^2)}_{BC2}\right)$.

4. **The '*UN*biased' numerator & *B*ias *C*orrected denominator (UNBC) estimators:** The final estimator makes use of the fact that the ratio $\hat{\mu}_3^2/\hat{\mu}_2^2$ is an unbiased estimator of the parameter $\mathrm{E}\left(\hat{\mu}_3^2/\hat{\mu}_2^2\right)$. This fact leads us to a completely unbiased estimator of the numerator in (5.8). The denominator is then estimated using the bias correction techniques discussed in the previous methods. Once again we will have two versions of this statistic: the first will estimate the denominator up to order $1/n$ and the second will estimate it up to order $1/n^2$. The estimators will be denoted with the subscripts **UNBC1** and **UNBC2** which are used to indicate that the numerator is *UN*biased, and the denominator is *B*ias *C*orrected up to order 1 and 2 respectively. The estimators are then given by:

**UNBC1:**

$$\hat{m}_{1,UNBC1} := \frac{n}{6} \cdot \frac{\hat{\mu}_3^2/\hat{\mu}_2^2}{\tilde{C}(\hat{\hat{\mu}}_{BC1}) + \frac{1}{n}\tilde{D}(\hat{\hat{\mu}}_{BC0})},$$

where

$$\tilde{C}(\hat{\hat{\mu}}_{BC1}) = \widehat{(\mu_2^3\mu_4)}_{BC1} - 3\widehat{(\mu_2^5)}_{BC1} - \widehat{(\mu_2^2\mu_3^2)}_{BC1},$$

$$\begin{aligned}\tilde{D}(\hat{\hat{\mu}}_{BC0}) &= 3\widehat{(\mu_2^3\mu_4)}_{BC0} - \widehat{(\mu_2^2\mu_6)}_{BC0} - 3\widehat{(\mu_2^5)}_{BC0} + 3\widehat{(\mu_2\mu_3\mu_5)}_{BC0} + \widehat{(\mu_2\mu_4^2)}_{BC0} \\ &\quad -3\widehat{(\mu_3^2\mu_4)}_{BC0} - 5\widehat{(\mu_2^2\mu_3^2)}_{BC0},\end{aligned}$$

and where $\hat{\hat{\mu}}_{BC0}$ is the vector consisting of the uncorrected products of sample moments, $\hat{\hat{\mu}}_{BC0} = \left(\widehat{(\mu_2^2\mu_3^2)}_{BC0}, \widehat{(\mu_2^2\mu_6)}_{BC0}, \widehat{(\mu_2^3\mu_4)}_{BC0}, \widehat{(\mu_2^5)}_{BC0}, \widehat{(\mu_3^2\mu_4)}_{BC0}, \widehat{(\mu_2\mu_3\mu_5)}_{BC0}, \widehat{(\mu_2\mu_4^2)}_{BC0}\right)$ and $\hat{\hat{\mu}}_{BC1}$ is the vector consisting of the bias corrected products of sample moments to the first order, $\hat{\hat{\mu}}_{BC1} = \left(\widehat{(\mu_2^2\mu_3^2)}_{BC1}, \widehat{(\mu_2^2\mu_6)}_{BC1}, \widehat{(\mu_2^3\mu_4)}_{BC1}, \widehat{(\mu_2^5)}_{BC1}, \widehat{(\mu_3^2\mu_4)}_{BC1}, \widehat{(\mu_2\mu_3\mu_5)}_{BC1}, \widehat{(\mu_2\mu_4^2)}_{BC1}\right)$ as defined previously.

**UNBC2:**

$$\hat{m}_{1,UNBC2} := \frac{n}{6} \cdot \frac{\hat{\mu}_3^2/\hat{\mu}_2^2}{\tilde{C}(\hat{\hat{\mu}}_{BC2}) + \frac{1}{n}\tilde{D}(\hat{\hat{\mu}}_{BC1})},$$

where

$$\tilde{C}(\hat{\hat{\mu}}_{BC2}) = \widehat{(\mu_2^3\mu_4)}_{BC2} - 3\widehat{(\mu_2^5)}_{BC2} - \widehat{(\mu_2^2\mu_3^2)}_{BC2},$$

$$\begin{aligned}\tilde{D}(\hat{\hat{\mu}}_{BC1}) &= 3\widehat{(\mu_2^3\mu_4)}_{BC1} - \widehat{(\mu_2^2\mu_6)}_{BC1} - 3\widehat{(\mu_2^5)}_{BC1} + 3\widehat{(\mu_2\mu_3\mu_5)}_{BC1} + \widehat{(\mu_2\mu_4^2)}_{BC1} \\ &\quad -3\widehat{(\mu_3^2\mu_4)}_{BC1} - 5\widehat{(\mu_2^2\mu_3^2)}_{BC1},\end{aligned}$$

and where $\hat{\hat{\mu}}_{BC1}$ is the vector consisting of the bias corrected products of sample moments to the first order, $\hat{\hat{\mu}}_{BC1} = \left(\widehat{(\mu_2^2\mu_3^2)}_{BC1}, \widehat{(\mu_2^2\mu_6)}_{BC1}, \widehat{(\mu_2^3\mu_4)}_{BC1}, \widehat{(\mu_2^5)}_{BC1}, \widehat{(\mu_3^2\mu_4)}_{BC1}, \widehat{(\mu_2\mu_3\mu_5)}_{BC1}, \widehat{(\mu_2\mu_4^2)}_{BC1}\right)$ and $\hat{\hat{\mu}}_{BC2}$ is the vector consisting of the bias corrected products of sample moments to the second order, $\hat{\hat{\mu}}_{BC2} = \left(\widehat{(\mu_2^2\mu_3^2)}_{BC2}, \widehat{(\mu_2^2\mu_6)}_{BC2}, \widehat{(\mu_2^3\mu_4)}_{BC2}, \widehat{(\mu_2^5)}_{BC2}, \widehat{(\mu_3^2\mu_4)}_{BC2}, \widehat{(\mu_2\mu_3\mu_5)}_{BC2}, \widehat{(\mu_2\mu_4^2)}_{BC2}\right)$.

### 5.6.3 Truncation

The nature of the sampling variation of the $m$-out-of-$n$ bootstrap resample size estimators listed in the previous section is such that, occasionally, they lead to very small choices of the resample size. This is obviously problematic and it must be rectified. The solution to the problem that we use involves using a simple truncation of the estimated $m$ values.

**Truncation for the choice of the resample size for $\tilde{\theta}_{brag,1}$:** This truncation is given by

$$\hat{m}_{1,X,trunc} = \begin{cases} np_n, & \text{if } \hat{m}_{1,X} \leq np_n \\ \hat{m}_{1,X}, & \text{if } np_n < \hat{m}_{1,X} \leq n \\ n, & \text{if } \hat{m}_{1,X} > n, \end{cases} \tag{5.14}$$

where $p_n$ is some function of $n$ decreasing to zero at a slow rate such that $np_n \to \infty$, and $\hat{m}_{1,X}$ is any one of the data-dependent choices of $m$ discussed in the previous section, i.e., $X$ can be BS1, BS2, BS3, NB, BC0, BC1, BC2, UNBC1 or UNBC2. One possible choice, suggested by a referee, is to choose $p_n = q^{j_n}$, with $j_n \to \infty$ (slowly enough so that $np_n = nq^{j_n} \to \infty$) and $q \in (0, 1)$. However, there are a multitude of other choices for $p_n$.

**Remark:**

- ⋆ According to, among others, del Barrio, Cuesta-Albertos and Matrán (2002), the choice of $m$ should not be too small because it can lead to instability of the estimators.
- ⋆ With the above remark in mind, and for the practical purposes of the Monte-Carlo studies, we will set the value $p_n$ to be equal to the constant values 0.1 (or 10% of the original sample size) or 0.2 (or 20% of the original sample size).

**Truncation for the choice of the resample size for $\tilde{\theta}_{brag,2}$:** The truncation applied to $\tilde{\theta}_{brag,2}$, the *corrected* $m$-out-of-$n$ bootstrap estimator, is slightly different from the one applied to $\tilde{\theta}_{brag,1}$, the $m$-out-of-$n$ bootstrap estimator, since this we will define the estimator as

$$\hat{m}_{2,X} = \frac{1}{n}\left[\hat{m}_{1,X,trunc}\right]^2,$$

where $X$ can be any one of BS1, BS2, BS3, NB, BC0, BC1, BC2, UNBC1 or UNBC2.

Clearly, the estimator $\hat{m}_{2,X}$ will always be less than or equal to the original truncated estimator $\hat{m}_{1,X,trunc}$, which follows from the relationship between $\hat{m}_{2,X}$ and $\hat{m}_{1,X,trunc}$ stated above. The result is that a different truncation has to be applied to the corrected $m$-out-of-$n$ estimator:

$$\hat{m}_{2,X,trunc} = \begin{cases} 1, & \text{if } \hat{m}_{2,X} \leq 1 \\ \hat{m}_{2,X}, & \text{if } 1 < \hat{m}_{2,X} \leq n \\ n, & \text{if } \hat{m}_{2,X} > n. \end{cases}$$

## 5.7 Monte-Carlo simulation results

The performance of the data-dependent choices of $m$ discussed in the previous section is evaluated in this section by using Monte-Carlo simulations. As in Section 5.4, the performance of the $m$-out-of-$n$ bootstrap and the corrected $m$-out-of-$n$ bootstrap using the various choices of $m$ is measured by calculating the ratios

$$\zeta(\tilde{\theta}_{brag,1}) = \frac{\text{MSE}(\bar{X}_n)}{\text{MSE}(\tilde{\theta}_{brag,1})},$$

where $\tilde{\theta}_{brag,1}$ is calculated using one of the estimated choices of $m$. If $\zeta(\tilde{\theta}_{brag,1})$ is greater than 1, then it indicates that $\tilde{\theta}_{brag,1}$ performs better (in a mean squared error sense) than the sample mean when estimating the population mean. Similarly, if $\zeta(\tilde{\theta}_{brag,2})$ is greater than 1 it indicates that $\tilde{\theta}_{brag,2}$ performs better than the sample mean in estimating the population mean.

The configurations for this set of Monte-Carlo simulations are:

⋆ The number of Monte-Carlo simulations performed for each calculation in the tables is $MC = 3000$.

⋆ The sample sizes used are $n = 20, 50, 100, 500$ and 1000.

⋆ The distributions used are the double exponential distribution, $F$-distribution, normal distribution and the contaminated normal distribution (the specific parameter choices for each of these distributions are given in each table's caption).

⋆ The values of $p_n = 0.1$ and 0.2 are chosen as the lower bound truncation proportions (as discussed in equation (5.14)), i.e., the estimated resample size lower bound for the $m$-out-of-$n$ bootstrap estimate is $np_n$.

Displayed in each of these tables is the value $\zeta(\tilde{\theta}_{brag,1})$ or $\zeta(\tilde{\theta}_{brag,2})$ for the different choices of $m$. The mean of the data-dependent choice of $m$ and its standard error calculated over the Monte-Carlo iterations are also provided in a separate table.

| MSE ratios for the different estimates of $m$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $n$ | BS1 | BS2 | BS3 | NB | BC0 | BC1 | BC2 | UNBC1 | UNBC2 |
| 20 | 1.294 | 1.239 | 1.200 | 1.221 | 1.224 | 1.292 | 1.182 | 1.278 | 1.231 |
| 50 | 1.167 | 1.149 | 1.118 | 1.131 | 1.122 | 1.163 | 1.152 | 1.137 | 1.113 |
| 100 | 1.089 | 1.106 | 1.083 | 1.088 | 1.085 | 1.094 | 1.072 | 1.086 | 1.063 |
| 500 | 1.018 | 1.014 | 1.018 | 1.025 | 1.019 | 1.028 | 1.053 | 1.026 | 1.032 |
| 1000 | 1.012 | 0.989 | 1.012 | 0.994 | 0.992 | 1.013 | 1.041 | 1.002 | 1.021 |

| Mean and standard error of chosen $m$ values | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Stat | $n$ | BS1 | BS2 | BS3 | NB | BC0 | BC1 | BC2 | UNBC1 | UNBC2 |
| $\bar{m}$ | 20 | 2.524 | 3.29 | 4.14 | 3.685 | 4.139 | 2.234 | 8.067 | 2.343 | 4.243 |
| $SE(\bar{m})$ | 20 | 0.042 | 0.062 | 0.081 | 0.074 | 0.089 | 0.03 | 0.12 | 0.031 | 0.092 |
| $\bar{m}$ | 50 | 5.501 | 6.249 | 7.273 | 7.054 | 7.097 | 5.279 | 6.788 | 5.913 | 7.51 |
| $SE(\bar{m})$ | 50 | 0.062 | 0.099 | 0.129 | 0.123 | 0.121 | 0.045 | 0.124 | 0.061 | 0.127 |
| $\bar{m}$ | 100 | 10.233 | 10.509 | 10.827 | 10.876 | 10.826 | 10.203 | 10.553 | 10.627 | 11.303 |
| $SE(\bar{m})$ | 100 | 0.062 | 0.095 | 0.090 | 0.110 | 0.100 | 0.065 | 0.410 | 0.107 | 0.123 |
| $\bar{m}$ | 500 | 50 | 50 | 50.005 | 50 | 50 | 50 | 50 | 50 | 50 |
| $SE(\bar{m})$ | 500 | 0 | 0 | 0.005 | 0 | 0 | 0 | 0 | 0 | 0 |
| $\bar{m}$ | 1000 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| $SE(\bar{m})$ | 1000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

**Table 5.6:** *MOON: Data generated: Double exponential* $\mu = 0$, $\sigma = \sqrt{2}$. *The values for the lower bound of truncation of* $\hat{m}$, *i.e., lower bound* $= np_n$ *where* $p_n =$ *0.1.*

| MSE ratios for the different estimates of $m$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $n$ | BS1 | BS2 | BS3 | NB | BC0 | BC1 | BC2 | UNBC1 | UNBC2 |
| 20 | 1.641 | 1.931 | 1.820 | 1.673 | 1.680 | 1.638 | 1.614 | 2.020 | 1.690 |
| 50 | 1.594 | 1.906 | 1.804 | 1.405 | 1.488 | 1.574 | 1.156 | 1.828 | 1.687 |
| 100 | 1.541 | 1.742 | 1.715 | 1.375 | 1.385 | 1.520 | 1.204 | 1.390 | 1.663 |
| 500 | 1.411 | 1.484 | 1.499 | 1.364 | 1.294 | 1.446 | 1.119 | 1.284 | 1.444 |
| 1000 | 1.344 | 1.364 | 1.336 | 1.152 | 1.092 | 1.385 | 1.115 | 1.076 | 1.381 |

| Mean and standard error of chosen $m$ values | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Stat | $n$ | BS1 | BS2 | BS3 | NB | BC0 | BC1 | BC2 | UNBC1 | UNBC2 |
| $\bar{m}$ | 20 | 2.797 | 5.173 | 7.294 | 7.278 | 4.826 | 2.758 | 2.306 | 3.968 | 2.249 |
| $SE(\bar{m})$ | 20 | 0.053 | 0.089 | 0.113 | 0.126 | 0.118 | 0.051 | 0.042 | 0.051 | 0.037 |
| $\bar{m}$ | 50 | 7.341 | 14.976 | 21.963 | 24.820 | 32.202 | 7.520 | 25.831 | 20.715 | 30.568 |
| $SE(\bar{m})$ | 50 | 0.130 | 0.221 | 0.264 | 0.296 | 0.359 | 0.147 | 0.377 | 0.148 | 0.379 |
| $\bar{m}$ | 100 | 14.242 | 30.059 | 46.470 | 50.500 | 71.083 | 16.124 | 46.515 | 62.286 | 64.111 |
| $SE(\bar{m})$ | 100 | 0.225 | 0.393 | 0.481 | 0.506 | 0.519 | 0.308 | 0.738 | 0.292 | 0.698 |
| $\bar{m}$ | 500 | 57.728 | 112.375 | 173.691 | 162.815 | 211.517 | 79.462 | 190.512 | 325.148 | 264.910 |
| $SE(\bar{m})$ | 500 | 0.549 | 1.218 | 1.628 | 1.120 | 1.436 | 0.947 | 3.356 | 1.985 | 2.742 |
| $\bar{m}$ | 1000 | 108.892 | 200.623 | 307.611 | 280.581 | 347.364 | 164.227 | 347.911 | 560.689 | 454.655 |
| $SE(\bar{m})$ | 1000 | 0.710 | 2.266 | 2.986 | 1.800 | 2.568 | 2.016 | 6.369 | 4.360 | 4.741 |

**Table 5.7:** *MOON: Data generated=* $F(8,5)$*. The values for the lower bound of truncation of* $\hat{m}$*, i.e., lower bound* $= np_n$ *where* $p_n = 0.1$*.*

| MSE ratios for the different estimates of $m$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $n$ | BS1 | BS2 | BS3 | NB | BC0 | BC1 | BC2 | UNBC1 | UNBC2 |
| 20 | 0.920 | 0.914 | 0.922 | 0.917 | 0.919 | 0.918 | 0.915 | 0.920 | 0.910 |
| 50 | 0.965 | 0.968 | 0.975 | 0.971 | 0.977 | 0.959 | 0.994 | 0.965 | 0.994 |
| 100 | 1.035 | 1.035 | 1.033 | 1.036 | 1.039 | 1.036 | 1.000 | 1.034 | 0.993 |
| 500 | 1.004 | 0.996 | 1.000 | 0.998 | 0.995 | 0.998 | 1.015 | 0.998 | 0.997 |
| 1000 | 1.001 | 0.997 | 0.996 | 0.995 | 0.986 | 0.988 | 1.019 | 0.987 | 1.027 |

| Mean and standard error of chosen $m$ values | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Stat | $n$ | BS1 | BS2 | BS3 | NB | BC0 | BC1 | BC2 | UNBC1 | UNBC2 |
| $\bar{m}$ | 20 | 2.699 | 2.894 | 3.159 | 2.957 | 2.637 | 2.413 | 5.758 | 2.509 | 3.057 |
| $SE(\bar{m})$ | 20 | 0.051 | 0.057 | 0.065 | 0.061 | 0.052 | 0.040 | 0.113 | 0.042 | 0.064 |
| $\bar{m}$ | 50 | 6.315 | 6.712 | 6.872 | 7.073 | 6.184 | 5.780 | 8.127 | 6.275 | 6.614 |
| $SE(\bar{m})$ | 50 | 0.109 | 0.124 | 0.131 | 0.140 | 0.108 | 0.086 | 0.165 | 0.106 | 0.121 |
| $\bar{m}$ | 100 | 11.956 | 12.713 | 13.188 | 13.449 | 11.518 | 11.695 | 13.006 | 12.260 | 12.057 |
| $SE(\bar{m})$ | 100 | 0.186 | 0.227 | 0.239 | 0.253 | 0.164 | 0.183 | 0.233 | 0.207 | 0.193 |
| $\bar{m}$ | 500 | 55.801 | 54.912 | 56.844 | 57.280 | 53.493 | 54.119 | 54.806 | 54.402 | 54.645 |
| $SE(\bar{m})$ | 500 | 0.784 | 0.667 | 0.800 | 0.821 | 0.570 | 0.671 | 0.706 | 0.697 | 0.701 |
| $\bar{m}$ | 1000 | 106.878 | 107.155 | 107.097 | 109.686 | 104.694 | 107.131 | 104.831 | 107.351 | 104.773 |
| $SE(\bar{m})$ | 1000 | 1.141 | 1.188 | 1.183 | 1.366 | 0.917 | 1.252 | 0.950 | 1.270 | 0.943 |

**Table 5.8:** *MOON: Data generated= Standard Normal. The values for the lower bound of truncation of* $\hat{m}$*, i.e., lower bound* $= np_n$ *where* $p_n = 0.1$*.*

| MSE ratios for the different estimates of $m$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $n$ | **BS1** | **BS2** | **BS3** | **NB** | **BC0** | **BC1** | **BC2** | **UNBC1** | **UNBC2** |
| 20 | 4.626 | 2.884 | 2.007 | 1.963 | 2.981 | 4.227 | 3.787 | 3.316 | 4.137 |
| 50 | 3.198 | 2.164 | 1.662 | 1.633 | 1.472 | 3.257 | 1.809 | 1.551 | 1.586 |
| 100 | 2.097 | 1.776 | 1.528 | 1.538 | 1.416 | 2.082 | 1.766 | 1.297 | 1.337 |
| 500 | 1.131 | 1.131 | 1.124 | 1.121 | 1.111 | 1.122 | 1.138 | 1.118 | 1.121 |
| 1000 | 1.044 | 1.046 | 1.057 | 1.049 | 1.046 | 1.048 | 1.019 | 1.041 | 1.014 |

| Mean and standard error of chosen $m$ values | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Stat | $n$ | **BS1** | **BS2** | **BS3** | **NB** | **BC0** | **BC1** | **BC2** | **UNBC1** | **UNBC2** |
| $\bar{\hat{m}}$ | 20 | 2.200 | 3.024 | 4.835 | 5.672 | 6.663 | 2.190 | 5.239 | 2.746 | 3.456 |
| $SE(\bar{\hat{m}})$ | 20 | 0.025 | 0.046 | 0.077 | 0.106 | 0.129 | 0.027 | 0.113 | 0.031 | 0.079 |
| $\bar{\hat{m}}$ | 50 | 5.058 | 6.301 | 9.282 | 9.724 | 15.438 | 6.382 | 18.871 | 11.952 | 14.067 |
| $SE(\bar{\hat{m}})$ | 50 | 0.022 | 0.064 | 0.137 | 0.172 | 0.277 | 0.129 | 0.351 | 0.146 | 0.286 |
| $\bar{\hat{m}}$ | 100 | 10.017 | 11.091 | 13.842 | 13.597 | 17.380 | 13.080 | 25.309 | 24.353 | 24.998 |
| $SE(\bar{\hat{m}})$ | 100 | 0.006 | 0.074 | 0.173 | 0.179 | 0.306 | 0.273 | 0.562 | 0.397 | 0.482 |
| $\bar{\hat{m}}$ | 500 | 50 | 50.043 | 50.301 | 50.258 | 50.[illegible]02 | 50.032 | 50.023 | 50.758 | 50.978 |
| $SE(\bar{\hat{m}})$ | 500 | 0 | 0.027 | 0.057 | 0.056 | 0.065 | 0.023 | 0.013 | 0.114 | 0.132 |
| $\bar{\hat{m}}$ | 1000 | 100 | 100 | 100.020 | 100.018 | 100.018 | 100 | 100 | 100.038 | 100.044 |
| $SE(\bar{\hat{m}})$ | 1000 | 0 | 0 | 0.014 | 0.013 | 0.014 | 0 | 0 | 0.025 | 0.019 |

**Table 5.9:** *MOON: Data generated= Contaminated Normal* $\mu_1 = 0,\ \sigma_1 = 1,\ \mu_2 = 0,\ \sigma_2 = 8,\ p = 0.1$. *The values for the lower bound of truncation of* $\hat{m}$, *i.e., lower bound* $= np_n$ *where* $p_n = 0.1$.

| MSE ratios for the different estimates of $m$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $n$ | **BS1** | **BS2** | **BS3** | **NB** | **BC0** | **BC1** | **BC2** | **UNBC1** | **UNBC2** |
| 20 | 1.824 | 1.545 | 1.502 | 1.566 | 1.718 | 1.834 | 1.708 | 1.603 | 1.882 |
| 50 | 1.287 | 1.177 | 1.122 | 1.138 | 1.122 | 1.280 | 1.222 | 1.147 | 1.116 |
| 100 | 1.095 | 1.067 | 1.051 | 1.047 | 1.038 | 1.082 | 1.071 | 1.042 | 1.030 |
| 500 | 0.959 | 0.966 | 0.964 | 0.952 | 0.956 | 0.949 | 0.983 | 0.960 | 0.988 |
| 1000 | 0.967 | 0.968 | 0.972 | 0.958 | 0.968 | 0.966 | 0.992 | 0.971 | 0.997 |

| Mean and standard error of chosen $m$ values | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Stat | $n$ | **BS1** | **BS2** | **BS3** | **NB** | **BC0** | **BC1** | **BC2** | **UNBC1** | **UNBC2** |
| $\bar{\hat{m}}$ | 20 | 2.273 | 3.360 | 4.448 | 4.286 | 5.272 | 2.201 | 7.675 | 2.526 | 4.296 |
| $SE(\bar{\hat{m}})$ | 20 | 0.029 | 0.064 | 0.081 | 0.086 | 0.109 | 0.024 | 0.124 | 0.026 | 0.095 |
| $\bar{\hat{m}}$ | 50 | 5.157 | 6.442 | 8.231 | 8.090 | 9.599 | 5.311 | 7.912 | 7.065 | 11.107 |
| $SE(\bar{\hat{m}})$ | 50 | 0.029 | 0.095 | 0.151 | 0.149 | 0.194 | 0.048 | 0.161 | 0.082 | 0.211 |
| $\bar{\hat{m}}$ | 100 | 10.105 | 10.960 | 12.449 | 12.180 | 12.739 | 10.333 | 18.637 | 10.924 | 13.323 |
| $SE(\bar{\hat{m}})$ | 100 | 0.036 | 0.101 | 0.171 | 0.153 | 0.182 | 0.051 | 0.108 | 0.113 | 0.171 |
| $\bar{\hat{m}}$ | 500 | 50 | 50.005 | 50.341 | 50.046 | 50.045 | 50.009 | 50.067 | 50.063 | 50.175 |
| $SE(\bar{\hat{m}})$ | 500 | 0 | 0.003 | 0.072 | 0.013 | 0.013 | 0.005 | 0.016 | 0.016 | 0.033 |
| $\bar{\hat{m}}$ | 1000 | 100 | 100 | 100.363 | 100.004 | 100.003 | 100 | 100 | 100.005 | 100.003 |
| $SE(\bar{\hat{m}})$ | 1000 | 0 | 0 | 0.135 | 0.002 | 0.002 | 0 | 0 | 0.003 | 0.003 |

**Table 5.10:** *MOON: Data generated= Contaminated Normal* $\mu_1 = 0,\ \sigma_1 = 1,\ \mu_2 = 1,\ \sigma_2 = 5,\ p = 0.3$. *The values for the lower bound of truncation of* $\hat{m}$, *i.e., lower bound* $= np_n$ *where* $p_n = 0.1$.

| MSE ratios for the different estimates of $m$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $n$ | **BS1** | **BS2** | **BS3** | **NB** | **BC0** | **BC1** | **BC2** | **UNBC1** | **UNBC2** |
| 20 | 1.336 | 1.294 | 1.285 | 1.302 | 1.376 | 1.323 | 1.276 | 1.265 | 1.343 |
| 50 | 1.079 | 1.036 | 1.069 | 1.088 | 1.054 | 1.077 | 1.032 | 1.028 | 1.019 |
| 100 | 1.082 | 1.080 | 1.072 | 1.073 | 1.072 | 1.085 | 1.011 | 1.026 | 1.004 |
| 500 | 0.983 | 0.979 | 0.981 | 0.970 | 0.970 | 0.968 | 0.964 | 0.971 | 0.969 |
| 1000 | 0.989 | 0.987 | 0.994 | 0.993 | 0.972 | 0.992 | 1.002 | 0.983 | 1.014 |

| Mean and standard error of chosen $m$ values | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Stat | $n$ | **BS1** | **BS2** | **BS3** | **NB** | **BC0** | **BC1** | **BC2** | **UNBC1** | **UNBC2** |
| $\bar{m}$ | 20 | 2.398 | 3.270 | 4.117 | 3.821 | 4.104 | 2.391 | 7.778 | 2.586 | 4.003 |
| $SE(\bar{m})$ | 20 | 0.036 | 0.064 | 0.080 | 0.078 | 0.089 | 0.037 | 0.118 | 0.038 | 0.086 |
| $\bar{m}$ | 50 | 5.576 | 6.792 | 8.001 | 8.044 | 8.634 | 5.769 | 7.443 | 6.678 | 8.442 |
| $SE(\bar{m})$ | 50 | 0.072 | 0.118 | 0.152 | 0.155 | 0.176 | 0.078 | 0.135 | 0.096 | 0.154 |
| $\bar{m}$ | 100 | 10.271 | 11.478 | 12.833 | 12.589 | 12.894 | 11.016 | 11.310 | 11.930 | 12.236 |
| $SE(\bar{m})$ | 100 | 0.051 | 0.146 | 0.198 | 0.178 | 0.195 | 0.120 | 0.119 | 0.143 | 0.152 |
| $\bar{m}$ | 500 | 50.003 | 50.304 | 51.003 | 50.126 | 50.121 | 50.065 | 50.035 | 50.108 | 50.064 |
| $SE(\bar{m})$ | 500 | 0.003 | 0.168 | 0.259 | 0.028 | 0.028 | 0.018 | 0.013 | 0.025 | 0.018 |
| $\bar{m}$ | 1000 | 100 | 100.181 | 101.230 | 100.002 | 100.002 | 100 | 100 | 100.002 | 100 |
| $SE(\bar{m})$ | 1000 | 0 | 0.091 | 0.475 | 0.002 | 0.002 | 0 | 0 | 0.002 | 0 |

**Table 5.11:** *MOON: Data generated= Contaminated Normal* $\mu_1 = 0$, $\sigma_1 = 1$, $\mu_2 = 2$, $\sigma_2 = 8$, $p = 0.5$. *The values for the lower bound of truncation of* $\hat{m}$, *i.e., lower bound* $= np_n$ *where* $p_n = 0.1$.

| MSE ratios for the different estimates of $m$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $n$ | **BS1** | **BS2** | **BS3** | **NB** | **BC0** | **BC1** | **BC2** | **UNBC1** | **UNBC2** |
| 20 | 1.185 | 1.172 | 1.140 | 1.152 | 1.146 | 1.185 | 1.056 | 1.190 | 1.088 |
| 50 | 1.079 | 1.077 | 1.078 | 1.074 | 1.084 | 1.077 | 1.037 | 1.076 | 1.036 |
| 100 | 1.074 | 1.074 | 1.076 | 1.074 | 1.085 | 1.080 | 1.028 | 1.075 | 1.029 |
| 500 | 0.998 | 0.994 | 0.991 | 0.994 | 0.991 | 0.995 | 0.993 | 0.993 | 0.999 |
| 1000 | 1.028 | 1.024 | 1.034 | 1.028 | 1.030 | 1.029 | 0.990 | 1.029 | 0.987 |

| Mean and standard error of chosen $m$ values | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Stat | $n$ | **BS1** | **BS2** | **BS3** | **NB** | **BC0** | **BC1** | **BC2** | **UNBC1** | **UNBC2** |
| $\bar{m}$ | 20 | 4.469 | 4.779 | 5.338 | 5.142 | 5.742 | 4.160 | 9.297 | 4.160 | 6.031 |
| $SE(\bar{m})$ | 20 | 0.042 | 0.051 | 0.067 | 0.064 | 0.080 | 0.025 | 0.112 | 0.024 | 0.086 |
| $\bar{m}$ | 50 | 10.336 | 10.790 | 11.188 | 11.273 | 11.402 | 10.218 | 11.319 | 10.348 | 11.546 |
| $SE(\bar{m})$ | 50 | 0.056 | 0.087 | 0.099 | 0.105 | 0.105 | 0.048 | 0.117 | 0.053 | 0.111 |
| $\bar{m}$ | 100 | 20.158 | 20.431 | 20.562 | 20.408 | 20.388 | 20.071 | 20.471 | 20.149 | 20.671 |
| $SE(\bar{m})$ | 100 | 0.057 | 0.092 | 0.099 | 0.083 | 0.080 | 0.030 | 0.106 | 0.039 | 0.111 |
| $\bar{m}$ | 500 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| $SE(\bar{m})$ | 500 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $\bar{m}$ | 1000 | 200 | 200 | 200 | 200 | 200 | 200 | 200 | 200 | 200 |
| $SE(\bar{m})$ | 1000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

**Table 5.12:** *MOON: Data generated= Double exponential* $\mu = 0$, $\sigma = \sqrt{2}$. *The values for the lower bound of truncation of* $\hat{m}$, *i.e., lower bound* $= np_n$ *where* $p_n = 0.2$.

| MSE ratios for the different estimates of $m$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $n$ | BS1 | BS2 | BS3 | NB | BC0 | BC1 | BC2 | UNBC1 | UNBC2 |
| 20 | 2.238 | 2.131 | 2.096 | 1.999 | 2.344 | 2.255 | 2.298 | 2.238 | 2.293 |
| 50 | 1.875 | 1.877 | 1.824 | 1.510 | 1.620 | 1.885 | 1.051 | 1.827 | 1.955 |
| 100 | 1.799 | 1.774 | 1.651 | 1.295 | 1.343 | 1.794 | 0.997 | 1.271 | 1.690 |
| 500 | 1.534 | 1.515 | 1.472 | 1.265 | 1.201 | 1.531 | 0.997 | 1.192 | 1.487 |
| 1000 | 1.451 | 1.457 | 1.426 | 1.328 | 1.251 | 1.460 | 0.996 | 1.240 | 1.424 |

| Mean and standard error of chosen $m$ values | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Stat | $n$ | BS1 | BS2 | BS3 | NB | BC0 | BC1 | BC2 | UNBC1 | UNBC2 |
| $\bar{m}$ | 20 | 4.558 | 5.858 | 7.992 | 8.487 | 6.582 | 4.557 | 4.325 | 4.760 | 4.264 |
| $SE(\bar{m})$ | 20 | 0.046 | 0.075 | 0.099 | 0.112 | 0.107 | 0.045 | 0.041 | 0.048 | 0.036 |
| $\bar{m}$ | 50 | 11.369 | 16.267 | 23.212 | 25.778 | 33.369 | 11.965 | 28.345 | 21.044 | 32.496 |
| $SE(\bar{m})$ | 50 | 0.111 | 0.202 | 0.245 | 0.277 | 0.321 | 0.133 | 0.340 | 0.149 | 0.338 |
| $\bar{m}$ | 100 | 22.057 | 31.880 | 47.031 | 50.807 | 71.142 | 24.339 | 51.031 | 62.791 | 66.652 |
| $SE(\bar{m})$ | 100 | 0.180 | 0.380 | 0.462 | 0.494 | 0.491 | 0.265 | 0.669 | 0.293 | 0.619 |
| $\bar{m}$ | 500 | 101.772 | 123.634 | 180.072 | 166.125 | 213.640 | 114.873 | 214.415 | 325.596 | 275.131 |
| $SE(\bar{m})$ | 500 | 0.341 | 1.067 | 1.693 | 1.146 | 1.458 | 0.682 | 3.071 | 1.987 | 2.414 |
| $\bar{m}$ | 1000 | 202.524 | 229.966 | 318.062 | 283.561 | 343.077 | 227.539 | 378.677 | 549.803 | 486.157 |
| $SE(\bar{m})$ | 1000 | 0.569 | 1.710 | 3.133 | 1.609 | 2.441 | 1.467 | 5.675 | 4.266 | 4.310 |

**Table 5.13:** *MOON: Data generated=* $F(8,5)$. *The values for the lower bound of truncation of* $\hat{m}$, *i.e., lower bound* $= np_n$ *where* $p_n = 0.2$.

| MSE ratios for the different estimates of $m$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $n$ | BS1 | BS2 | BS3 | NB | BC0 | BC1 | BC2 | UNBC1 | UNBC2 |
| 20 | 0.990 | 0.988 | 0.988 | 0.989 | 0.991 | 0.982 | 0.984 | 0.984 | 0.977 |
| 50 | 0.987 | 0.979 | 0.979 | 0.985 | 0.989 | 0.976 | 1.003 | 0.986 | 1.005 |
| 100 | 0.993 | 0.991 | 0.983 | 0.995 | 0.991 | 0.988 | 0.979 | 0.996 | 0.972 |
| 500 | 0.969 | 0.961 | 0.968 | 0.965 | 0.964 | 0.965 | 0.991 | 0.968 | 0.993 |
| 1000 | 0.978 | 0.980 | 0.982 | 0.975 | 0.976 | 0.978 | 1.009 | 0.978 | 1.012 |

| Mean and standard error of chosen $m$ values | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Stat | $n$ | BS1 | BS2 | BS3 | NB | BC0 | BC1 | BC2 | UNBC1 | UNBC2 |
| $\bar{m}$ | 20 | 4.463 | 4.711 | 4.806 | 4.709 | 4.459 | 4.304 | 7.104 | 4.358 | 4.953 |
| $SE(\bar{m})$ | 20 | 0.042 | 0.052 | 0.055 | 0.053 | 0.042 | 0.034 | 0.100 | 0.036 | 0.061 |
| $\bar{m}$ | 50 | 10.852 | 11.158 | 11.487 | 11.659 | 10.829 | 10.594 | 12.236 | 11.007 | 11.139 |
| $SE(\bar{m})$ | 50 | 0.088 | 0.104 | 0.119 | 0.128 | 0.090 | 0.075 | 0.142 | 0.096 | 0.102 |
| $\bar{m}$ | 100 | 21.829 | 21.619 | 22.120 | 22.270 | 21.193 | 21.091 | 22.042 | 21.477 | 21.472 |
| $SE(\bar{m})$ | 100 | 0.194 | 0.173 | 0.201 | 0.209 | 0.151 | 0.150 | 0.197 | 0.172 | 0.167 |
| $\bar{m}$ | 500 | 103.919 | 104.846 | 105.136 | 105.811 | 102.861 | 103.731 | 102.276 | 104.050 | 102.250 |
| $SE(\bar{m})$ | 500 | 0.640 | 0.675 | 0.723 | 0.754 | 0.539 | 0.623 | 0.455 | 0.657 | 0.455 |
| $\bar{m}$ | 1000 | 204.165 | 206.231 | 204.469 | 205.680 | 202.220 | 202.849 | 206.049 | 202.973 | 205.995 |
| $SE(\bar{m})$ | 1000 | 0.862 | 1.158 | 0.930 | 1.028 | 0.670 | 0.729 | 1.102 | 0.743 | 1.103 |

**Table 5.14:** *MOON: Data generated= Standard Normal. The values for the lower bound of truncation of* $\hat{m}$, *i.e., lower bound* $= np_n$ *where* $p_n = 0.2$.

| MSE ratios for the different estimates of $m$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $n$ | BS1 | BS2 | BS3 | NB | BC0 | BC1 | BC2 | UNBC1 | UNBC2 |
| 20 | 3.115 | 2.622 | 1.934 | 1.758 | 2.268 | 3.119 | 2.706 | 3.005 | 2.762 |
| 50 | 1.957 | 1.839 | 1.566 | 1.527 | 1.356 | 1.940 | 1.504 | 1.488 | 1.438 |
| 100 | 1.328 | 1.306 | 1.254 | 1.245 | 1.190 | 1.327 | 1.326 | 1.322 | 1.233 |
| 500 | 1.068 | 1.066 | 1.073 | 1.065 | 1.068 | 1.069 | 1.049 | 1.067 | 1.046 |
| 1000 | 1.014 | 1.009 | 1.018 | 1.014 | 1.014 | 1.013 | 0.987 | 1.017 | 0.996 |

| Mean and standard error of chosen $m$ values | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Stat | $n$ | BS1 | BS2 | BS3 | NB | BC0 | BC1 | BC2 | UNBC1 | UNBC2 |
| $\bar{m}$ | 20 | 4.137 | 4.520 | 5.536 | 6.772 | 8.107 | 4.076 | 6.546 | 4.108 | 5.134 |
| $SE(\bar{m})$ | 20 | 0.023 | 0.043 | 0.062 | 0.093 | 0.118 | 0.017 | 0.096 | 0.019 | 0.068 |
| $\bar{m}$ | 50 | 10.091 | 10.404 | 12.124 | 12.987 | 18.008 | 11.247 | 21.548 | 13.891 | 17.630 |
| $SE(\bar{m})$ | 50 | 0.030 | 0.048 | 0.110 | 0.146 | 0.245 | 0.118 | 0.311 | 0.116 | 0.256 |
| $\bar{m}$ | 100 | 20.002 | 20.321 | 21.977 | 21.906 | 24.974 | 22.393 | 32.698 | 29.520 | 31.479 |
| $SE(\bar{m})$ | 100 | 0.002 | 0.048 | 0.151 | 0.145 | 0.259 | 0.233 | 0.494 | 0.328 | 0.434 |
| $\bar{m}$ | 500 | 100 | 100 | 100.006 | 100 | 100 | 100 | 100.133 | 100.069 | 100.122 |
| $SE(\bar{m})$ | 500 | 0 | 0 | 0.006 | 0 | 0 | 0 | 0.133 | 0.025 | 0.039 |
| $\bar{m}$ | 1000 | 200 | 200 | 200 | 200 | 200 | 200 | 200 | 200 | 200 |
| $SE(\bar{m})$ | 1000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

**Table 5.15:** *MOON: Data generated= Contaminated Normal* $\mu_1 = 0,\ \sigma_1 = 1,\ \mu_2 = 0,\ \sigma_2 = 8,\ p = 0.1$. *The values for the lower bound of truncation of* $\hat{m}$, *i.e., lower bound* $= np_n$ *where* $p_n = 0.2$.

| MSE ratios for the different estimates of $m$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $n$ | BS1 | BS2 | BS3 | NB | BC0 | BC1 | BC2 | UNBC1 | UNBC2 |
| 20 | 1.385 | 1.283 | 1.214 | 1.234 | 1.267 | 1.388 | 1.353 | 1.371 | 1.404 |
| 50 | 1.126 | 1.112 | 1.084 | 1.089 | 1.063 | 1.126 | 1.153 | 1.102 | 1.118 |
| 100 | 1.098 | 1.099 | 1.092 | 1.088 | 1.083 | 1.100 | 0.993 | 1.089 | 0.990 |
| 500 | 0.997 | 0.995 | 0.997 | 0.996 | 1.001 | 0.996 | 1.045 | 0.995 | 1.037 |
| 1000 | 0.993 | 0.989 | 0.974 | 0.981 | 0.978 | 0.969 | 0.973 | 0.968 | 0.979 |

| Mean and standard error of chosen $m$ values | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Stat | $n$ | BS1 | BS2 | BS3 | NB | BC0 | BC1 | BC2 | UNBC1 | UNBC2 |
| $\bar{m}$ | 20 | 4.157 | 4.753 | 5.811 | 5.828 | 6.721 | 4.137 | 8.522 | 4.155 | 5.855 |
| $SE(\bar{m})$ | 20 | 0.024 | 0.049 | 0.074 | 0.078 | 0.098 | 0.022 | 0.112 | 0.022 | 0.084 |
| $\bar{m}$ | 50 | 10.093 | 10.651 | 11.978 | 12.013 | 13.306 | 10.172 | 12.009 | 10.838 | 13.791 |
| $SE(\bar{m})$ | 50 | 0.028 | 0.067 | 0.124 | 0.129 | 0.167 | 0.037 | 0.141 | 0.057 | 0.171 |
| $\bar{m}$ | 100 | 20.044 | 20.343 | 21.065 | 20.996 | 21.450 | 20.110 | 20.606 | 20.717 | 21.780 |
| $SE(\bar{m})$ | 100 | 0.026 | 0.068 | 0.119 | 0.116 | 0.145 | 0.031 | 0.102 | 0.072 | 0.143 |
| $\bar{m}$ | 500 | 100 | 100.013 | 100.123 | 100 | 100 | 100 | 100 | 100 | 100.002 |
| $SE(\bar{m})$ | 500 | 0 | 0.013 | 0.073 | 0 | 0 | 0 | 0 | 0 | 0.002 |
| $\bar{m}$ | 1000 | 200 | 200 | 200.001 | 200 | 200 | 200 | 200 | 200 | 200 |
| $SE(\bar{m})$ | 1000 | 0 | 0 | 0.001 | 0 | 0 | 0 | 0 | 0 | 0 |

**Table 5.16:** *MOON: Data generated= Contaminated Normal* $\mu_1 = 0,\ \sigma_1 = 1,\ \mu_2 = 1,\ \sigma_2 = 5,\ p = 0.3$. *The values for the lower bound of truncation of* $\hat{m}$, *i.e., lower bound* $= np_n$ *where* $p_n = 0.2$.

| MSE ratios for the different estimates of $m$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $n$ | **BS1** | **BS2** | **BS3** | **NB** | **BC0** | **BC1** | **BC2** | **UNBC1** | **UNBC2** |
| 20 | 1.090 | 1.063 | 1.061 | 1.057 | 1.073 | 1.098 | 1.105 | 1.094 | 1.118 |
| 50 | 1.112 | 1.104 | 1.099 | 1.091 | 1.097 | 1.103 | 0.998 | 1.096 | 0.988 |
| 100 | 1.024 | 1.023 | 1.025 | 1.026 | 1.025 | 1.028 | 0.960 | 1.029 | 0.962 |
| 500 | 0.986 | 0.985 | 0.992 | 0.990 | 0.990 | 0.990 | 0.986 | 0.987 | 0.981 |
| 1000 | 0.950 | 0.948 | 0.946 | 0.949 | 0.953 | 0.947 | 1.034 | 0.949 | 1.028 |

| Mean and standard error of chosen $m$ values | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Stat | $n$ | **BS1** | **BS2** | **BS3** | **NB** | **BC0** | **BC1** | **BC2** | **UNBC1** | **UNBC2** |
| $\bar{m}$ | 20 | 4.335 | 4.827 | 5.498 | 5.383 | 5.677 | 4.273 | 8.403 | 4.286 | 5.510 |
| $SE(\bar{m})$ | 20 | 0.036 | 0.054 | 0.071 | 0.07 | 0.078 | 0.032 | 0.106 | 0.032 | 0.075 |
| $\bar{m}$ | 50 | 10.406 | 11.095 | 12 | 12.107 | 12.428 | 10.525 | 11.782 | 10.922 | 12.467 |
| $SE(\bar{m})$ | 50 | 0.064 | 0.098 | 0.130 | 0.137 | 0.145 | 0.068 | 0.126 | 0.081 | 0.139 |
| $\bar{m}$ | 100 | 20.362 | 20.839 | 21.479 | 21.284 | 21.405 | 20.562 | 20.639 | 20.862 | 21.036 |
| $SE(\bar{m})$ | 100 | 0.086 | 0.121 | 0.155 | 0.143 | 0.146 | 0.096 | 0.100 | 0.115 | 0.113 |
| $\bar{m}$ | 500 | 100 | 100.040 | 100.380 | 100 | 100 | 100 | 100 | 100 | 100 |
| $SE(\bar{m})$ | 500 | 0 | 0.040 | 0.197 | 0 | 0 | 0 | 0 | 0 | 0 |
| $\bar{m}$ | 1000 | 200 | 200.296 | 200.556 | 200 | 200 | 200 | 200 | 200 | 200 |
| $SE(\bar{m})$ | 1000 | 0 | 0.268 | 0.242 | 0 | 0 | 0 | 0 | 0 | 0 |

**Table 5.17:** *MOON: Data generated= Contaminated Normal* $\mu_1 = 0$, $\sigma_1 = 1$, $\mu_2 = 2$, $\sigma_2 = 8$, $p = 0.5$. *The values for the lower bound of truncation of* $\hat{m}$, *i.e., lower bound* $= np_n$ *where* $p_n$ = *0.2.*

| MSE ratios for the different estimates of $m$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $n$ | **BS1** | **BS2** | **BS3** | **NB** | **BC0** | **BC1** | **BC2** | **UNBC1** | **UNBC2** |
| 20 | 1.190 | 1.189 | 1.179 | 1.179 | 1.185 | 1.191 | 1.111 | 1.192 | 1.140 |
| 50 | 1.089 | 1.088 | 1.089 | 1.089 | 1.086 | 1.088 | 1.135 | 1.086 | 1.128 |
| 100 | 1.088 | 1.087 | 1.087 | 1.086 | 1.088 | 1.089 | 1.104 | 1.088 | 1.102 |
| 500 | 1.063 | 1.002 | 1.001 | 1.000 | 0.999 | 1.001 | 0.999 | 0.999 | 1.001 |
| 1000 | 1.037 | 1.038 | 1.037 | 1.039 | 1.036 | 1.036 | 1.051 | 1.040 | 1.051 |

| Mean and standard error of chosen $m$ values | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Stat | $n$ | **BS1** | **BS2** | **BS3** | **NB** | **BC0** | **BC1** | **BC2** | **UNBC1** | **UNBC2** |
| $\bar{m}$ | 20 | 1.319 | 1.765 | 2.393 | 2.107 | 2.651 | 1.160 | 5.727 | 1.165 | 2.788 |
| $SE(\bar{m})$ | 20 | 0.040 | 0.059 | 0.079 | 0.070 | 0.087 | 0.029 | 0.127 | 0.029 | 0.092 |
| $\bar{m}$ | 50 | 1.312 | 1.800 | 2.438 | 2.305 | 2.276 | 1.163 | 2.271 | 1.344 | 2.476 |
| $SE(\bar{m})$ | 50 | 0.056 | 0.094 | 0.124 | 0.119 | 0.112 | 0.040 | 0.124 | 0.050 | 0.117 |
| $\bar{m}$ | 100 | 1.163 | 1.376 | 1.415 | 1.544 | 1.475 | 1.167 | 1.460 | 1.314 | 1.729 |
| $SE(\bar{m})$ | 100 | 0.057 | 0.093 | 0.078 | 0.106 | 0.092 | 0.067 | 0.113 | 0.072 | 0.110 |
| $\bar{m}$ | 500 | 5 | 5 | 5.001 | 5 | 5 | 5 | 5 | 5 | 5 |
| $SE(\bar{m})$ | 500 | 0 | 0 | 0.001 | 0 | 0 | 0 | 0 | 0 | 0 |
| $\bar{m}$ | 1000 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 |
| $SE(\bar{m})$ | 1000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

**Table 5.18:** *CMOON: Data generated= Double exponential* $\mu = 0$, $\sigma = \sqrt{2}$. *Based on* $\hat{m}$ *values that used* $p_n = 0.1$ *as the lower bound of truncation, i.e., lower bound* $= np_n$ *where* $p_n$ = *0.1.*

| MSE ratios for the different estimates of $m$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $n$ | BS1 | BS2 | BS3 | NB | BC0 | BC1 | BC2 | UNBC1 | UNBC2 |
| 20 | 1.308 | 1.306 | 1.406 | 1.251 | 1.311 | 1.306 | 1.816 | 1.303 | 1.810 |
| 50 | 1.229 | 1.258 | 1.387 | 1.168 | 1.157 | 1.233 | 1.229 | 1.579 | 1.366 |
| 100 | 1.208 | 1.289 | 1.39 | 1.220 | 1.170 | 1.212 | 1.267 | 1.072 | 1.369 |
| 500 | 1.318 | 1.368 | 1.406 | 1.365 | 1.307 | 1.341 | 1.085 | 1.284 | 1.158 |
| 1000 | 1.128 | 1.165 | 1.210 | 1.149 | 1.102 | 1.146 | 1.142 | 1.081 | 1.192 |

| Mean and standard error of chosen $m$ values | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Stat | $n$ | BS1 | BS2 | BS3 | NB | BC0 | BC1 | BC2 | UNBC1 | UNBC2 |
| $\bar{m}$ | 20 | 1.513 | 2.931 | 4.921 | 5.447 | 3.940 | 1.480 | 1.314 | 1.596 | 1.245 |
| $SE(\bar{m})$ | 20 | 0.050 | 0.090 | 0.119 | 0.129 | 0.124 | 0.047 | 0.044 | 0.049 | 0.038 |
| $\bar{m}$ | 50 | 2.468 | 7.484 | 13.909 | 17.670 | 28.614 | 2.845 | 22.062 | 9.913 | 27.490 |
| $SE(\bar{m})$ | 50 | 0.125 | 0.237 | 0.300 | 0.339 | 0.396 | 0.141 | 0.419 | 0.174 | 0.410 |
| $\bar{m}$ | 100 | 3.549 | 13.658 | 28.525 | 33.176 | 58.610 | 5.449 | 37.975 | 41.347 | 55.722 |
| $SE(\bar{m})$ | 100 | 0.214 | 0.418 | 0.572 | 0.606 | 0.663 | 0.296 | 0.823 | 0.380 | 0.762 |
| $\bar{m}$ | 500 | 8.476 | 34.151 | 76.233 | 60.543 | 101.845 | 18.014 | 140.185 | 235.073 | 185.458 |
| $SE(\bar{m})$ | 500 | 0.457 | 1.169 | 1.800 | 1.015 | 1.624 | 0.563 | 3.791 | 2.862 | 2.992 |
| $\bar{m}$ | 1000 | 13.371 | 55.649 | 121.374 | 88.446 | 140.445 | 39.166 | 242.736 | 371.395 | 274.131 |
| $SE(\bar{m})$ | 1000 | 0.571 | 2.170 | 3.178 | 1.574 | 2.795 | 1.638 | 7.231 | 5.907 | 5.187 |

**Table 5.19:** *CMOON: Data generated=* $F(8,5)$. *Based on* $\hat{m}$ *values that used* $p_n = 0.1$ *as the lower bound of truncation, i.e., lower bound* $= np_n$ *where* $p_n = 0.1$.

| MSE ratios for the different estimates of $m$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $n$ | BS1 | BS2 | BS3 | NB | BC0 | BC1 | BC2 | UNBC1 | UNBC2 |
| 20 | 0.942 | 0.935 | 0.935 | 0.941 | 0.941 | 0.932 | 0.941 | 0.939 | 0.935 |
| 50 | 0.980 | 0.981 | 0.980 | 0.981 | 0.982 | 0.979 | 0.993 | 0.982 | 0.988 |
| 100 | 1.043 | 1.044 | 1.044 | 1.048 | 1.047 | 1.046 | 1.011 | 1.048 | 1.008 |
| 500 | 1.010 | 1.010 | 1.011 | 1.010 | 1.009 | 1.012 | 0.986 | 1.011 | 0.983 |
| 1000 | 1.006 | 1.006 | 1.006 | 1.004 | 1.006 | 1.003 | 1.040 | 1.004 | 1.043 |

| Mean and standard error of chosen $m$ values | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Stat | $n$ | BS1 | BS2 | BS3 | NB | BC0 | BC1 | BC2 | UNBC1 | UNBC2 |
| $\bar{m}$ | 20 | 1.485 | 1.605 | 1.821 | 1.698 | 1.494 | 1.285 | 4.111 | 1.326 | 1.778 |
| $SE(\bar{m})$ | 20 | 0.049 | 0.054 | 0.062 | 0.059 | 0.050 | 0.037 | 0.114 | 0.039 | 0.061 |
| $\bar{m}$ | 50 | 1.962 | 2.267 | 2.406 | 2.603 | 1.934 | 1.581 | 3.353 | 1.906 | 2.188 |
| $SE(\bar{m})$ | 50 | 0.107 | 0.121 | 0.130 | 0.138 | 0.108 | 0.083 | 0.166 | 0.103 | 0.120 |
| $\bar{m}$ | 100 | 2.470 | 3.156 | 3.449 | 3.725 | 2.133 | 2.367 | 3.325 | 2.790 | 2.571 |
| $SE(\bar{m})$ | 100 | 0.182 | 0.226 | 0.235 | 0.252 | 0.159 | 0.182 | 0.229 | 0.205 | 0.191 |
| $\bar{m}$ | 500 | 9.920 | 8.703 | 10.305 | 10.605 | 7.670 | 8.556 | 8.995 | 8.833 | 8.917 |
| $SE(\bar{m})$ | 500 | 0.801 | 0.645 | 0.795 | 0.805 | 0.554 | 0.683 | 0.718 | 0.715 | 0.715 |
| $\bar{m}$ | 1000 | 15.331 | 15.713 | 15.669 | 17.633 | 13.485 | 16.182 | 13.697 | 16.361 | 13.647 |
| $SE(\bar{m})$ | 1000 | 1.144 | 1.185 | 1.171 | 1.352 | 0.911 | 1.295 | 0.960 | 1.315 | 0.948 |

**Table 5.20:** *CMOON: Data generated= Standard Normal. Based on* $\hat{m}$ *values that used* $p_n = 0.1$ *as the lower bound of truncation, i.e., lower bound* $= np_n$ *where* $p_n = 0.1$.

| MSE ratios for the different estimates of $m$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $n$ | **BS1** | **BS2** | **BS3** | **NB** | **BC0** | **BC1** | **BC2** | **UNBC1** | **UNBC2** |
| 20 | 1.469 | 1.45 | 1.417 | 1.383 | 1.432 | 1.463 | 1.409 | 1.469 | 1.419 |
| 50 | 1.279 | 1.309 | 1.316 | 1.291 | 1.245 | 1.276 | 1.221 | 1.409 | 1.188 |
| 100 | 1.152 | 1.180 | 1.194 | 1.185 | 1.171 | 1.153 | 1.175 | 1.184 | 1.162 |
| 500 | 1.114 | 1.116 | 1.116 | 1.114 | 1.116 | 1.116 | 1.125 | 1.116 | 1.125 |
| 1000 | 1.073 | 1.074 | 1.073 | 1.074 | 1.073 | 1.074 | 1.047 | 1.072 | 1.046 |

| Mean and standard error of chosen $m$ values | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Stat | $n$ | **BS1** | **BS2** | **BS3** | **NB** | **BC0** | **BC1** | **BC2** | **UNBC1** | **UNBC2** |
| $\bar{m}$ | 20 | 1.105 | 1.408 | 2.453 | 3.769 | 5.233 | 1.130 | 3.896 | 1.161 | 2.222 |
| $SE(\bar{m})$ | 20 | 0.021 | 0.043 | 0.074 | 0.107 | 0.134 | 0.025 | 0.115 | 0.028 | 0.078 |
| $\bar{m}$ | 50 | 1.038 | 1.402 | 3.096 | 3.953 | 9.626 | 2.280 | 14.806 | 4.324 | 9.191 |
| $SE(\bar{m})$ | 50 | 0.023 | 0.049 | 0.120 | 0.161 | 0.289 | 0.132 | 0.382 | 0.124 | 0.298 |
| $\bar{m}$ | 100 | 1.005 | 1.393 | 2.818 | 2.807 | 5.823 | 3.952 | 15.895 | 10.653 | 13.232 |
| $SE(\bar{m})$ | 100 | 0.002 | 0.047 | 0.138 | 0.138 | 0.277 | 0.286 | 0.601 | 0.365 | 0.487 |
| $\bar{m}$ | 500 | 5 | 5.013 | 5.080 | 5.071 | 5.086 | 5.009 | 5.006 | 5.230 | 5.302 |
| $SE(\bar{m})$ | 500 | 0 | 0.010 | 0.017 | 0.020 | 0.024 | 0.007 | 0.003 | 0.046 | 0.054 |
| $\bar{m}$ | 1000 | 10 | 10 | 10.005 | 10.004 | 10.004 | 10 | 10 | 10.010 | 10.010 |
| $SE(\bar{m})$ | 1000 | 0 | 0 | 0.003 | 0.003 | 0.003 | 0 | 0 | 0.007 | 0.004 |

Table 5.21: *CMOON: Data generated= Contaminated Normal* $\mu_1 = 0,\ \sigma_1 = 1,\ \mu_2 = 0,\ \sigma_2 = 8,\ p = 0.1$. *Based on* $\hat{m}$ *values that used* $p_n = 0.1$ *as the lower bound of truncation, i.e., lower bound* $= np_n$ *where* $p_n = 0.1$.

| MSE ratios for the different estimates of $m$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $n$ | **BS1** | **BS2** | **BS3** | **NB** | **BC0** | **BC1** | **BC2** | **UNBC1** | **UNBC2** |
| 20 | 1.252 | 1.216 | 1.213 | 1.221 | 1.231 | 1.247 | 1.296 | 1.249 | 1.335 |
| 50 | 1.145 | 1.125 | 1.105 | 1.102 | 1.098 | 1.142 | 1.143 | 1.128 | 1.114 |
| 100 | 1.078 | 1.070 | 1.064 | 1.064 | 1.062 | 1.077 | 1.071 | 1.067 | 1.057 |
| 500 | 0.989 | 0.987 | 0.984 | 0.983 | 0.982 | 0.981 | 1.013 | 0.981 | 1.010 |
| 1000 | 0.986 | 0.986 | 0.985 | 0.987 | 0.986 | 0.984 | 0.969 | 0.984 | 0.969 |

| Mean and standard error of chosen $m$ values | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Stat | $n$ | **BS1** | **BS2** | **BS3** | **NB** | **BC0** | **BC1** | **BC2** | **UNBC1** | **UNBC2** |
| $\bar{m}$ | 20 | 1.149 | 1.802 | 2.475 | 2.585 | 3.741 | 1.096 | 5.610 | 1.113 | 2.912 |
| $SE(\bar{m})$ | 20 | 0.026 | 0.061 | 0.078 | 0.083 | 0.111 | 0.020 | 0.130 | 0.020 | 0.096 |
| $\bar{m}$ | 50 | 1.073 | 1.791 | 3.081 | 3.002 | 4.455 | 1.185 | 3.190 | 1.749 | 5.440 |
| $SE(\bar{m})$ | 50 | 0.022 | 0.085 | 0.145 | 0.140 | 0.191 | 0.045 | 0.161 | 0.057 | 0.208 |
| $\bar{m}$ | 100 | 1.060 | 1.509 | 2.429 | 2.185 | 2.611 | 1.144 | 1.545 | 1.842 | 2.655 |
| $SE(\bar{m})$ | 100 | 0.034 | 0.087 | 0.154 | 0.131 | 0.159 | 0.033 | 0.098 | 0.070 | 0.128 |
| $\bar{m}$ | 500 | 5 | 5.001 | 5.100 | 5.010 | 5.010 | 5.002 | 5.015 | 5.014 | 5.041 |
| $SE(\bar{m})$ | 500 | 0 | 0.001 | 0.024 | 0.003 | 0.003 | 0.001 | 0.004 | 0.004 | 0.008 |
| $\bar{m}$ | 1000 | 10 | 10 | 10.127 | 10.001 | 10.001 | 10 | 10 | 10.001 | 10.001 |
| $SE(\bar{m})$ | 1000 | 0 | 0 | 0.055 | 0.001 | 0 | 0 | 0 | 0.001 | 0.001 |

Table 5.22: *CMOON: Data generated= Contaminated Normal* $\mu_1 = 0,\ \sigma_1 = 1,\ \mu_2 = 1,\ \sigma_2 = 5,\ p = 0.3$. *Based on* $\hat{m}$ *values that used* $p_n = 0.1$ *as the lower bound of truncation, i.e., lower bound* $= np_n$ *where* $p_n = 0.1$.

| MSE ratios for the different estimates of $m$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $n$ | **BS1** | **BS2** | **BS3** | **NB** | **BC0** | **BC1** | **BC2** | **UNBC1** | **UNBC2** |
| 20 | 1.305 | 1.288 | 1.269 | 1.275 | 1.304 | 1.291 | 1.227 | 1.282 | 1.284 |
| 50 | 1.164 | 1.148 | 1.129 | 1.135 | 1.132 | 1.136 | 1.118 | 1.129 | 1.109 |
| 100 | 1.177 | 1.155 | 1.139 | 1.140 | 1.139 | 1.160 | 1.075 | 1.143 | 1.063 |
| 500 | 0.978 | 0.977 | 0.979 | 0.980 | 0.980 | 0.975 | 0.972 | 0.977 | 0.970 |
| 1000 | 1.001 | 0.997 | 1.001 | 0.997 | 1.000 | 1.000 | 1.026 | 1.000 | 1.022 |

| Mean and standard error of chosen $m$ values | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Stat | $n$ | **BS1** | **BS2** | **BS3** | **NB** | **BC0** | **BC1** | **BC2** | **UNBC1** | **UNBC2** |
| $\bar{m}$ | 20 | 1.228 | 1.788 | 2.366 | 2.267 | 2.657 | 1.250 | 5.401 | 1.262 | 2.540 |
| $SE(\bar{m})$ | 20 | 0.032 | 0.062 | 0.076 | 0.076 | 0.088 | 0.035 | 0.125 | 0.035 | 0.085 |
| $\bar{m}$ | 50 | 1.411 | 2.170 | 3.037 | 3.112 | 3.725 | 1.499 | 2.594 | 1.841 | 3.219 |
| $SE(\bar{m})$ | 50 | 0.070 | 0.112 | 0.147 | 0.150 | 0.174 | 0.074 | 0.128 | 0.085 | 0.143 |
| $\bar{m}$ | 100 | 1.134 | 1.956 | 2.826 | 2.535 | 2.806 | 1.643 | 1.705 | 2.039 | 2.190 |
| $SE(\bar{m})$ | 100 | 0.040 | 0.138 | 0.184 | 0.161 | 0.177 | 0.113 | 0.102 | 0.125 | 0.127 |
| $\bar{m}$ | 500 | 5.001 | 5.231 | 5.606 | 5.030 | 5.02. | 5.015 | 5.008 | 5.026 | 5.015 |
| $SE(\bar{m})$ | 500 | 0.001 | 0.171 | 0.246 | 0.007 | 0.007 | 0.004 | 0.003 | 0.006 | 0.004 |
| $\bar{m}$ | 1000 | 10 | 10.061 | 10.924 | 10 | 10 | 10 | 10 | 10 | 10 |
| $SE(\bar{m})$ | 1000 | 0 | 0.034 | 0.461 | 0 | 0 | 0 | 0 | 0 | 0 |

**Table 5.23:** *CMOON: Data generated= Contaminated Normal* $\mu_1 = 0,\ \sigma_1 = 1,\ \mu_2 = 2,\ \sigma_2 = 8,\ p = 0.5$. *Based on* $\hat{m}$ *values that used* $p_n = 0.1$ *as the lower bound of truncation, i.e., lower bound* $= np_n$ *where* $p_n = 0.1$.

| MSE ratios for the different estimates of $m$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $n$ | **BS1** | **BS2** | **BS3** | **NB** | **BC0** | **BC1** | **BC2** | **UNBC1** | **UNBC2** |
| 20 | 1.190 | 1.189 | 1.179 | 1.179 | 1.185 | 1.191 | 1.076 | 1.192 | 1.104 |
| 50 | 1.089 | 1.083 | 1.079 | 1.080 | 1.076 | 1.088 | 1.059 | 1.086 | 1.053 |
| 100 | 1.088 | 1.087 | 1.087 | 1.086 | 1.088 | 1.089 | 1.039 | 1.088 | 1.039 |
| 500 | 1.043 | 1.042 | 1.041 | 1.040 | 1.040 | 1.041 | 1.001 | 0.999 | 1.002 |
| 1000 | 1.037 | 1.038 | 1.037 | 1.039 | 1.036 | 1.036 | 0.991 | 1.040 | 0.991 |

| Mean and standard error of chosen $m$ values | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Stat | $n$ | **BS1** | **BS2** | **BS3** | **NB** | **BC0** | **BC1** | **BC2** | **UNBC1** | **UNBC2** |
| $\bar{m}$ | 20 | 1.443 | 1.703 | 2.244 | 2.096 | 2.761 | 1.153 | 6.277 | 1.143 | 3.086 |
| $SE(\bar{m})$ | 20 | 0.046 | 0.056 | 0.074 | 0.071 | 0.090 | 0.027 | 0.132 | 0.026 | 0.098 |
| $\bar{m}$ | 50 | 2.323 | 2.781 | 3.089 | 3.207 | 3.266 | 2.227 | 3.388 | 2.309 | 3.404 |
| $SE(\bar{m})$ | 50 | 0.061 | 0.098 | 0.108 | 0.117 | 0.115 | 0.055 | 0.136 | 0.059 | 0.121 |
| $\bar{m}$ | 100 | 4.162 | 4.429 | 4.524 | 4.369 | 4.350 | 4.055 | 4.525 | 4.106 | 4.646 |
| $SE(\bar{m})$ | 100 | 0.066 | 0.103 | 0.108 | 0.090 | 0.088 | 0.028 | 0.124 | 0.038 | 0.126 |
| $\bar{m}$ | 500 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 |
| $SE(\bar{m})$ | 500 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $\bar{m}$ | 1000 | 40 | 40 | 40 | 40 | 40 | 40 | 40 | 40 | 40 |
| $SE(\bar{m})$ | 1000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

**Table 5.24:** *CMOON: Data generated= Double exponential* $\mu = 0,\ \sigma = \sqrt{2}$. *Based on* $\hat{m}$ *values that used* $p_n = 0.2$ *as the lower bound of truncation, i.e., lower bound* $= np_n$ *where* $p_n = 0.2$.

| MSE ratios for the different estimates of $m$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $n$ | BS1 | BS2 | BS3 | NB | BC0 | BC1 | BC2 | UNBC1 | UNBC2 |
| 20 | 1.569 | 1.554 | 1.586 | 1.465 | 1.570 | 1.572 | 1.495 | 1.573 | 1.496 |
| 50 | 1.344 | 1.348 | 1.433 | 1.261 | 1.262 | 1.351 | 1.052 | 1.593 | 1.290 |
| 100 | 1.310 | 1.313 | 1.375 | 1.208 | 1.175 | 1.312 | 0.996 | 1.047 | 1.170 |
| 500 | 1.316 | 1.317 | 1.337 | 1.263 | 1.212 | 1.322 | 0.996 | 1.196 | 1.138 |
| 1000 | 1.322 | 1.322 | 1.335 | 1.316 | 1.264 | 1.326 | 0.991 | 1.243 | 1.107 |

| Mean and standard error of chosen $m$ values | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Stat | $n$ | BS1 | BS2 | BS3 | NB | BC0 | BC1 | BC2 | UNBC1 | UNBC2 |
| $\bar{m}$ | 20 | 1.533 | 2.688 | 4.748 | 5.601 | 4.035 | 1.529 | 1.380 | 1.642 | 1.297 |
| $SE(\bar{m})$ | 20 | 0.051 | 0.084 | 0.116 | 0.131 | 0.126 | 0.050 | 0.048 | 0.052 | 0.041 |
| $\bar{m}$ | 50 | 3.329 | 7.747 | 14.376 | 17.908 | 28.470 | 3.929 | 22.992 | 10.181 | 27.965 |
| $SE(\bar{m})$ | 50 | 0.125 | 0.229 | 0.296 | 0.338 | 0.391 | 0.149 | 0.411 | 0.184 | 0.404 |
| $\bar{m}$ | 100 | 5.835 | 14.488 | 28.531 | 33.147 | 57.841 | 8.030 | 39.481 | 42.005 | 55.915 |
| $SE(\bar{m})$ | 100 | 0.196 | 0.426 | 0.564 | 0.608 | 0.653 | 0.290 | 0.808 | 0.391 | 0.739 |
| $\bar{m}$ | 500 | 21.413 | 37.403 | 82.043 | 63.078 | 104.035 | 29.186 | 148.530 | 235.703 | 186.367 |
| $SE(\bar{m})$ | 500 | 0.359 | 1.088 | 1.917 | 1.071 | 1.661 | 0.536 | 3.726 | 2.864 | 2.879 |
| $\bar{m}$ | 1000 | 41.989 | 61.655 | 130.612 | 88.172 | 135.583 | 58.232 | 239.998 | 356.878 | 292.074 |
| $SE(\bar{m})$ | 1000 | 0.601 | 1.722 | 3.427 | 1.381 | 2.666 | 1.511 | 6.874 | 5.795 | 5.245 |

**Table 5.25:** *CMOON: Data generated=* $F(8,5)$. *Based on* $\hat{m}$ *values that used* $p_n = 0.2$ *as the lower bound of truncation, i.e., lower bound* $= np_n$ *where* $p_n = 0.2$.

| MSE ratios for the different estimates of $m$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $n$ | BS1 | BS2 | BS3 | NB | BC0 | BC1 | BC2 | UNBC1 | UNBC2 |
| 20 | 0.975 | 0.972 | 0.976 | 0.974 | 0.975 | 0.978 | 0.971 | 0.980 | 0.973 |
| 50 | 0.983 | 0.981 | 0.982 | 0.984 | 0.984 | 0.981 | 1.010 | 0.983 | 1.010 |
| 100 | 1.000 | 0.997 | 0.997 | 0.996 | 0.995 | 0.996 | 0.978 | 0.997 | 0.981 |
| 500 | 0.973 | 0.972 | 0.975 | 0.971 | 0.971 | 0.973 | 1.003 | 0.974 | 1.000 |
| 1000 | 0.981 | 0.984 | 0.986 | 0.982 | 0.981 | 0.980 | 1.018 | 0.982 | 1.015 |

| Mean and standard error of chosen $m$ values | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Stat | $n$ | BS1 | BS2 | BS3 | NB | BC0 | BC1 | BC2 | UNBC1 | UNBC2 |
| $\bar{m}$ | 20 | 1.441 | 1.690 | 1.793 | 1.715 | 1.450 | 1.294 | 4.141 | 1.335 | 1.969 |
| $SE(\bar{m})$ | 20 | 0.046 | 0.058 | 0.062 | 0.060 | 0.047 | 0.038 | 0.114 | 0.040 | 0.069 |
| $\bar{m}$ | 50 | 2.822 | 3.137 | 3.488 | 3.706 | 2.830 | 2.581 | 4.201 | 2.979 | 3.109 |
| $SE(\bar{m})$ | 50 | 0.097 | 0.116 | 0.134 | 0.146 | 0.101 | 0.083 | 0.159 | 0.107 | 0.114 |
| $\bar{m}$ | 100 | 5.890 | 5.568 | 6.101 | 6.268 | 5.175 | 5.124 | 6.022 | 5.501 | 5.446 |
| $SE(\bar{m})$ | 100 | 0.220 | 0.190 | 0.225 | 0.235 | 0.168 | 0.170 | 0.221 | 0.196 | 0.187 |
| $\bar{m}$ | 500 | 24.058 | 24.723 | 25.245 | 25.804 | 22.906 | 23.848 | 22.163 | 24.239 | 22.150 |
| $SE(\bar{m})$ | 500 | 0.725 | 0.751 | 0.821 | 0.854 | 0.613 | 0.701 | 0.498 | 0.744 | 0.499 |
| $\bar{m}$ | 1000 | 43.913 | 46.556 | 44.404 | 45.476 | 42.241 | 42.742 | 46.101 | 42.853 | 46.082 |
| $SE(\bar{m})$ | 1000 | 0.950 | 1.326 | 1.044 | 1.146 | 0.772 | 0.787 | 1.237 | 0.803 | 1.241 |

**Table 5.26:** *CMOON: Data generated= Standard Normal. Based on* $\hat{m}$ *values that used* $p_n = 0.2$ *as the lower bound of truncation, i.e., lower bound* $= np_n$ *where* $p_n = 0.2$.

| MSE ratios for the different estimates of $m$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $n$ | BS1 | BS2 | BS3 | NB | BC0 | BC1 | BC2 | UNBC1 | UNBC2 |
| 20 | 1.439 | 1.427 | 1.407 | 1.368 | 1.393 | 1.445 | 1.367 | 1.440 | 1.372 |
| 50 | 1.384 | 1.371 | 1.360 | 1.333 | 1.281 | 1.382 | 1.306 | 1.434 | 1.246 |
| 100 | 1.234 | 1.222 | 1.206 | 1.198 | 1.175 | 1.231 | 1.272 | 1.176 | 1.209 |
| 500 | 1.092 | 1.091 | 1.091 | 1.094 | 1.094 | 1.09 | 1.072 | 1.092 | 1.074 |
| 1000 | 1.022 | 1.025 | 1.024 | 1.025 | 1.025 | 1.024 | 1.002 | 1.024 | 1.003 |

| Mean and standard error of chosen $m$ values | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Stat | $n$ | BS1 | BS2 | BS3 | NB | BC0 | BC1 | BC2 | UNBC1 | UNBC2 |
| $\bar{m}$ | 20 | 1.128 | 1.480 | 2.229 | 3.719 | 5.492 | 1.072 | 3.687 | 1.089 | 2.184 |
| $SE(\bar{m})$ | 20 | 0.025 | 0.048 | 0.068 | 0.106 | 0.138 | 0.019 | 0.110 | 0.020 | 0.077 |
| $\bar{m}$ | 50 | 2.092 | 2.302 | 3.659 | 4.653 | 10.084 | 3.367 | 15.093 | 4.667 | 10.149 |
| $SE(\bar{m})$ | 50 | 0.034 | 0.046 | 0.113 | 0.157 | 0.281 | 0.137 | 0.371 | 0.116 | 0.298 |
| $\bar{m}$ | 100 | 4.001 | 4.200 | 5.512 | 5.430 | 8.246 | 6.647 | 18.009 | 11.937 | 15.572 |
| $SE(\bar{m})$ | 100 | 0.001 | 0.039 | 0.153 | 0.139 | 0.269 | 0.272 | 0.577 | 0.339 | 0.492 |
| $\bar{m}$ | 500 | 20 | 20 | 20.003 | 20 | 20 | 20 | 20.160 | 20.031 | 20.058 |
| $SE(\bar{m})$ | 500 | 0 | 0 | 0.003 | 0 | 0 | 0 | 0.160 | 0.011 | 0.019 |
| $\bar{m}$ | 1000 | 40 | 40 | 40 | 40 | 40 | 40 | 40 | 40 | 40 |
| $SE(\bar{m})$ | 1000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

**Table 5.27:** *CMOON: Data generated= Contaminated Normal* $\mu_1 = 0,\ \sigma_1 = 1,\ \mu_2 = 0,\ \sigma_2 = 8,\ p = 0.1$. *Based on* $\hat{m}$ *values that used* $p_n = 0.2$ *as the lower bound of truncation, i.e., lower bound* $= np_n$ *where* $p_n = 0.2$.

| MSE ratios for the different estimates of $m$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $n$ | BS1 | BS2 | BS3 | NB | BC0 | BC1 | BC2 | UNBC1 | UNBC2 |
| 20 | 1.269 | 1.244 | 1.221 | 1.227 | 1.235 | 1.257 | 1.269 | 1.257 | 1.295 |
| 50 | 1.181 | 1.166 | 1.135 | 1.134 | 1.122 | 1.179 | 1.198 | 1.159 | 1.169 |
| 100 | 1.145 | 1.137 | 1.129 | 1.129 | 1.124 | 1.142 | 1.032 | 1.131 | 1.023 |
| 500 | 1.002 | 1.004 | 1.001 | 1.002 | 1.003 | 1.006 | 1.053 | 1.003 | 1.050 |
| 1000 | 0.976 | 0.979 | 0.978 | 0.974 | 0.977 | 0.975 | 0.981 | 0.974 | 0.981 |

| Mean and standard error of chosen $m$ values | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Stat | $n$ | BS1 | BS2 | BS3 | NB | BC0 | BC1 | BC2 | UNBC1 | UNBC2 |
| $\bar{m}$ | 20 | 1.148 | 1.659 | 2.656 | 2.765 | 3.855 | 1.127 | 5.606 | 1.129 | 2.932 |
| $SE(\bar{m})$ | 20 | 0.027 | 0.053 | 0.082 | 0.088 | 0.113 | 0.024 | 0.131 | 0.024 | 0.096 |
| $\bar{m}$ | 50 | 2.083 | 2.535 | 3.794 | 3.878 | 5.214 | 2.150 | 4.071 | 2.542 | 5.567 |
| $SE(\bar{m})$ | 50 | 0.029 | 0.069 | 0.135 | 0.141 | 0.187 | 0.038 | 0.161 | 0.053 | 0.188 |
| $\bar{m}$ | 100 | 4.038 | 4.277 | 4.865 | 4.811 | 5.233 | 4.073 | 4.558 | 4.449 | 5.353 |
| $SE(\bar{m})$ | 100 | 0.025 | 0.071 | 0.122 | 0.120 | 0.154 | 0.026 | 0.111 | 0.057 | 0.139 |
| $\bar{m}$ | 500 | 20 | 20.006 | 20.081 | 20 | 20 | 20 | 20 | 20 | 20.001 |
| $SE(\bar{m})$ | 500 | 0 | 0.006 | 0.055 | 0 | 0 | 0 | 0 | 0 | 0.001 |
| $\bar{m}$ | 1000 | 40 | 40 | 40.001 | 40 | 40 | 40 | 40 | 40 | 40 |
| $SE(\bar{m})$ | 1000 | 0 | 0 | 0.001 | 0 | 0 | 0 | 0 | 0 | 0 |

**Table 5.28:** *CMOON: Data generated= Contaminated Normal* $\mu_1 = 0,\ \sigma_1 = 1,\ \mu_2 = 1,\ \sigma_2 = 5,\ p = 0.3$. *Based on* $\hat{m}$ *values that used* $p_n = 0.2$ *as the lower bound of truncation, i.e., lower bound* $= np_n$ *where* $p_n = 0.2$.

| MSE ratios for the different estimates of $m$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $n$ | BS1 | BS2 | BS3 | NB | BC0 | BC1 | BC2 | UNBC1 | UNBC2 |
| 20 | 1.243 | 1.224 | 1.214 | 1.212 | 1.240 | 1.241 | 1.208 | 1.236 | 1.269 |
| 50 | 1.202 | 1.185 | 1.166 | 1.163 | 1.169 | 1.194 | 1.061 | 1.171 | 1.051 |
| 100 | 1.042 | 1.036 | 1.034 | 1.033 | 1.031 | 1.039 | 0.97 | 1.036 | 0.968 |
| 500 | 0.995 | 0.993 | 0.994 | 0.995 | 0.994 | 0.995 | 0.988 | 0.996 | 0.986 |
| 1000 | 0.955 | 0.953 | 0.952 | 0.955 | 0.956 | 0.954 | 1.036 | 0.952 | 1.038 |

| Mean and standard error of chosen $m$ values | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Stat | $n$ | BS1 | BS2 | BS3 | NB | BC0 | BC1 | BC2 | UNBC1 | UNBC2 |
| $\bar{m}$ | 20 | 1.329 | 1.774 | 2.418 | 2.348 | 2.677 | 1.259 | 5.283 | 1.258 | 2.526 |
| $SE(\bar{m})$ | 20 | 0.041 | 0.060 | 0.079 | 0.079 | 0.088 | 0.035 | 0.124 | 0.034 | 0.085 |
| $\bar{m}$ | 50 | 2.414 | 3.042 | 3.892 | 4.050 | 4.351 | 2.493 | 3.728 | 2.779 | 4.269 |
| $SE(\bar{m})$ | 50 | 0.073 | 0.108 | 0.144 | 0.154 | 0.162 | 0.074 | 0.141 | 0.086 | 0.152 |
| $\bar{m}$ | 100 | 4.369 | 4.779 | 5.331 | 5.141 | 5.223 | 4.505 | 4.557 | 4.748 | 4.809 |
| $SE(\bar{m})$ | 100 | 0.097 | 0.131 | 0.169 | 0.155 | 0.156 | 0.102 | 0.110 | 0.122 | 0.118 |
| $\bar{m}$ | 500 | 20 | 20.026 | 20.385 | 20 | 20 | 20 | 20 | 20 | 20 |
| $SE(\bar{m})$ | 500 | 0 | 0.026 | 0.229 | 0 | 0 | 0 | 0 | 0 | 0 |
| $\bar{m}$ | 1000 | 40 | 40.335 | 40.399 | 40 | 40 | 40 | 40 | 40 | 40 |
| $SE(\bar{m})$ | 1000 | 0 | 0.320 | 0.195 | 0 | 0 | 0 | 0 | 0 | 0 |

**Table 5.29:** *CMOON: Data generated= Contaminated Normal* $\mu_1 = 0,\ \sigma_1 = 1,\ \mu_2 = 2,\ \sigma_2 = 8,\ p = 0.5$. *Based on* $\bar{m}$ *values that used* $p_n = 0.2$ *as the lower bound of truncation, i.e., lower bound* $= np_n$ *where* $p_n = 0.2$

### 5.7.1 Conclusions drawn from the tables

The conclusions drawn from the tables will be broken down into two sections; the first section deals with the performance of the MOON estimators, and the second section deals with the performance of the CMOON estimators. We make conclusions concerning the classes of distributions used and comment on the overall performance of the estimators. An indication of which ones performed the best will also be given.

⋆ When we consider the statistic $\zeta(\tilde{\theta}_{brag,1})$ for the MOON estimators in Tables 5.6 to 5.17 we find that the performance of the estimators is very good (i.e., producing values larger than 1) for small to moderate sample sizes, and that it converges to 1 as the sample size becomes larger. However, the tendency differs for the various distributions considered. When we categorise the distributions according to general properties we find the following:

  - **Symmetric, heavy-tailed distributions:** The distributions that are symmetric and heavy-tailed include the double exponential distribution and the unimodal contaminated normal distribution (displayed in Tables 5.6, 5.12, 5.9 and 5.15). These distributions show massive improvements over the sample mean when estimating the population mean for small sample sizes (in Table 5.9, we find that the maximum recorded improvement is 362% for small sample sizes!). It is clear from these tables that the choice of a smaller truncation lower bound produces better results for small to moderate sample sizes.

    For the double exponential distribution in Tables 5.6 and 5.12 one would be hard pressed to choose which estimators perform the best (although in Table 5.12 we see that **BC2** and **UNBC2** have rather poorer performances than the other estimators). When we consider the symmetric contaminated normal distribution in Tables 5.9 and 5.15 we find that **BS1**, **BC1** and **UNBC2** have very good performance relative to the other estimators.

- **Heavily skewed, heavy-tailed distributions:** The only distribution we considered that is both heavily skewed and heavy-tailed is the $F$ distribution (found in Tables 5.7 and 5.13). The estimators applied to this distribution improved on the performance of the sample mean by up to 134% for small to moderate sample sizes. For this distribution we obtain better performance by choosing a larger truncation lower bound.

  Deciding which estimator performs best in Table 5.13 is rather difficult since they mostly all have the same performance (except **NB** which performs slightly less well). In Table 5.7, on the other hand, three estimators stand out as being better than the rest, viz., **BS2**, **BS3** and **UNBC1**.

- **Bimodal, heavy-tailed distributions:** The distributions we considered that are bimodal and heavy-tailed include the contaminated normal distribution with $\mu_1 \neq \mu_2$ (found in Tables 5.10, 5.16, 5.11, and 5.17). For these distributions we obtain results which exceeded the performance of the sample mean by more than 80% in some cases (see for example the $n = 20$ row of Table 5.10). In this instance we obtained better results with the smaller truncation lower bound, i.e., $p_n = 0.1$.

  For $p_n = 0.2$ (i.e., Tables 5.16 and 5.17) all of the estimators work roughly equally well for these distributions. For $p_n = 0.1$ in Table 5.11 we also find that the estimators have equal performance, but in Table 5.10 we find that **BS1**, **BC1** and **UNBC2** all have slightly better performance than the others.

- **Standard normal distribution:** The results for the standard normal distribution are displayed in Tables 5.8 and 5.14. Since the sample mean is admissible as an estimator for the mean of a normal distribution we can not expect to see any improvement over the sample mean in this case. However, the estimators using the various choices of $m$ for this distribution routinely achieve at least 91% of the performance of the sample mean when the truncation lower bound is taken as $p_n = 0.1$, and at least 96% when the truncation lower bound is taken as $p_n = 0.2$. We can also see that the estimators improve as the sample size increases.

  As one might expect, all of the estimators perform equally well for this distribution (for $p_n = 0.1$ and $p_n = 0.2$).

In general, we see a marked improvement over the sample mean when the underlying data has heavier tails. This improvement is most acute when working with small to moderate sample sizes, but is less impressive when one has larger samples (except in the case of the $F$ distribution where we still see improvements of up to 40% for samples as large as 1000).

In most cases the data-dependent choices of $m$ are very close to the truncation lower bound (i.e., $np_n$ where $p_n = 0.1$ or $p_n = 0.2$) as the sample size becomes larger.

⋆ When we consider the statistic $\zeta(\tilde{\theta}_{brag,2})$ for the CMOON estimators in Tables 5.18 to 5.29 we can make similar conclusions to those made concerning the MOON estimators. Once again the estimators are found to perform very well for small to moderate sample sizes, but the results are not as impressive as those observed in the MOON estimators' results. The tendency of the estimators' performance is that, as the sample size increases, the performance starts to match the performance of the sample mean. When we categorise the distributions according to general properties we find that we can make the following conclusions:

- **Symmetric, heavy-tailed distributions:** As before, the distributions that are symmetric and heavy-tailed include the double exponential distribution and the unimodal contaminated normal distribution (found in Tables 5.18, 5.24, 5.21 and 5.27). The tables for these distributions show nearly equivalent behaviour of the performance of the estimators for both choices of the truncation lower bound. For the contaminated normal distribution we obtain performance up to 46% better than the sample mean, and for the double exponential we get an improvement of up to 19%.

All of the estimators seem to perform equally well for these distributions (for $p_n = 0.1$ and $p_n = 0.2$).

- **Heavily skewed, heavy-tailed distributions:** The $F$ distribution (found in Tables 5.19 and 5.25), also shows a significant increase in performance over the sample mean. However, the improvement in this case was not restricted to small sample sizes, because we can clearly see that even large sample sizes display an improvement between 8% and 21%, depending on the statistic when $p_n = 0.1$. For $p_n = 0.2$ we find that the performance of the estimators is better still.

  This distribution is an exception among the CMOON estimators, because it is the only one where different estimators have significantly different performance. We see in Table 5.19 that the estimator **BC2** and **UNBC2** have a slight advantage over the other estimators for small sample sizes. However, in Table 5.25, we find that no estimators stand out among the rest.

- **Bimodal, heavy-tailed distributions:** Bimodal and heavy-tailed distributions (found in Tables 5.22, 5.28, 5.23, and 5.29) also show that the estimators perform better than the sample mean when estimating the population mean (up to 30% or more in some cases). The choice of the truncation lower bound $p_n$ does not seem to significantly affect the results for these distributions.

  The estimators all perform almost equally well.

- **Standard normal distribution:** The results related to the standard normal distribution are displayed in Tables 5.8 and 5.14. Again, we do not expect to see any improvement over the sample mean for the standard normal distribution. Most results were found to be within 7% of the performance of the population mean (we obtained slightly better results with a larger truncation lower bound). An interesting result was that the convergence of the estimators' performance to the performance of the sample mean happened very quickly (i.e., for smaller samples than other distributions).

  As before, all of the estimators perform equally well for this distribution (for $p_n = 0.1$ and $p_n = 0.2$).

In general, we see that these estimators perform not nearly as well as the MOON estimators, but the choice of truncation lower bound does not seem to be as critical as it is for the MOON estimators.

The choice of which estimator performs best in the CMOON context is much more difficult because it would appear that they all perform equally well across the distributions. However, this may be attributed to the fact that most of them choose very similar resample sizes.

**Remarks:**

1. *It is difficult to choose a single estimator which performs well in all of the situations considered, but, if a choice has to be made, then it would be fair to choose the following estimators from the three classes of estimators:*

   ⋆ *For the* ***BS*** *class of estimators one would prefer to use the* ***BS1*** *estimator.*

* ⋆ *For the* ***BC*** *class of estimators we can see that the* ***BC1*** *estimator performs well for all of the situations considered.*
* ⋆ *For the* ***UNBC*** *class of estimators one would most likely choose to use* ***UNBC2*** *based on the results provided.*

*The* ***NB*** *estimator, while not bad, seems to have the worst performance relative to the other estimators. This is not surprising since it was based on a naive estimation of products of population moments.*

2. *The relatively small resample sizes obtained for the MOON and CMOON estimators are similar to the findings of Chung and Lee (2001) (recall that the CMOON estimator for $m$ is closely related to the estimator proposed by Chung and Lee (2001) for the estimation of the resample size used to determine a bootstrap confidence bound). If one calculates the estimated resample size for the confidence bound that they propose, one can easily see that it chooses very small $m$ values when we work with the mean. See the example on page 231 of Chung and Lee (2001) where they determine a 95% upper confidence bound for the population mean using the formula given on page 229. Unfortunately, this estimator for the resample size is shown to be independent of the sample data. The resample size is then given by*

$$\hat{m} = \left\lfloor n\left(\frac{1-z_\alpha^2}{1+2z_\alpha^2}\right)^2\right\rfloor = \left\lfloor n\left(\frac{1-1.645^2}{1+2\cdot 1.645^2}\right)^2\right\rfloor = \lfloor 0.07n\rfloor .$$

*This choice, while much smaller than the choices given in the above tables, indicates that one might expect to find very small values of $m$ when working with the mean.*

3. *At the request of a referee, the computational time of the m-out-of-n bootstrap procedures required to calculate the estimator are given in Table 5.30. The table also shows the time taken to calculate the choice of the data-dependent resample sizes.*

| $n$ | | $m=n$ | BS1 | BS2 | BS3 | NB | BC0 | BC1 | BC2 | UNBC1 | UNBC2 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 20 | CPU time for $\hat{m}_{1,X}$ | <0.01 | 0.25 | 0.26 | 0.27 | <0.01 | <0.01 | 0.01 | 0.01 | <0.01 | 0.01 |
| | CPU time for $\tilde{\theta}_{brag,1}$ | 0.05 | 0.04 | 0.04 | 0.05 | 0.04 | 0.04 | 0.04 | 0.05 | 0.04 | 0.04 |
| 50 | CPU time for $\hat{m}_{1,X}$ | <0.01 | 0.27 | 0.28 | 0.29 | <0.01 | <0.01 | <0.01 | 0.02 | <0.01 | 0.01 |
| | CPU time for $\tilde{\theta}_{brag,1}$ | 0.05 | 0.04 | 0.04 | 0.05 | 0.06 | 0.05 | 0.04 | 0.04 | 0.04 | 0.05 |
| 100 | CPU time for $\hat{m}_{1,X}$ | <0.01 | 0.31 | 0.31 | 0.33 | <0.01 | <0.01 | <0.01 | 0.02 | <0.01 | 0.01 |
| | CPU time for $\tilde{\theta}_{brag,1}$ | 0.07 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.04 | 0.05 | 0.05 |
| 1000 | CPU time for $\hat{m}_{1,X}$ | <0.01 | 0.42 | 0.59 | 0.75 | 0.01 | <0.01 | 0.01 | 0.05 | 0.01 | 0.03 |
| | CPU time for $\tilde{\theta}_{brag,1}$ | 0.26 | 0.07 | 0.06 | 0.06 | 0.06 | 0.07 | 0.06 | 0.06 | 0.06 | 0.06 |

**Table 5.30:** *Computational time (given in seconds) taken to calculate the resample size $\hat{m}_{1,X}$, and the corresponding estimator $\tilde{\theta}_{brag,1}$.*

*All calculations were performed in* R *version 2.4.1 using a 1.7 GHz Intel Centrino processor. The number of bootstrap replications chosen for this table is $B = 1000$.*

*We can see from this table that the computational times required to calculate the statistic* $\tilde{\theta}_{brag,1}$ *for sample sizes* $n = 20$ *to* $n = 100$ *are not substantially different for the traditional and m-out-of-n bootstrap procedures. However, if one considers the time taken to calculate the resample size* $\hat{m}_{1,X}$*, then the procedures used to obtain* $\hat{m}_{1,X}$ *through the bootstrap (i.e.,* ***BS1, BS2,*** *and* ***BS3****) perform slower than the traditional bootstrap (although, to be fair, all of the procedures run in less than a second).*

*The m-out-of-n bootstrap procedures that determine the resample size* $\hat{m}_{1,X}$ *using the sample moment calculations (i.e.,* ***NB, BC0, BC1, BC2, UNBC1*** *and* ***UNBC2****) proved to be able to calculate both the choice of* $\hat{m}_{1,X}$ and *the estimator* $\tilde{\theta}_{brag,1}$ *in roughly the same amount of time as the traditional bootstrap for the sample sizes* $n = 20$ *to* $n = 100$*, and they actually perform* better *than the traditional bootstrap for larger samples (*$n = 1000$*). This is possibly due to the fact that these procedures are able to select the resample size using elementary calculations that do not require iterations,* and *because the bootstrap calculation of* $\tilde{\theta}_{brag,1}$ *is based on fewer observations.*

*For* practical *purposes we see that even for samples as large as* $n = 1000$*, the m-out-of-n bootstrap procedures for calculating the estimated resample size* and *the estimator itself are still very quick. For* $n = 1000$ *we see that the time taken to obtain the result (i.e., performing both calculations) ranges from 0.07 seconds to 0.81 seconds.*

# Chapter 6

# The $m$-out-of-$n$ bootstrap applied to hypothesis testing

## 6.1 Introduction

The construction of bootstrap confidence intervals is linked to the execution of bootstrap hypothesis tests because of the duality between confidence intervals and hypothesis testing, i.e., the null hypothesis is rejected if, and only if, the hypothesized value under the null hypothesis lies outside the confidence interval. However, Shao and Tu (1995) provide two reasons why it is important to consider bootstrap hypothesis testing separately: *" 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."*

Although there is a great deal of literature on the bootstrap, only a small proportion of it is devoted to bootstrap-based hypothesis testing. This is because the literature typically focuses on deriving bootstrap standard errors and constructing confidence intervals rather than hypothesis testing. 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. In their second guideline, they recommend that bootstrap hypothesis tests should be based on test statistics that are (at least asymptotically) pivotal. In a recent paper, Martin (2007) investigated bootstrap hypothesis testing for some common statistical problems. Allison (2008) developed a new method to evaluate the performance of bootstrap-based tests, and found that this new evaluation method can, among other things, detect defects of bootstrap critical values that can not be observed if the traditional evaluation method is employed.

The traditional $n$-out-of-$n$ bootstrap was used in all of the above-mentioned papers. Sakov (1998) applied the $m$-out-of-$n$ bootstrap in hypothesis testing to estimate the critical values and $p$-values of a test. However, it should be noted that Sakov (1998) only attempted to use the $m$-out-of-$n$ bootstrap to correct the test when one violates the first guideline stated in Hall and Wilson (1991). Moreover, Sakov (1998) only made use of non-pivotal statistics, which contradicts the second guideline of Hall and Wilson (1991).

In the next section two data-dependent methods of choosing the bootstrap resample size $m$, when testing hypotheses using the bootstrap, are derived. It would appear that this topic has been neglected in the bootstrap literature.

## 6.2 Two data-based methods for choosing the bootstrap resample size

Before introducing these two new data-dependent methods for choosing $m$ in hypothesis testing, some new notation needs to be established.

Let $\boldsymbol{X}_n = (X_1, X_2, \ldots, X_n)$ be a sample of i.i.d. random variables with unknown distribution function $F$. Consider, without loss of generality, the right-sided alternative hypothesis

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

where the parameter $\theta(F)$ is some functional of $F$.

Suppose that the test rejects $H_0$ if and only if

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

where $C_n(\alpha)$ is defined as the critical value satisfying

$$\mathrm{P}(T_n(\boldsymbol{X}_n) \geq C_n(\alpha) \mid H_0) \approx \alpha,$$

with $T_n(\boldsymbol{X}_n)$ an appropriate test statistic, and $\alpha$ $(0 < \alpha < 1)$ the prescribed nominal significance level of the test.

Unfortunately, the critical value $C_n(\alpha)$ is unknown, since $F$ is unknown and so we will need to estimate this quantity. Two possible bootstrap estimators for $C_n(\alpha)$ will be considered, namely $C_n(\alpha; \boldsymbol{X}_n)$ derived from the traditional bootstrap and $C_m(\alpha; \boldsymbol{X}_n)$ obtained from the $m$-out-of-$n$ bootstrap. Define $C_m(\alpha; \boldsymbol{X}_n)$ as follows:
Let $\boldsymbol{X}_m^* = (X_1^*, X_2^*, \ldots, X_m^*)$ be a bootstrap random sample drawn with replacement from the EDF of $\boldsymbol{X}_n$, $F_n$. In the bootstrap world we require that $\theta(F_n) = \theta_0$. However, this is seldom if ever the case, hence we need to transform the data $X_1, X_2, \ldots, X_n$. Denote the transformed variables by

$$W_i^0 \equiv W_i^0(\boldsymbol{X}_n), \qquad i = 1, 2, \ldots, n.$$

These variables are chosen such that $\theta(G_n) = \theta_0$ in the bootstrap world, where $G_n$ is the EDF of $\boldsymbol{W}_n^0 = (W_1^0, W_2^0, \ldots, W_n^0)$. The bootstrap random sample is now given by $\boldsymbol{W}_m^{0*} = (W_1^{0*}, W_2^{0*}, \ldots, W_m^{0*})$ drawn with replacement from $G_n$. Usually $\boldsymbol{W}_m^{0*}$ will depend on both $\boldsymbol{X}_n$ and $\boldsymbol{X}_m^*$. The bootstrap estimator $C_m(\alpha; \boldsymbol{X}_n)$ is now chosen such that

$$\mathrm{P}^*(T_m(\boldsymbol{W}_m^{0*}) \geq C_m(\alpha; \boldsymbol{X}_n)) \approx \alpha,$$

where $\mathrm{P}^*$ denotes the conditional probability law of $\boldsymbol{W}_m^{0*}$ given $\boldsymbol{X}_n$. The bootstrap critical value $C_n(\alpha; \boldsymbol{X}_n)$ is obtained by simply setting $m = n$ in the above formulation. Note that $C_m(\alpha; \boldsymbol{X}_n)$ and $C_n(\alpha; \boldsymbol{X}_n)$ can be approximated by a simple Monte-Carlo algorithm suggested by Efron (1979). These approximations are denoted by $C_{m,B}(\alpha; \boldsymbol{X}_n)$ and $C_{n,B}(\alpha; \boldsymbol{X}_n)$ respectively.

In addition to bootstrap critical values one can also define bootstrap $p$-values. Recall that the $p$-value of the test is given by:

$$\pi^0(\mathrm{x}_n) := \mathrm{P}\left(T_n(\boldsymbol{X}_n) \geq T_n(\mathrm{x}_n) \mid H_0\right),$$

for each realisation $\mathrm{x}_n = (x_1, x_2, \ldots, x_n)$. Hence the $m$-out-of-$n$ bootstrap estimate of $\pi^0(\cdot)$ is given by:

$$p_{boot,m} := \mathrm{P}^*(T_m(\boldsymbol{W}_m^{0*}) \geq T_n(\boldsymbol{X}_n)).$$

Once again we define the traditional bootstrap $p$-value, $p_{boot,n}$, by simply substituting $m = n$ in the above expression. The Monte-Carlo approximations of the above ideal bootstrap quantities are denoted throughout by $p_{boot,m,B}$ and $p_{boot,n,B}$ respectively.

### 6.2.1 A data-dependent choice of $m$ based on critical values

We base our first choice of the bootstrap resample size $m$ when applied to bootstrap hypothesis testing on a measure used to assess the accuracy of bootstrap critical values. This measure considers the size of the test using this bootstrap critical value, i.e.,

$$\mathrm{P}\left(T_n(\boldsymbol{X}_n) \geq C_m(\alpha; \boldsymbol{X}_n) \mid H_0\right), \tag{6.1}$$

and it deems the bootstrap critical value to be accurate if this size is approximately equal to the nominal significance level $\alpha$. This method of evaluating the traditional bootstrap critical value is used in a number of articles in the bootstrap literature, see for example Boos and Brownie (1989), Boos, Janssen and Veraverbeke (1989), Fisher and Hall (1990), Nankervis and Savin (1996), Cao and van Keilegom (2006), Davidson and MacKinnon (2004), and Martin (2007).

The procedure we propose is based on the following $m$-out-of-$n$ bootstrap estimator of the probability given in (6.1):

$$a_m^* := \mathrm{P}^*(T_m(\boldsymbol{W}_m^{0*}) \geq C_m(\alpha; \boldsymbol{W}_m^{0*})), \tag{6.2}$$

where we define $C_m(\alpha; \boldsymbol{W}_m^{0*})$ as the value satisfying

$$\mathrm{P}^{**}\left(T_m(\boldsymbol{V}_m^{0**}) \geq C_m(\alpha; \boldsymbol{W}_m^{0*})\right) \approx \alpha.$$

Here $\mathrm{P}^{**}$ refers to the conditional probability law of $\boldsymbol{V}_m^{0**}$ given $\boldsymbol{W}_m^{0*}$, and $\boldsymbol{V}_m^{0**}$ is defined as follows: Consider the bootstrap sample $\boldsymbol{W}_m^{0*}$. In the double bootstrap world we would like to have that $\theta(G_m^*) = \theta_0$, where $G_m^*$ is the EDF of $\boldsymbol{W}_m^{0*}$. Unfortunately, this is seldom (if ever) the case, hence we need to transform $W_1^{0*}, W_2^{0*}, \ldots, W_m^{0*}$. Denote the transformed variables by

$$V_i^{0*} \equiv V_i^{0*}(\boldsymbol{W}_m^{0*}), \qquad i = 1, 2, \ldots, m.$$

These variables are chosen such that $\theta(H_m^*) = \theta_0$, in the double bootstrap world, where $H_m^*$ is the EDF of $\boldsymbol{V}_m^{0*} = (V_1^{0*}, V_2^{0*}, \ldots, V_m^{0*})$. The double bootstrap sample is now given by $\boldsymbol{V}_m^{0**} = (V_1^{0**}, V_2^{0**}, \ldots, V_m^{0**})$ drawn with replacement from $H_m^*$.

Various choices of $m$ are used in (6.2) and the choice which leads to a value closest to the nominal significance level $\alpha$ is considered to be the best choice of $m$. We can now define our first data-based method of choosing $m$ as follows:

$$\hat{m} := \arg \min_{n_0 \leq m \leq n} |a_m^* - \alpha|, \tag{6.3}$$

where $n_0$ is some suitably chosen non-zero positive integer.

The ideal bootstrap estimator $\hat{m}$ given in (6.3) can be approximated by the following Monte-Carlo algorithm:

**Approximating $\hat{m}$ for bootstrap hypothesis testing (using the critical value approach)**

1. Given the sample $X_1, X_2, \ldots, X_n$, transform the values such that they reflect the conditions stated in the null hypothesis, i.e., we now have $W_1^0, W_2^0, \ldots, W_n^0$. Next, for some fixed value $n_0$, fix a set of resample sizes $n_0 = m_1 < m_2 < \cdots < m_S = n$ where $S \geq 2$.

2. For a chosen value $s \in \{1, 2, \ldots, S\}$ proceed as follows:

   a. Obtain the bootstrap sample $W_1^{0*}, W_2^{0*}, \ldots, W_{m_s}^{0*}$ by sampling $m_s$ observations with replacement from the transformed sample $W_1^0, W_2^0, \ldots, W_n^0$. Calculate the statistic

$$T_{m_s}^* := T_{m_s}(W_1^{0*}, W_2^{0*}, \ldots, W_{m_s}^{0*}).$$

   b. Now, transform $W_1^{0*}, W_2^{0*}, \ldots, W_{m_s}^{0*}$ to obtain $V_1^{0*}, V_2^{0*}, \ldots, V_{m_s}^{0*}$ so that the transformed bootstrap sample reflects the conditions stated in the null hypothesis. Using this sample:

      i. Obtain a double bootstrap sample $V_1^{0**}, V_2^{0**}, \ldots, V_{m_s}^{0**}$ by sampling $m_s$ observations with replacement from the transformed bootstrap sample $V_1^{0*}, V_2^{0*}, \ldots, V_{m_s}^{0*}$.

      ii. For the double bootstrap sample calculate the statistic

$$T_{m_s}^{**} := T_{m_s}(V_1^{0**}, V_2^{0**}, \ldots, V_{m_s}^{0**}).$$

      iii. Repeat steps 2.b.i and 2.b.ii $R$ times to obtain the $R$ statistics denoted by $T_{m_s,1}^{**}, T_{m_s,2}^{**}, \ldots, T_{m_s,R}^{**}$. Sort these statistics and denote them by $T_{m_s,(1)}^{**} \leq T_{m_s,(2)}^{**} \leq \cdots \leq T_{m_s,(R)}^{**}$.

      iv. Approximate the double bootstrap critical value $C_{m_s}(\alpha; \boldsymbol{W}_{m_s}^{0*})$ by

$$C_{m_s,R}(\alpha; \boldsymbol{W}_{m_s}^{0*}) = T_{m_s,(\lfloor(1-\alpha)R\rfloor)}^{**}.$$

   c. Repeat steps 2.a and 2.b $B$ times so that we now have $T_{m_s,1}^*, T_{m_s,2}^*, \ldots, T_{m_s,B}^*$, and $C_{m_s,R,1}(\alpha; \boldsymbol{W}_{m_s}^{0*}), C_{m_s,R,2}(\alpha; \boldsymbol{W}_{m_s}^{0*}), \ldots, C_{m_s,R,B}(\alpha; \boldsymbol{W}_{m_s}^{0*})$.

   d. Calculate

$$\Delta(m_s) := \left| \frac{1}{B} \sum_{b=1}^{B} \mathrm{I}\left(T_{m_s,b}^* \geq C_{m_s,R,b}(\alpha; \boldsymbol{W}_{m_s}^{0*})\right) - \alpha \right|.$$

3. Repeat step 2 for each $s = 1, 2, \ldots, S$ to obtain

$$\Delta(m_1), \Delta(m_2), \ldots, \Delta(m_S).$$

4. Calculate,

$$\hat{m}_{B,R} = \arg \min_{m \in \{m_1, m_2, \ldots, m_S\}} \Delta(m).$$

In order to evaluate the performance of the estimated sample size $\hat{m}$ the following criterion is used:

$$\mathrm{P}(T_n(\boldsymbol{X}_n) \geq C_{\hat{m}}(\alpha; \boldsymbol{X}_n) \mid H_0), \tag{6.4}$$

and is compare to the behaviour of the traditional bootstrap using the criterion:

$$\mathrm{P}(T_n(\boldsymbol{X}_n) \geq C_n(\alpha; \boldsymbol{X}_n) \mid H_0). \tag{6.5}$$

### 6.2.2 A data-dependent choice of $m$ based on $p$-values

The second data-based method of choosing $m$ is derived by making use of the fact that the $p$-value of a test (considered as a random variable with respect to $\boldsymbol{X}_n$) is uniformly distributed on the interval $(0, 1)$ under the null hypothesis. We employ a double bootstrap procedure to obtain bootstrap replications of the bootstrap estimator of the $p$-value and then use these values in a goodness-of-fit test for uniformity.

The estimator that we propose for $m$ is

$$\hat{m} := \arg \min_{n_0 \leq m \leq n} \gamma(\mathcal{L}(p^*_{boot,m}), \mathcal{U}), \tag{6.6}$$

where $\mathcal{U}$ denotes the uniform $(0,1)$ distribution, $p^*_{boot,m}$ is the double bootstrap $p$-value, based on a double bootstrap sample of size $m$, defined as

$$p^*_{boot,m} := \mathrm{P}^{**}\left(T_m(\boldsymbol{V}^{0**}_m) \geq T_m(\boldsymbol{W}^{0*}_m)\right),$$

$\mathcal{L}(p^*_{boot,m})$ denotes the bootstrap distribution of $p^*_{boot,m}$, and $\gamma(\cdot\,,\cdot)$ is some discrepancy measure which is able to measure the difference between two distributions.

The ideal bootstrap estimator, $\hat{m}$, given in (6.6) can be approximated by the following Monte-Carlo algorithm:

**Approximating $\hat{m}$ for bootstrap hypothesis testing (using the $p$-value approach)**

1. Given the sample $X_1, X_2, \ldots, X_n$, transform the values such that they reflect the conditions stated in the null hypothesis, i.e., we now have $W^0_1, W^0_2, \ldots, W^0_n$. Next, for some fixed value $n_0$, fix a set of resample sizes $n_0 = m_1 < m_2 < \cdots < m_S = n$ where $S \geq 2$.

2. For a chosen value $s \in \{1, 2, \ldots, S\}$ proceed as follows:

   a. Obtain the bootstrap sample $W^{0*}_1, W^{0*}_2, \ldots, W^{0*}_{m_s}$ by sampling $m_s$ observations with replacement from the transformed sample $W^0_1, W^0_2, \ldots, W^0_n$. Calculate $T^*_{m_s} := T_{m_s}(W^{0*}_1, W^{0*}_2, \ldots, W^{0*}_{m_s})$.

   b. Now, transform $W^{0*}_1, W^{0*}_2, \ldots, W^{0*}_{m_s}$ to obtain $V^{0*}_1, V^{0*}_2, \ldots, V^{0*}_{m_s}$ so that the transformed bootstrap sample reflects the conditions stated in the null hypothesis. Using this sample:

      i. Obtain a double bootstrap sample $V^{0**}_1, V^{0**}_2, \ldots, V^{0**}_{m_s}$ by sampling $m_s$ observations with replacement from the transformed bootstrap sample $V^{0*}_1, V^{0*}_2, \ldots, V^{0*}_{m_s}$.

      ii. For the double bootstrap sample calculate the statistic

      $$T^{**}_{m_s} := T_{m_s}(V^{0**}_1, V^{0**}_2, \ldots, V^{0**}_{m_s}).$$

      iii. Repeat steps 2.b.i and 2.b.ii $R$ times to obtain the $R$ statistics denoted by $T^{**}_{m_s,1}, T^{**}_{m_s,2}, \ldots, T^{**}_{m_s,R}$.

      iv. Approximate the double bootstrap $p$-value, $p^*_{boot,m_s}$, by

      $$p^*_{boot,m_s,R} := \frac{1}{R} \sum_{r=1}^{R} \mathrm{I}\left(T^{**}_{m_s,r} \geq T^*_{m_s}\right).$$

   c. Repeat steps 2.a and 2.b $B$ times to obtain $p^*_{boot,m_s,R,1}, p^*_{boot,m_s,R,2}, \ldots, p^*_{boot,m_s,R,B}$.

   d. Calculate

   $$\gamma_{m_s} := \gamma(F^*_{B,m_s}, \mathcal{U}),$$

   where $F^*_{B,m_s}$ is the EDF of $p^*_{boot,m_s,R,b}$, $b = 1, 2, \ldots, B$.

3. Repeat step 2 for each $s = 1, 2, \ldots, S$ to obtain
$$\gamma_{m_1}, \gamma_{m_2}, \ldots, \gamma_{m_S}.$$
4. Calculate,
$$\hat{m}_{B,R} = \arg \min_{m \in \{m_1, m_2, \ldots, m_S\}} \gamma_m.$$

There exist several choices for the function $\gamma$ in the above algorithm. In the simulation study below, we consider using the well known Kolmogorov-Smirnov (KS) and Cramér-von Mises (CvM) discrepancy measures. These measures are defined (see, for example, D'agostino and Stephens (1986)) as

**Kolmogorov-Smirnov:**
$$\gamma_{m_s,\mathrm{KS}} := \max\left\{\max_b \left(\frac{b}{n} - Z_{(b)}\right) \ , \ \max_b \left(Z_{(b)} - \frac{b-1}{n}\right)\right\},$$

**Cramér-von Mises:**
$$\gamma_{m_s,\mathrm{CvM}} := \sum_{b=1}^{B} \left\{Z_{(b)} - \frac{b - \frac{1}{2}}{B}\right\}^2 + \frac{1}{12B},$$

where $Z_{(b)}$ denotes the $b^{\text{th}}$ order statistic of $p^*_{boot,m_s,R,b}$, $b = 1, 2, \ldots, B$, in the algorithm stated above.

To evaluate the performance of the estimated sample size $\hat{m}$ we use the following criterion:
$$A_{moon} := \mathrm{E}\left(\left|p_{boot,\hat{m}} - \pi^0(\boldsymbol{X}_n)\right|\right), \tag{6.7}$$
and compare this to
$$A_{trad} := \mathrm{E}\left(\left|p_{boot,n} - \pi^0(\boldsymbol{X}_n)\right|\right), \tag{6.8}$$
where We will approximate $\pi^0(\mathbf{x}_n)$ using a Monte-Carlo simulation for each realized sample $\mathbf{x}_n$.

**Remark:** Sakov (1998) introduced the following data-based choices of $m$ based on a grid of $m$-values, $n_0 = m_1 < m_2 < \cdots < m_S = n$, $S \geq 2$:

1. *A data-dependent choice of $m$ based on critical values:*
   Choose $\hat{m}$ as follows:
$$\hat{m} := \arg \min_{\{m_j: j=2,3,\ldots,S\}} \left|C_{m_j}(\alpha; \boldsymbol{X}_n) - C_{m_{j-1}}(\alpha; \boldsymbol{X}_n)\right|. \tag{6.9}$$

2. *A data-dependent choice of $m$ based on p-values:*
   Choose $\hat{m}$ as follows:
$$\hat{m} := \arg \min_{\{m_j: j=2,3,\ldots,S\}} \left|p_{boot,m_j} - p_{boot,m_{j-1}}\right|. \tag{6.10}$$

In both of the cases stated above the value $\hat{m}$ is obtained from Monte-Carlo approximations as suggested by Efron (1979).

## 6.3 Monte-Carlo study

In this section we present the results of two Monte-Carlo studies. The first study compares the traditional bootstrap critical value $C_n(\alpha; \boldsymbol{X}_n)$ to the $\hat{m}$-out-of-$n$ bootstrap critical value $C_{\hat{m}}(\alpha; \boldsymbol{X}_n)$, based on both choices of $m$ given in (6.3) and (6.9). The second study compares the traditional bootstrap $p$-value $p_{boot,n}$ and the $\hat{m}$-out-of-$n$ bootstrap $p$-value $p_{boot,\hat{m}}$, again using our data-dependent choice of $m$ and the choice proposed by Sakov (1998), defined in (6.6) and (6.10) respectively.

For the first Monte-Carlo study the size estimates were calculated as the proportion of $MC = 4000$ Monte-Carlo samples that resulted in the rejection of $H_0$ using the bootstrap critical values $C_{n,B}(\alpha; \boldsymbol{X}_n)$ and $C_{\hat{m},B}(\alpha; \boldsymbol{X}_n)$. The size estimate based on $C_{n,B}(\alpha; \boldsymbol{X}_n)$ is denoted by $S_{trad}$, the size estimate based on $C_{\hat{m},B}(\alpha; \boldsymbol{X}_n)$ using the choice of $m$ given in (6.3) is denoted by $S_{new}$, and the size estimation based on $C_{\hat{m},B}(\alpha; \boldsymbol{X}_n)$ using the choice of $m$ given in (6.9) is denoted by $S_{sak}$. The standard errors of these estimated sizes are less than or equal to $\sqrt{0.25/4000} = 0.0079$.

For the second Monte-Carlo study, $A_{moon}$ defined in equation (6.7), was approximated as the average (over $MC = 1500$ replications) of the absolute differences between the $\hat{m}$-out-of-$n$ bootstrap $p$-value, $p_{boot,\hat{m},B}$, and the theoretical $p$-value, $\pi^0$. A similar calculation, using the same set of data, was done for $A_{trad}$ defined in equation (6.8). These approximations are denoted by $A_{moon,MC}$ and $A_{trad,MC}$ respectively. To facilitate comparison we then calculate the following ratio

$$RATIO := \frac{A_{moon,MC}}{A_{trad,MC}} \times 100\%.$$

Note that when the calculated value of $RATIO$ is greater than 100, then it implies that the traditional bootstrap performs better than the $\hat{m}$-out-of-$n$ bootstrap. When $RATIO$ is calculated using the data-dependent choice of $m$ given in (6.6), then we denote the result by $R_{new}$, and when it is calculated using the data-dependent choice given in (6.10) it is denoted by $R_{sak}$. The maximum standard error of these absolute differences was found to be 0.0008 (for samples of size $n = 20$).

Various configurations of distributions, sample sizes and testing scenarios (viz. testing the mean and the variance of a population) were used in both Monte-Carlo studies. A nominal significance level of $\alpha = 0.05$ was adopted. Bootstrap critical values and $p$-values were based on $B = 1000$ independent bootstrap replications and $R = 200$ independent double bootstrap replications. The sample sizes, $n$, which were employed differed depending on the study being performed, but the grid of resample sizes $m_1, m_2, \ldots, m_9$ which were utilized for $m$-out-of-$n$ bootstrap calculations were defined throughout as $m_s = n\left(\frac{s+1}{10}\right)$, $s = 1, 2, \ldots, 9$. Calculations were performed using R and FORTRAN (using the IMSL library and in double precision).

### 6.3.1 The mean in the univariate case

Let $\boldsymbol{X}_n = (X_1, X_2, \ldots, X_n)$ denote a random sample from some unknown univariate distribution function $F$ with finite mean $\mu$. For simplicity we consider testing the hypothesis

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

The test is based on the following asymptotically pivotal test statistic

$$T_n(\boldsymbol{X}_n) = \frac{\sqrt{n}(\bar{X}_n - \mu_0)}{S_n(\boldsymbol{X}_n)},$$

where $S_n^2(\boldsymbol{X}_n) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$ and $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. We define the quantities $W_i^0$, $W_i^{0*}$, $V_i^{0*}$ and $V_i^{0**}$, which appear in the Monte-Carlo algorithm, as

$$W_i^0 = X_i - \bar{X}_n + \mu_0, \qquad i = 1, 2, \ldots, n,$$

$$W_i^{0*} = X_i^* - \bar{X}_n + \mu_0, \qquad i = 1, 2, \ldots, m,$$
$$V_i^{0*} = W_i^{0*} - \bar{W}_m^{0*} + \mu_0, \qquad i = 1, 2, \ldots, m,$$

and

$$V_i^{0**} = W_i^{0**} - \bar{W}_m^{0*} + \mu_0, \qquad i = 1, 2, \ldots, m.$$

The corresponding test statistics are then defined as

$$T_m(\boldsymbol{W}_m^{0*}) = \frac{\sqrt{m}(\bar{W}_m^{0*} - \mu_0)}{S_m(\boldsymbol{W}_m^{0*})} = \frac{\sqrt{m}(\bar{X}_m^* - \bar{X}_n)}{S_m(\boldsymbol{X}_m^*)},$$

where $S_m^2(\boldsymbol{W}_m^{0*}) = \frac{1}{m}\sum_{i=1}^m (W_i^{0*} - \bar{W}_m^{0*})^2$, $S_m^2(\boldsymbol{X}_m^*) = \frac{1}{m}\sum_{i=1}^m (X_i^* - \bar{X}_m^*)^2$, $\bar{W}_m^{0*} = \frac{1}{m}\sum_{i=1}^m W_i^{0*}$, $\bar{X}_m^* = \frac{1}{m}\sum_{i=1}^m X_i^*$, and

$$T_m(\boldsymbol{V}_m^{0**}) = \frac{\sqrt{m}(\bar{V}_m^{0**} - \mu_0)}{S_m(\boldsymbol{V}_m^{0**})} = \frac{\sqrt{m}(\bar{X}_m^{**} - \bar{X}_m^*)}{S_m(\boldsymbol{X}_m^{**})},$$

where $S_m^2(\boldsymbol{V}_m^{0**}) = \frac{1}{m}\sum_{i=1}^m (V_i^{0**} - \bar{V}_m^{0**})^2$, $S_m^2(\boldsymbol{X}_m^{**}) = \frac{1}{m}\sum_{i=1}^m (X_i^{**} - \bar{X}_m^{**})^2$, $\bar{V}_m^{0**} = \frac{1}{m}\sum_{i=1}^m V_i^{0**}$, and $\bar{X}_m^{**} = \frac{1}{m}\sum_{i=1}^m X_i^{**}$.

The configurations for this Monte-Carlo study were:

- ⋆ Sample sizes: $n = 20, 30$ and 50.
- ⋆ Distributions: standard normal, standard exponential, and double exponential with mean 0 and variance 2.

| | Normal | | | Exponential | | | Double Exponential | | |
|---|---|---|---|---|---|---|---|---|---|
| $n$ | $\mathbf{S}_{trad}$ | $\mathbf{S}_{new}$ | $\mathbf{S}_{sak}$ | $\mathbf{S}_{trad}$ | $\mathbf{S}_{new}$ | $\mathbf{S}_{sak}$ | $\mathbf{S}_{trad}$ | $\mathbf{S}_{new}$ | $\mathbf{S}_{sak}$ |
| 20 | 0.047 | 0.046 | 0.043 | 0.037 | 0.036 | 0.036 | 0.062 | 0.054 | 0.058 |
| 30 | 0.051 | 0.048 | 0.044 | 0.043 | 0.042 | 0.042 | 0.059 | 0.053 | 0.056 |
| 50 | 0.049 | 0.049 | 0.047 | 0.046 | 0.048 | 0.047 | 0.055 | 0.052 | 0.053 |

**Table 6.1:** *The estimated sizes of the test $H_0 : \mu = \mu_0$ using the traditional bootstrap and the $\hat{m}$-out-of-n bootstrap.*

| | Normal | | | Exponential | | | Double Exponential | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\mathbf{R}_{sak}$ | $\mathbf{R}_{new}$ | | $\mathbf{R}_{sak}$ | $\mathbf{R}_{new}$ | | $\mathbf{R}_{sak}$ | $\mathbf{R}_{new}$ | |
| $n$ | | KS | CvM | | KS | CvM | | KS | CvM |
| 20 | 117.06 | 99.06 | 98.67 | 106.29 | 91.03 | 87.82 | 103.48 | 93.62 | 93.36 |
| 30 | 110.91 | 98.73 | 96.98 | 111.73 | 85.18 | 85.09 | 103.92 | 94.13 | 92.12 |
| 50 | 95.59 | 90.42 | 87.73 | 109.41 | 90.19 | 88.05 | 96.87 | 89.45 | 90.13 |

**Table 6.2:** *The ratio of $A_{moon,MC}$ to $A_{trad,MC}$ (expressed as a percentage) using different choices of $\hat{m}$ for the test $H_0 : \mu = \mu_0$.*

**Conclusions for the mean in the univariate case:**

⋆ From Table 6.1:

- In the normal case the three test procedures perform satisfactorily, although $S_{trad}$ and $S_{new}$ are slightly closer to the prescribed significance level than $S_{sak}$.
- For the exponential distribution, all three procedures yield almost identical estimated sizes.
- As far as the double exponential distribution is concerned, $S_{new}$ noticeably outperforms both $S_{trad}$ and $S_{sak}$.
- We note that, for this rather easy problem of testing the population mean, all three tests have estimated sizes close to the prescribed significance level of 5% for all distributions considered (even for sample sizes as small as $n = 30$).
- The average resample sizes obtained for the calculations found in this table are roughly $n/2$.

⋆ From Table 6.2:

- The most striking feature of this table is that $R_{new}$ outperforms both the traditional bootstrap and $R_{sak}$ for all distributions and sample sizes considered.
- In most cases $R_{sak}$ performs worse than even the traditional bootstrap.
- In the normal case the improvement of $R_{new}$ over $R_{sak}$ is between 8% and 19%, for the exponential case the improvement is between 19% and 26%, and for the double exponential case it is between 6% and 11%.
- Another feature which is also quite evident in this table is that $R_{new}$ based on the Cramér-von Mises discrepancy measure (CvM) is better than $R_{new}$ based on the Kolmogorov-Smirnov (KS) discrepancy measure in all cases except for the double exponential distribution with $n = 50$, where it is only slightly inferior.
- For this table we find that the average resample sizes vary between $2n/3$ and $3n/4$.

### 6.3.2 The variance in the univariate case

Let $\boldsymbol{X}_n = (X_1, X_2, \ldots, X_n)$ denote a random sample from some unknown univariate distribution function $F$ with variance $\sigma^2$, $0 < \sigma^2 < \infty$. In this scenario we test the hypothesis

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

The asymptotically pivotal test statistic which is used to test this hypothesis is

$$T_n(\boldsymbol{X}_n) = \frac{\sqrt{n}(S_n^2(\boldsymbol{X}_n) - \sigma_0^2)}{\sqrt{\hat{\mu}_4(\boldsymbol{X}_n) - S_n^4(\boldsymbol{X}_n)}},$$

where

$$\hat{\mu}_4(\boldsymbol{X}_n) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^4.$$

We define the quantities $W_i^0$, $W_i^{0*}$, $V_i^{0*}$ and $V_i^{0**}$, which appear in the Monte-Carlo algorithm, as

$$W_i^0 = \frac{X_i\sigma_0}{S_n(\boldsymbol{X}_n)}, \qquad i = 1, 2, \ldots, n,$$

$$W_i^{0*} = \frac{X_i^* \sigma_0}{S_n(\boldsymbol{X}_n)}, \qquad i = 1, 2, \ldots, m,$$

$$V_i^{0*} = \frac{W_i^{0*} \sigma_0}{S_n(\boldsymbol{W}_m^{0*})}, \qquad i = 1, 2, \ldots, m,$$

and

$$V_i^{0**} = \frac{W_i^{0**} \sigma_0}{S_n(\boldsymbol{W}_m^{0*})}, \qquad i = 1, 2, \ldots, m.$$

The corresponding test statistics are then defined as

$$T_m(\boldsymbol{W}_m^{0*}) = \frac{\sqrt{m}(S_m^2(\boldsymbol{W}_m^{0*}) - \sigma_0^2)}{\sqrt{\hat{\mu}_4(\boldsymbol{W}_m^{0*}) - S_m^4(\boldsymbol{W}_m^{0*})}} = \frac{\sqrt{m}(S_m^2(\boldsymbol{X}_m^*) - S_n^2(\boldsymbol{X}_n))}{\sqrt{\hat{\mu}_4(\boldsymbol{X}_m^*) - S_m^4(\boldsymbol{X}_m^*)}},$$

where

$$\hat{\mu}_4(\boldsymbol{W}_m^{0*}) = \frac{1}{m}\sum_{i=1}^{m}(W_i^{0*} - \bar{W}_m^{0*})^4 \qquad \text{and} \qquad \hat{\mu}_4(\boldsymbol{X}_m^*) = \frac{1}{m}\sum_{i=1}^{m}(X_i^* - \bar{X}_m^*)^4,$$

and

$$T_m(\boldsymbol{V}_m^{0**}) = \frac{\sqrt{m}(S_m^2(\boldsymbol{V}_m^{0**}) - \sigma_0^2)}{\sqrt{\hat{\mu}_4(\boldsymbol{V}_m^{0**}) - S_m^4(\boldsymbol{V}_m^{0**})}} = \frac{\sqrt{m}(S_m^2(\boldsymbol{X}_m^{**}) - S_m^2(\boldsymbol{X}_m^*))}{\sqrt{\hat{\mu}_4(\boldsymbol{X}_m^{**}) - S_m^4(\boldsymbol{X}_m^{**})}},$$

where

$$\hat{\mu}_4(\boldsymbol{V}_m^{0**}) = \frac{1}{m}\sum_{i=1}^{m}(V_i^{0**} - \bar{V}_m^{0**})^4 \qquad \text{and} \qquad \hat{\mu}_4(\boldsymbol{X}_m^{**}) = \frac{1}{m}\sum_{i=1}^{m}(X_i^{**} - \bar{X}_m^{**})^4.$$

The configurations for this Monte-Carlo study were:

- ⋆ Sample sizes: $n = 50, 100$ and 150.
- ⋆ Distributions: standard normal, standard exponential, and double exponential with mean 0 and variance 2.

| | Normal | | | Exponential | | | Double Exponential | | |
|---|---|---|---|---|---|---|---|---|---|
| $n$ | $\mathrm{S}_{trad}$ | $\mathrm{S}_{new}$ | $\mathrm{S}_{sak}$ | $\mathrm{S}_{trad}$ | $\mathrm{S}_{new}$ | $\mathrm{S}_{sak}$ | $\mathrm{S}_{trad}$ | $\mathrm{S}_{new}$ | $\mathrm{S}_{sak}$ |
| 50 | 0.043 | 0.047 | 0.047 | 0.027 | 0.035 | 0.034 | 0.034 | 0.042 | 0.039 |
| 100 | 0.051 | 0.053 | 0.055 | 0.036 | 0.045 | 0.044 | 0.040 | 0.049 | 0.051 |
| 150 | 0.053 | 0.052 | 0.055 | 0.039 | 0.047 | 0.046 | 0.041 | 0.050 | 0.048 |

**Table 6.3:** *The estimated sizes of the test $H_0 : \sigma^2 = \sigma_0^2$ using the traditional bootstrap and the $\hat{m}$-out-of-n bootstrap.*

**Conclusions for the variance in the univariate case:**

- ⋆ From Table 6.3:
  - For the normal distribution all three methods yielded estimated sizes quite close to the prescribed level (even for $n = 50$).

| | Normal | | | Exponential | | | Double Exponential | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\mathbf{R}_{sak}$ | $\mathbf{R}_{new}$ | | $\mathbf{R}_{sak}$ | $\mathbf{R}_{new}$ | | $\mathbf{R}_{sak}$ | $\mathbf{R}_{new}$ | |
| $n$ | | KS | CvM | | KS | CvM | | KS | CvM |
| 50 | 95.32 | 83.27 | 82.76 | 89.61 | 93.38 | 93.20 | 78.92 | 87.41 | 86.79 |
| 100 | 99.91 | 87.21 | 86.38 | 84.87 | 93.09 | 92.79 | 83.38 | 86.92 | 86.80 |
| 150 | 99.97 | 93.70 | 91.25 | 92.94 | 91.90 | 90.29 | 86.99 | 85.74 | 85.70 |

**Table 6.4:** *The ratio of $A_{moon,MC}$ to $A_{trad,MC}$ (expressed as a percentage) using different choices of $\hat{m}$ for the test $H_0 : \sigma^2 = \sigma_0^2$.*

- For both the exponential and the double exponential distributions we find that $S_{sak}$ and $S_{new}$ have very similar estimated sizes, and that they both outperform $S_{trad}$ significantly.
- Once again we find that the average resample sizes used are taken to be roughly $n/2$.

⋆ From Table 6.4:

- In all cases $R_{sak}$ and $R_{new}$ outperform the traditional bootstrap $p$-values.
- For the normal distribution the improvement of $R_{new}$ over $R_{sak}$ is between 8% and 13%.
- For the exponential and the double exponential distributions, $R_{sak}$ performs slightly better than $R_{new}$ for sample sizes $n = 50$ and $n = 100$, but $R_{new}$ is slightly better than $R_{sak}$ when $n = 150$.
- Once again, $R_{new}$ based on CvM is slightly better than $R_{new}$ based on KS.
- The average resample sizes which are used in these calculations are typically in the region of $4n/5$.

## 6.4 Conclusions

In this chapter we considered the problem of choosing a data-dependent choice of $m$ when applying the $m$-out-of-$n$ bootstrap to hypothesis testing, a topic which has been largely neglected in the bootstrap statistical literature. The methods we proposed are quite general, and can be applied in a variety of testing scenarios. The testing methods which employed the data-dependent choices of $m$ (developed in this chapter), were found to be very effective when compared to both the traditional bootstrap test and a testing procedure where $m$ was chosen using an alternative method suggested in the literature.

Finally, when using the $p$-value approach to derive data-dependent choices of $m$, we strongly recommend that the procedure based on the Cramér-von Mises discrepancy measure should be used.

# Appendix A

# The Edgeworth and Cornish-Fisher expansion

This section will be devoted to the explanation and derivation of the Edgeworth expansion. This expansion is inverted to form the Cornish-Fisher expansion which is mentioned in Chapter 5. The derivations provided here are simply a summary of the work found in Hall (1992) and Chung and Lee (2001).

## A.1 Edgeworth expansions of the sum of i.i.d random variables

The derivation of the Edgeworth expansion is a fairly involved process, incorporating, among other things, characteristic functions, cumulants, Taylor expansions, Hermite polynomials and Fourier-Stieltjes transforms. Definitions of some of these concepts can be found in the appendices that follow.

For the beginning of the discussion we will focus on the Edgeworth expansion of the distribution of the simple statistic $\sum_{i=1}^{n} X_i$, with i.i.d. $X_1, X_2, \ldots, X_n$. This may seem like a trivial starting point, but many statistics of interest have a form similar to this and they typically also have limiting normal distributions (consider $U$-statistics for example).

To begin the derivation a few definitions are required.

- ⋆ Let $X_1, X_2, \ldots, X_n$ be i.i.d random variables with mean $\mu$ and standard deviation $\sigma$.
- ⋆ Let $\bar{X}_n$ be defined as the sample mean of the $X_i$'s:

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

- ⋆ Let $Y_1, Y_2, \ldots, Y_n$ be defined as the standardized variables $Y_i = (X_i - \mu)/\sigma$. (i.e., the $Y_i$'s have mean 0 and standard deviation 1.)
- ⋆ Let $T_n$ be defined as the standardized version of $\bar{X}_n$

$$\begin{aligned} T_n &= \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y_i. \end{aligned}$$

- ⋆ Let $\chi_Y(t)$ be the characteristic function of the $Y_i$'s.
- ⋆ Let $\kappa_Y(t)$ be the cumulant function of the $Y_i$'s.
- ⋆ Let $\chi_{T_n}(t)$ be the characteristic function of $T_n$.
- ⋆ Let $\kappa_j$ be the $j^{\text{th}}$ cumulant of the $Y_i$'s.

We know from the Central Limit Theorem that $T_n$ is asymptotically standard normal. Therefore, the characteristic function of $T_n$, $\chi_{T_n}(t)$, will converge to the characteristic function of the standard normal distribution, i.e.,

$$\chi_{T_n}(t) = \mathrm{E}\left(e^{itT_n}\right) \to \mathrm{E}\left(e^{itZ}\right) = e^{-t^2/2}, \quad -\infty < t < \infty, \tag{1.1}$$

where $Z \sim \mathrm{N}(0,1)$. This convergence is useful, but it will be more useful to know at what rate this happens. To show this we will need the following result:

$$\chi_{T_n}(t) = \left[\chi_Y\left(t/\sqrt{n}\right)\right]^n, \tag{1.2}$$

the proof of which is fairly simple.

To show the rate of convergence we will apply series expansions of the right-hand side of equation (1.2), and then equate that series to the left-hand side of the equation, providing an expansion for $\chi_{T_n}(t)$.

We will now show that $\chi_{T_n}(t)$ can be expanded into the following expression:

$$\chi_{T_n}(t) = e^{-t^2/2} + n^{-1/2}e^{-t^2/2}r_1(it) + n^{-1}e^{-t^2/2}r_2(it) + \cdots + n^{-j/2}e^{-t^2/2}r_j(it) + \ldots \tag{1.3}$$

**Proof:**
The Taylor series expansion of the cumulant function, $\kappa_Y(t) = \ln \chi_Y(t)$, is:

$$\begin{aligned} \kappa_Y(t) &= \sum_{j=0}^{\infty} \frac{\left(\frac{\partial}{\partial t}\right)^j \kappa_Y(t)|_{t=0}}{j!} t^j \\ &= \sum_{j=0}^{\infty} \frac{i^j \kappa_j}{j!} t^j \\ &= it\kappa_1 + \tfrac{1}{2}(it)^2\kappa_2 + \cdots + \frac{1}{j!}(it)^j\kappa_j + \ldots \end{aligned}$$

(where $i^j\kappa_j = \frac{\partial^j}{\partial t^j}\kappa_Y(t)|_{t=0}$), therefore, an expansion of the characteristic function, $\chi_Y(t)$, is:

$$\begin{aligned} \chi_Y(t) &= e^{\kappa_Y(t)} \\ &= \exp\left(it\kappa_1 + \tfrac{1}{2}(it)^2\kappa_2 + \cdots + \frac{1}{j!}(it)^j\kappa_j + \ldots\right) \end{aligned} \tag{1.4}$$

Combining equations (1.2) and (1.4), and using the fact that $\mathrm{E}(Y) = \kappa_1 = 0$ and $\mathrm{Var}(Y) = \kappa_2 = 1$, we get:

$$\begin{aligned} \chi_{T_n}(t) &= \left[\chi_Y\left(t/\sqrt{n}\right)\right]^n \\ &= \left[\exp\left(i\frac{t}{\sqrt{n}}\kappa_1 + \tfrac{1}{2}\left(i\frac{t}{\sqrt{n}}\right)^2\kappa_2 + \frac{1}{6}\left(i\frac{t}{\sqrt{n}}\right)^3\kappa_3 + \cdots + \frac{1}{j!}\left(i\frac{t}{\sqrt{n}}\right)^j\kappa_j + \ldots\right)\right]^n \end{aligned}$$

$$
\begin{aligned}
&= \left[\exp\left(\tfrac{1}{2}\left(i\frac{t}{\sqrt{n}}\right)^2 + \frac{1}{6}\left(i\frac{t}{\sqrt{n}}\right)^3\kappa_3 + \dots + \frac{1}{j!}\left(i\frac{t}{\sqrt{n}}\right)^j\kappa_j + \dots\right)\right]^n \\
&= \left[\exp\left(\frac{-t^2}{2n} + \frac{1}{6}\left(i\frac{t}{\sqrt{n}}\right)^3\kappa_3 + \dots + \frac{1}{j!}\left(i\frac{t}{\sqrt{n}}\right)^j\kappa_j + \dots\right)\right]^n \\
&= \exp\left(-\frac{1}{2}t^2 + n^{-1/2}\frac{1}{6}(it)^3\kappa_3 + \dots + \frac{1}{j!}(it)^j n^{(-j+2)/2}\kappa_j + \dots\right) \qquad (1.5) \\
&= \exp\left(-\frac{1}{2}t^2\right)\underbrace{\exp\left(n^{-1/2}\left[\frac{1}{6}(it)^3\kappa_3\right] + \dots + n^{(-j+2)/2}\left[\frac{1}{j!}(it)^j\kappa_j\right] + \dots\right)}_{a(t)}
\end{aligned}
$$

A Taylor expansion and then some simplification of the $a(t)$ term above yields*:

$$
\begin{aligned}
a(t) &= 1 + n^{-1/2}\left[\frac{1}{6}(it)^3\kappa_3\right] + n^{-1}\left[\frac{1}{72}(it)^6\kappa_3^2 + \frac{1}{24}(it)^4\kappa_4\right] + n^{-3/2}\left[\frac{1}{144}(it)^7\kappa_3\kappa_4 + \dots\right] + \dots \\
&= 1 + n^{-1/2}r_1(it) + n^{-1}r_2(it) + \dots + n^{-j/2}r_j(it) + \dots
\end{aligned}
$$

where the $r_j$'s are polynomials with real coefficients, of degree $3j$, depending on $\kappa_3, \dots, \kappa_{j+2}$, but not on $n$. Of particular interest are the first two $r_j$'s, defined as:

$$r_1(u) = \frac{1}{6}u^3\kappa_3$$

and

$$r_2(u) = \frac{1}{72}u^6\kappa_3^2 + \frac{1}{24}u^4\kappa_4.$$

Finally, putting it together we get:

$$
\begin{aligned}
\chi_{T_n}(t) &= \exp\left(-\frac{1}{2}t^2\right)a(t) \\
&= \exp\left(-\frac{1}{2}t^2\right)\left(1 + n^{-1/2}r_1(it) + n^{-1}r_2(it) + \dots + n^{-j/2}r_j(it) + \dots\right) \\
&= e^{-t^2/2} + n^{-1/2}e^{-t^2/2}r_1(it) + n^{-1}e^{-t^2/2}r_2(it) + \dots + n^{-j/2}e^{-t^2/2}r_j(it) + \dots
\end{aligned}
$$

**Q.E.D.**■

**Note:** From this we can determine the rate of convergence of the limit in equation (1.1), that is,

$$\chi_{T_n}(t) = e^{-t^2/2} + O(n^{-1/2}).$$

We now note the following forms of the characteristic functions of $T_n$ and the standard normal variable $Z$:

$$\chi_{T_n}(t) = \mathrm{E}\left(e^{itT_n}\right) = \int_{-\infty}^{\infty} e^{itx}d\,\mathrm{P}(T_n \le x) \qquad (1.6)$$

and

$$e^{-t^2/2} = \mathrm{E}\left(e^{itZ}\right) = \int_{-\infty}^{\infty} e^{itx}d\Phi(x). \qquad (1.7)$$

*In order to obtain this expansion for $a(t)$ you will need to apply the Taylor expansion about zero up to at least *seven* terms. The expression appears the way it does because many of the first few terms resolve to zero. This process involves many simple, but tedious, calculations and will be omitted from this discussion.

We can also find, through the use of Fourier-Stieltjes transforms, functions $R_j$ such that:

$$e^{-t^2/2}r_j(it) = \int_{-\infty}^{\infty} e^{itx} dR_j(x) \tag{1.8}$$

(more detail on how these $R_j$ functions are obtained will be discussed momentarily)
Now, substituting equations (1.6), (1.7) and (1.8) into (1.3), we get:

$$\int_{-\infty}^{\infty} e^{itx} d\,\mathrm{P}(T_n \leq x) =$$
$$\int_{-\infty}^{\infty} e^{itx} d\Phi(x) + n^{-1/2}\int_{-\infty}^{\infty} e^{itx} dR_1(x) + n^{-1}\int_{-\infty}^{\infty} e^{itx} dR_2(x) + \cdots + n^{-j/2}\int_{-\infty}^{\infty} e^{itx} dR_j(x)\ldots$$

Which can be rewritten as:

$$\mathrm{P}(T_n \leq x) = \Phi(x) + n^{-1/2}R_1(x) + n^{-1}R_2(x) + \cdots + n^{-j/2}R_j(x)\ldots \tag{1.9}$$

This is the basic form of an Edgeworth expansion. The only thing left to do now is to determine what the functions $R_j$ are.

We will now show that the $R_j$ terms can be expressed as

$$R_j(x) = r_j\left(-\frac{d}{dx}\right)\cdot\Phi(x), \qquad j \geq 1. \tag{1.10}$$

**Proof:**
To find these $R_j$'s we start by looking at how the function $\mathrm{E}(e^{itT_n}) = \int e^{itx} d\Phi(x) = e^{-t^2/2}$ behaves under repeated integration by parts:

$$\begin{aligned}
\mathrm{E}(e^{itT_n}) = e^{-t^2/2} &= \int_{-\infty}^{\infty} e^{itx} d\Phi(x) \\
&= (-it)^{-1}\int_{-\infty}^{\infty} e^{itx} d\left(\frac{d}{dx}\Phi(x)\right) \\
&= (-it)^{-2}\int_{-\infty}^{\infty} e^{itx} d\left(\left(\frac{d}{dx}\right)^2\Phi(x)\right) \\
&\ \vdots \\
&= (-it)^{-j}\int_{-\infty}^{\infty} e^{itx} d\left(\left(\frac{d}{dx}\right)^j\Phi(x)\right) \\
&= (it)^{-j}\int_{-\infty}^{\infty} e^{itx} d\left(\left(-\frac{d}{dx}\right)^j\Phi(x)\right).
\end{aligned}$$

Therefore

$$\int_{-\infty}^{\infty} e^{itx} d\left(\left(-\frac{d}{dx}\right)^j\Phi(x)\right) = e^{-t^2/2}(it)^j.$$

We now need to find a way to change the above equation so that it equals the left hand side of equation (1.8). As it turns out, substituting $\left(-\frac{d}{dx}\right)^j$ with $r_j\left(-\frac{d}{dx}\right)$ will do this, i.e.,

$$\int_{-\infty}^{\infty} e^{itx} d\left(r_j\left(-\frac{d}{dx}\right)\Phi(x)\right) = e^{-t^2/2}r_j(it).$$

(One can confirm the validity of this equation by substituting in the polynomials $r_1$ and $r_2$.)

Therefore, the form of the function $R_j$ is:

$$R_j(x) = r_j\left(-\frac{d}{dx}\right)\Phi(x), \qquad j \geq 1.$$

**Q.E.D.■**

We will now determine the first two functions, $R_1$ and $R_2$ with the help of Hermite polynomials[†]. Hermite polynomials are like shortcuts that can be used when deriving the standard normal distribution function $\Phi(x)$. Using equation (1.10) and the identity

$$\left(-\frac{d}{dx}\right)^j \Phi(x) = -\operatorname{He}_{j-1}(x)\phi(x), \tag{1.11}$$

where $\operatorname{He}_j$ is the $j^{\text{th}}$ Hermite polynomial, and $\phi(x)$ is the standard normal density function, we get the following expression for $R_1$:

$$\begin{aligned} R_1(x) &= r_1\left(-\frac{d}{dx}\right)\Phi(x) \\ &= \frac{1}{6}\kappa_3\left[\left(-\frac{d}{dx}\right)^3\Phi(x)\right] \\ &= \frac{1}{6}\kappa_3\left[-\operatorname{He}_2(x)\phi(x)\right] \\ &= -\frac{1}{6}(x^2-1)\kappa_3\phi(x), \end{aligned}$$

and similarly for $R_2$:

$$R_2(x) = -x\left\{\frac{1}{24}(x^2-3)\kappa_4 + \frac{1}{72}(x^4-10x+15)\kappa_3^2\right\}\phi(x).$$

To make notation a bit simpler we will denote the polynomial that is multiplied by the function $\phi(x)$ in $R_j$ by the polynomial $p_j(x)$, so that we can write

$$R_1(x) = p_1(x)\phi(x) \qquad \text{and} \qquad R_2(x) = p_2(x)\phi(x),$$

where

$$p_1(x) = -\frac{1}{6}(x^2-1)\kappa_3 \qquad \text{and} \qquad p_2(x) = -x\left\{\frac{1}{24}(x^2-3)\kappa_4 + \frac{1}{72}(x^4-10x+15)\kappa_3^2\right\}. \tag{1.12}$$

In general we can say

$$R_j(x) = p_j(x)\phi(x).$$

Using these new forms of $R_j$, we can rewrite equation (1.9) as:

$$\mathrm{P}(T_n \leq x) = \Phi(x) + n^{-1/2}p_1(x)\phi(x) + n^{-1}p_2(x)\phi(x) + \cdots + n^{-j/2}p_j(x)\phi(x)\ldots \tag{1.13}$$

This is known as the *Edgeworth expansion* of the distribution function $\mathrm{P}(T_n \leq x)$.

---

[†]See Appendix B for more details on Hermite polynomials.

**Some comments**

- ⋆ The polynomial $p_1(x)$ in the Edgeworth expansion is known as a skewness correction and $p_2(x)$ is known as a correction for kurtosis and for the secondary effect of skewness.

- ⋆ For the remainder of these notes the term $p_j(x)$ will be used to denote the Edgeworth polynomial terms obtained from *standardized* statistics, while the term $q_j(x)$ will be used to denote the Edgeworth polynomial terms obtained from *studentized* statistics.

- ⋆ For i.i.d. random variables $X_1, X_2, \ldots, X_n$ with mean $\mu$ (estimated by $\bar{X}_n = n^{-1}\sum_{i=1}^{n} X_i$) and variance $\sigma^2 > 0$ (and $\sigma^2$ is estimated by $\hat{\sigma}^2 = n^{-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$), the form of the Edgeworth expansion for the statistic $T_n = \sqrt{n}(\bar{X}_n - \mu)/\hat{\sigma}$, the studentized sum of i.i.d random variables, where $\hat{\sigma}^2$ is a consistent estimate of the asymptotic variance $\sigma$ of $n^{1/2}\bar{X}_n$ is:

$$\mathrm{P}(T_n \leq x) = \Phi(x) + n^{-1/2}q_1(x)\phi(x) + n^{-1}q_2(x)\phi(x) + \cdots + n^{-j/2}q_j(x)\phi(x) + o(n^{-j/2}),$$

  where

$$q_1(x) = \frac{1}{6}(2x^2+1)\kappa_3 \quad \text{and} \quad q_2(x) = x\left\{\frac{1}{12}(x^2-3)\kappa_4 - \frac{1}{18}(x^4+2x-3)\kappa_3^2 - \frac{1}{4}(x^2+3)\right\},$$

  and $\kappa_j$ is the $j^{\text{th}}$ cumulant of $(X_i - \mu)/\sigma$.

- ⋆ The Edgeworth expansions discussed in this section can rarely be considered to be a converging infinite series. Only under certain conditions on the underlying data and the statistic will they converge. The series can, however, be seen as an asymptotic series, i.e., if the series is stopped after a certain number of terms, then the remainder is of smaller order than the last term that was included.

- ⋆ Also, these expansions only hold under certain conditions on the moments and continuity of the underlying data. These conditions are, fortunately, not very strict at all. For further reading see Hall (1992).

## A.2 Edgeworth Expansion for a slightly more general statistic

The extension of the concepts found in the previous section is as follows: The statistic of interest is now broadly defined as one with a limiting standard normal distribution. The statistic can be defined as either the standardized version: $T_n = n^{1/2}(\hat{\theta}_n - \theta)/\sigma$ or as the studentized version: $T_n = n^{1/2}(\hat{\theta}_n - \theta)/\hat{\sigma}$, where $\hat{\sigma}^2$ is a consistent estimate of the asymptotic variance $\sigma^2$ of $n^{1/2}\hat{\theta}_n$. Both versions of the statistic are assumed to have a limiting standard normal distribution.

The Edgeworth expansion of this slightly more general statistic follows in a similar fashion to the expansion found in the previous section.

Let

- ⋆ $\chi_{T_n}(u)$ denote the characteristic function of $T_n$, and

- ⋆ $\kappa_{j,T_n}$ denote the $j^{\text{th}}$ cumulant of $T_n$.

Now,

$$\chi_{T_n}(t) = \mathrm{E}\left(e^{itT_n}\right)$$

$$= \exp\left(\kappa_{1,T_n} it + \frac{1}{2!}\kappa_{2,T_n}(it)^2 + \cdots + \frac{1}{j!}\kappa_{j,T_n}(it)^j + \ldots\right), \quad (1.14)$$

(compare this with equation (1.5) and note the similarities). The cumulant terms, $\kappa_{j,T_n}$, can be expressed in terms of polynomials. These polynomials are of order $n^{-(j-2)/2}$. In general these polynomials have the following form:

$$\kappa_{j,T_n} = n^{-(j-2)/2}\left(k_{j,1} + n^{-1}k_{j,2} + n^{-2}k_{j,3} + \ldots\right), \quad (1.15)$$

where $k_{1,1} = 0$, $k_{2,1} = 1$ and the rest of the $k_{j,i}$ terms will depend on the statistic being calculated‡. The proof of this equation is rather complicated and will not be discussed here. Carrying on with the derivation: Combine equations (1.14) and (1.15) to get:

$$\begin{aligned}
\chi_{T_n}(t) &= \exp\left(\frac{1}{2!}t^2 + n^{-1/2}\left[k_{1,2}it + \frac{1}{3!}k_{3,1}(it)^3\right] + n^{-1}\left[\frac{1}{2!}k_{2,2}(it)^2 + \frac{1}{4!}k_{4,1}(it)^4\right] + \ldots\right) \\
&= e^{-t^2/2}\left(1 + n^{-1/2}\left[k_{1,2}it + \frac{1}{6}k_{3,1}(it)^3\right]\right. \\
&\quad + n^{-1}\left[\tfrac{1}{2}k_{2,2}(it)^2 + \frac{1}{24}k_{4,1}(it)^4 + \tfrac{1}{2}\left\{k_{1,2}it + \frac{1}{6}k_{3,1}(it)^3\right\}^2\right] \\
&\quad \left. + O\left(n^{-3/2}\right)\right) \\
&= e^{-t^2/2}\left(1 + n^{-1/2}\left[k_{1,2}it + \frac{1}{6}k_{3,1}(it)^3\right]\right. \\
&\quad + n^{-1}\left[\tfrac{1}{2}\left(k_{2,2} + k_{1,2}^2\right)(it)^2 + \frac{1}{24}\left(k_{4,1} + 4k_{1,2}k_{3,1}\right)(it)^4 + \frac{1}{72}k_{3,1}^2(it)^6\right] \\
&\quad \left. + O\left(n^{-3/2}\right)\right) \\
&= e^{-t^2/2} + n^{-1/2}\left[k_{1,2}it + \frac{1}{6}k_{3,1}(it)^3\right]e^{-t^2/2} \\
&\quad + n^{-1}\left[\tfrac{1}{2}\left(k_{2,2} + k_{1,2}^2\right)(it)^2 + \frac{1}{24}\left(k_{4,1} + 4k_{1,2}k_{3,1}\right)(it)^4 + \frac{1}{72}k_{3,1}^2(it)^6\right]e^{-t^2/2} \\
&\quad + O\left(n^{-3/2}\right).
\end{aligned} \quad (1.16)$$

The derivation of this is similar to the one used to get equation (1.3).

Equation (1.16) leads us to the general forms of $r_1$ and $r_2$, i.e.,

$$r_1(u) = k_{1,2}u + \frac{1}{6}k_{3,1}(u)^3, \quad (1.17)$$

and

$$r_2(u) = \tfrac{1}{2}\left(k_{2,2} + k_{1,2}^2\right)(u)^2 + \frac{1}{24}\left(k_{4,1} + 4k_{1,2}k_{3,1}\right)(u)^4 + \frac{1}{72}k_{3,1}^2(u)^6. \quad (1.18)$$

We now write $\chi_{T_n}(t)$ as

$$\chi_{T_n}(t) = e^{-t^2/2} + n^{-1/2}r_1(it)e^{-t^2/2} + n^{-1}r_2(it)e^{-t^2/2} + \cdots + n^{-j/2}r_j(it)e^{-t^2/2} + \ldots \quad (1.19)$$

‡If we look at the simple case where the statistic is $T_n$, the sum of i.i.d random variables, as discussed in the previous section, then we can see that cumulants also have the form described by (1.15). This expansion of the cumulants is critically important to the development of an Edgeworth expansion of the distribution of the statistic, because all of the polynomials related to an Edgeworth expansion have components which can be obtained from the above expansion.

In particular, if you look at the forms of (1.5) and (1.14), the form of the individual cumulants is clearly $\kappa_{j,T_n} = n^{-(j-2)/2}\kappa_j \ \forall\ j \geq 2$ where $\kappa_{j,T_n}$ is the $j^{\text{th}}$ cumulant of $T_n$, and $\kappa_j$ is the $j^{\text{th}}$ cumulant of $Y$.

This is the more general form of equation (1.3). Using the same steps already used to get from equation (1.3) to equation (1.9) we have:

$$\begin{aligned} \mathrm{P}(T_n \le x) &= \Phi(x) + n^{-1/2}R_1(x) + n^{-1}R_2(x) + \cdots + n^{-j/2}R_j(x)\ldots \\ &= \Phi(x) + n^{-1/2}p_1(x)\phi(x) + n^{-1}p_2(x)\phi(x) + \cdots + n^{-j/2}p_j(x)\phi(x)\ldots \end{aligned} \tag{1.20}$$

Now, using (1.17), (1.18), (1.10), the identity (1.11) and the fact that $R_j(x) = p_j(x)\phi(x)$, we get the general forms of $p_1$ and $p_2$:

$$\begin{aligned} p_1(x) &= -\left(k_{1,2} + \frac{1}{6}k_{3,1}\,\mathrm{He}_1(x)\right) \\ &= -\left(k_{1,2} + \frac{1}{6}k_{3,1}(x^2-1)\right), \end{aligned}$$

and

$$\begin{aligned} p_2(x) &= -\left(\tfrac{1}{2}(k_{2,2} + k_{1,2}^2)\,\mathrm{He}_1(x) + \frac{1}{24}(k_{4,1} + 4k_{1,2}k_{3,1})\,\mathrm{He}_3(x) + \frac{1}{72}k_{3,1}^2\,\mathrm{He}_5(x)\right) \\ &= -x\left(\tfrac{1}{2}(k_{2,2} + k_{1,2}^2) + \frac{1}{24}(k_{4,1} + 4k_{1,2}k_{3,1})(x^2-3) + \frac{1}{72}k_{3,1}^2(x^4 - 10x + 15)\right). \end{aligned}$$

We now have the somewhat more general form of the Edgeworth expansion up to the first two terms (further terms can also be derived):

$$\mathrm{P}(T_n \le x) = \Phi(x) + n^{-1/2}p_1(x)\phi(x) + n^{-1}p_2(x)\phi(x) + \cdots + n^{-j/2}p_j(x)\phi(x) + o(n^{-j/2}).$$

**Remark**: The fact that $k_{1,2} = 0$ and that $k_{3,1} = \kappa_3$ in the expansion given in (1.13) can be seen considering the general forms of the characteristic function of $T_n$, given in (1.14), and the cumulants of $T_n$, given in (1.15). These general forms can then be compared to the form of the characteristic function for the specific statistic $T_n$, given in (1.5). This leads to the conclusion that $k_{1,2} = 0$ and $k_{3,1} = \kappa_3$ for the expansion of $T_n$.

## A.3 Edgeworth expansions for statistics that satisfy the "smooth function model"

Finally, a general model which can be manipulated to accommodate a large variety of statistics and for which the rigorous derivation of the cumulant expansion presented in equation (1.15) and the Edgeworth expansion presented in equation (1.20) can be determined, will now be presented. The statistics which can be expressed in terms of this general formulation include the mean, the variance, difference between means, differences between variances, ratios of means, ratios of variances, correlations, and so on.

This model, called the "Smooth Function Model", was developed by Bhattacharya and Ghosh (1978) and is stated here in a form adapted from Hall (1992), who focused on using this model with the bootstrap.

***The Smooth Function Model:*** *Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$ be i.i.d. random column d-vectors from some distribution $F_d$ with mean $\boldsymbol{\mu}$. Define the mean of these variables as $\bar{\boldsymbol{X}}_n = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i$. Suppose that we are interested in estimating a parameter $\theta = g(\boldsymbol{\mu})$ with the statistic $\hat{\theta}_n = g(\bar{\boldsymbol{X}}_n)$ and define $(h(\boldsymbol{\mu}))^2$ as the asymptotic variance of $n^{1/2}\hat{\theta}_n$, where h is some known function. The quantity $h(\boldsymbol{\mu})$ is estimated by $h(\bar{\boldsymbol{X}}_n)$.*

*Let $A$ be a smooth function so that $A : \mathbb{R}^d \to \mathbb{R}$ and is defined such that $A(\boldsymbol{\mu}) = 0$. The function $A$ is typically given by:*

$$A(\mathbf{x}) = \frac{g(\mathbf{x}) - g(\boldsymbol{\mu})}{h(\boldsymbol{\mu})} \qquad or \qquad A(\mathbf{x}) = \frac{g(\mathbf{x}) - g(\boldsymbol{\mu})}{h(\mathbf{x})}.$$

As mentioned, this smooth function model can be used to express a large number of different scenarios, but requires an almost artful choice of the column vectors $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$ and the functions $g$ and $h$. If we define a sample from a population with mean $m$ and variance $\beta^2$ then Hall (1992) illustrates on pp. 52 and 53 how to choose the values of $d$, $\boldsymbol{\mu}$, $\boldsymbol{X}_i$, $g$ and $h$ so that one can find appropriate expressions for $\theta$, $\hat{\theta}_n$, $h(\boldsymbol{\mu})$ and $h(\bar{\boldsymbol{X}}_n)$ when estimating either $m$ or $\beta^2$.

Let the $i^{\text{th}}$ element of a $d$-vector $\mathbf{x}$ be denoted by $x^{(i)}$, and define the vector $\boldsymbol{Z}$ as

$$\boldsymbol{Z} = n^{1/2}(\bar{\boldsymbol{X}}_n - \boldsymbol{\mu}).$$

The Taylor expansion of the statistic $T_n$, defined at the beginning of Section A.2, is

$$\begin{aligned} T_n &= n^{1/2}(\hat{\theta}_n - \theta)/\sigma \\ &= n^{1/2}A(\bar{\boldsymbol{X}}_n) \\ &= T_{nr} + O_p(n^{-r/2}), \qquad r \geq 1, \end{aligned} \tag{1.21}$$

since $A(\boldsymbol{\mu}) = 0$ and $\boldsymbol{Z} = O_p(1)$. We define $T_{nr}$ as

$$T_{nr} = \sum_{i=1}^{d} a_i Z^{(i)} + n^{-1/2} \sum_{i_1=1}^{d} \sum_{i_2=1}^{d} a_{i_1 i_2} Z^{(i_1)} Z^{(i_2)} + \ldots$$

$$+ n^{-(r-1)/2} \sum_{i_1=1}^{d} \cdots \sum_{i_r=1}^{d} a_{i_1 \ldots i_r} Z^{(i_1)} \ldots Z^{(i_r)},$$

where $Z^{(1)}, Z^{(2)}, \ldots$ are the elements of the $d$-vector $\boldsymbol{Z}$, and

$$a_{i_1, i_2, \ldots, i_j} = \left. \frac{\partial^j}{\partial x^{(i_1)} \ldots \partial x^{(i_j)}} A(\mathbf{x}) \right|_{\mathbf{x}=\boldsymbol{\mu}}. \tag{1.22}$$

According to Theorem 2.1 stated in Hall (1992), the cumulants of the statistic $T_{nr}$ can be expanded as

$$\kappa_{j,n,r} = n^{-(j-2)/2} \left( k_{j,1} + n^{-1} k_{j,2} + n^{-2} k_{j,3} + \ldots \right), \tag{1.23}$$

where the constants $k_{j,l}$ depend only on the $a$ coefficients, the moments of $\boldsymbol{X}$ (up to the $jr^{\text{th}}$ moment), and on $r$. The series only has a finite number of nonzero terms. Under the assumption that the first $j$ moments of the $O_p(n^{-r/2})$ terms in (1.21) are all of order $n^{-r/2}$, then the $j^{\text{th}}$ cumulant of $T_n$ equals the $j^{\text{th}}$ cumulant of $T_{nr}$ plus a term of order $n^{-r/2}$ (which can be made arbitrarily small by choosing a large value for $r$).

Through another set of lengthy calculations we can also express the first three cumulants of $T_n$ as

$$\kappa_{1,n} = \mathrm{E}(T_n) = n^{-1/2} A_1 + O(n^{-1}),$$

$$\kappa_{2,n} = \mathrm{E}(T_n^2) - (\mathrm{E}(T_n))^2 = \sigma^2 + O(n^{-1}),$$

and

$$\kappa_{3,n} = \mathrm{E}(T_n^3) - 3\,\mathrm{E}(T_n^2)\,\mathrm{E}(T_n) + 2\left(\mathrm{E}(T_n)\right)^3 = n^{-1/2} A_2 + O(n^{-1}),$$

where

$$\mathrm{E}(T_n) = n^{-1/2}\frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d} a_{ij}\mu_{ij} + O(n^{-1}),$$

$$\mathrm{E}(T_n^2) = \sum_{i=1}^{d}\sum_{j=1}^{d} a_i a_j \mu_{ij} + O(n^{-1}),$$

$$\begin{aligned}\mathrm{E}(T_n^3) = n^{-1/2}\Big\{ & \textstyle\sum_{i=1}^{d}\sum_{j=1}^{d}\sum_{k=1}^{d} a_i a_j a_k \mu_{ijk} \\ & +\tfrac{3}{2}\textstyle\sum_{i=1}^{d}\sum_{j=1}^{d}\sum_{k=1}^{d}\sum_{l=1}^{d} a_i a_j a_{kl}\left(\mu_{ij}\mu_{kl} + \mu_{ik}\mu_{jl} + \mu_{il}\mu_{jk}\right)\Big\} \\ & +O(n^{-1}),\end{aligned}$$

$$\sigma^2 = \sum_{i=1}^{d}\sum_{j=1}^{d} a_i a_j \mu_{ij}, \tag{1.24}$$

$$A_1 = \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d} a_{ij}\mu_{ij}, \tag{1.25}$$

$$A_2 = \sum_{i=1}^{d}\sum_{j=1}^{d}\sum_{k=1}^{d} a_i a_j a_k \mu_{ijk} + 3\sum_{i=1}^{d}\sum_{j=1}^{d}\sum_{k=1}^{d}\sum_{l=1}^{d} a_i a_j a_{kl}\mu_{ik}\mu_{jl}, \tag{1.26}$$

and, finally, where

$$\mu_{i_1\ldots i_j} = \mathrm{E}\left((\boldsymbol{X}-\boldsymbol{\mu})^{(i_1)}\cdots(\boldsymbol{X}-\boldsymbol{\mu})^{(i_j)}\right), \quad j \geq 1. \tag{1.27}$$

If the $\sigma$ term, defined in (1.24) as the standard error of $T_n$, is not equal to one, then the statistic $T_n$ is redefined such that it has unit variance, i.e., we redefine it as follows:

$$T_n = n^{1/2}A(\bar{\boldsymbol{X}}_n)/\sigma.$$

The Edgeworth expansion of the distribution of $T_n$ can now be stated. This theorem is also stated as Theorem 2.2 in Hall (1992).

Before the theorem is stated, first define a $d$-vector $\boldsymbol{t} = (t^{(1)}, \ldots, t^{(d)})^T$, and let

$$\|\boldsymbol{t}\| = \sqrt{\left(t^{(1)}\right)^2 + \cdots + \left(t^{(d)}\right)^2}$$

and

$$\chi(\boldsymbol{t}) = \mathrm{E}\left(\exp\left\{i\sum_{j=1}^{d} t^{(j)}X^{(j)}\right\}\right).$$

**Theorem 1.** ***Edgeworth expansion for the smooth function model:***
*Assume that the function $A$ has $j+2$ continuous derivatives in a neighbourhood of $\boldsymbol{\mu} = \mathrm{E}(\boldsymbol{X})$, that $A(\boldsymbol{\mu}) = 0$, that $\mathrm{E}(\|\boldsymbol{X}\|^{j+2}) < \infty$, and that the characteristic function $\chi$ of $\boldsymbol{X}$ satisfies*

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

*(i.e., Cramer's condition). Define $a_{i_1\ldots i_r}$ and $\mu_{i_1\ldots i_r}$ as in equation (1.22) and (1.27), and $\sigma$, $A_1$ and $A_2$ as in equations (1.24), (1.25) and (1.26). Suppose that $\sigma > 0$. Then for $j \geq 1$,*

$$\mathrm{P}\left(n^{1/2}A(\bar{\boldsymbol{X}}_n)/\sigma \leq x\right) \quad = \quad \Phi(x) + n^{-1/2}p_1(x)\phi(x) + \ldots$$

$$+n^{-j/2}p_j(x)\phi(x) + o(n^{-j/2}), \tag{1.28}$$

*uniformly in $x$, where $p_j$ is a polynomial of degree at most $3j-1$, odd for even $j$ and even for odd $j$, with coefficients depending on the moments of $\boldsymbol{X}$ up to order $j+2$. In particular,*

$$p_1(x) = -\left\{A_1\sigma^{-1} + \frac{1}{6}A_2\sigma^{-3}(x^2-1)\right\}.$$

It is shown in Hall (1992) that it is possible to obtain the same polynomials, $p_1$ and $p_2$, which were derived in Sections A.1 and A.2 by applying this theorem.

## A.4 Cornish-Fisher expansion

We will now look at the "inverse" version of the Edgeworth expansion, known as the *Cornish-Fisher expansion*. The Cornish-Fisher expansion is an expansion for the quantile of a distribution of a statistic $T_n$. Note: In this discussion we assume that $T_n$ is a standardized statistic with a limiting standard normal distribution. Formally we define this quantile, $w(\alpha)$, as

$$w(\alpha) = \inf\{x : \mathrm{P}(T_n \leq x) \geq \alpha\}$$

i.e., the smallest value $x$ such that the probability that $T_n$ is smaller than $x$ is greater than or equal to $\alpha$. But, for simplicity's sake, we will simply define $w(\alpha)$ as:

$$\mathrm{P}(T_n \leq w(\alpha)) = \alpha,$$

for a given $0 < \alpha < 1$.

If we look at the Edgeworth expansion

$$\mathrm{P}(T_n \leq x) = \Phi(x) + n^{-1/2}p_1(x)\phi(x) + n^{-1}p_2(x)\phi(x) + \cdots + n^{-j/2}p_j(x)\phi(x) + o(n^{-j/2}),$$

we can 'invert' this expression so that it becomes an expansion for the quantile. The general form of this expansion is:

$$w(\alpha) = z(\alpha) + n^{-1/2}p_1^{CF}(z(\alpha)) + \cdots + n^{-j/2}p_j^{CF}(z(\alpha)) + o(n^{-j/2}),$$

where $z(\alpha) = \Phi^{-1}(\alpha)$, i.e., the $\alpha$ level quantile of the standard normal distribution, and $p_i^{CF}(u)$ are the Cornish-Fisher polynomials. These expansions are interpreted as an asymptotic series uniformly in $\varepsilon < \alpha < 1-\varepsilon$ for any $\varepsilon \in (0, \frac{1}{2})$.

The derivation of these Cornish-Fisher polynomials will not be discussed. One only has to note that they are functions of the Edgeworth polynomials $p_i(u)$, and the relationship, for the first two polynomials for the derivations discussed in Sections A.1 and A.2 are:

$$p_1^{CF}(x) = -p_1(x) \qquad \text{and} \qquad p_2^{CF}(x) = p_1(x)\left[\frac{d}{dx}p_1(x)\right] - \tfrac{1}{2}xp_1(x)^2 - p_2(x).$$

The formal theorem provided in Hall (1992) for the smooth function model will now be repeated here.

**Theorem 2.** ***Cornish-Fisher expansion for the smooth function model:***
*Assume the conditions of Theorem 1 on the function $A$ and the distribution of $\boldsymbol{X}$. Define*

$$w(\alpha) = \inf\left\{x : \mathrm{P}\left(n^{1/2}A(\bar{\boldsymbol{X}}_n) \leq x\right) \geq \alpha\right\}.$$

*Let $p_1, \dots, p_j$ denote the polynomials which appear in (1.28), and define $p_1^{CF}, \dots, p_j^{CF}$ as polynomials which satisfy the equality*

$$\Phi\left\{z(\alpha) + \sum_{j\geq 1} n^{-j/2} p_j^{CF}(z(\alpha))\right\}$$
$$+ \sum_{i\geq 1} n^{-i/2} p_i \left\{z(\alpha) + \sum_{j\geq 1} n^{-j/2} p_j^{CF}(z(\alpha))\right\}$$
$$\times \phi \left\{z(\alpha) + \sum_{j\geq 1} n^{-j/2} p_j^{CF}(z(\alpha))\right\} = \alpha, \quad 0 < \alpha < 1.$$

*Then,*

$$w(\alpha) = z(\alpha) + n^{-1/2} p_1^{CF}(z(\alpha)) + \cdots + n^{-j/2} p_j^{CF}(z(\alpha)) + o(n^{-j/2}),$$

*uniformly in $\varepsilon < \alpha < 1 - \varepsilon$ for each $\varepsilon > 0$.*

**Remark:** Similar expansions can be obtained for the quantiles of the distribution of studentized statistics.

## A.5 Edgeworth and Cornish-Fisher expansions for bootstrap distributions

This section will briefly state the Edgeworth and Cornish-Fisher expansions as they are applied to bootstrap estimated distributions. To make these expressions as general as possible we will consider the expansions given in Chung and Lee (2001) for the $m$-out-of-$n$ bootstrap statistics. These formulations are more general because they can be applied to both the traditional ($m = n$) and the $m$-out-of-$n$ ($m < n$) bootstrap resampling schemes.

We will limit our discussion to the parameters, statistics and distribution functions which are covered by the smooth function model. Let $X_1, X_2, \dots, X_n$ be data generated from the distribution function $F$ with mean $\mu$, and let $\theta = g(\mu)$ be estimated by $\hat{\theta}_n = g(\bar{X}_n)$ where $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. The asymptotic variance of $n^{1/2}\hat{\theta}_n$ is given by $h(\mu)$ and it is estimated by $h(\bar{X}_n)$. Define the following two distribution functions for the standardized and studentized statistics respectively:

$$H_n(x) = \mathrm{P}\left(n^{1/2}(\hat{\theta}_n - \theta)/h(\mu) \leq x\right),$$

and

$$K_n(x) = \mathrm{P}\left(n^{1/2}(\hat{\theta}_n - \theta)/h(\bar{X}_n) \leq x\right).$$

Next, define the bootstrap statistics as $\hat{\theta}_m^* = g(\bar{X}_m^*)$, $h(\bar{X}_m^*)$ and $\bar{X}_m^* = \frac{1}{m}\sum_{i=1}^m X_i^*$. The bootstrap versions of the standardized and studentized distribution functions are respectively:

$$\hat{H}_m(x) = \mathrm{P}^*\left(n^{1/2}(\hat{\theta}_m^* - \hat{\theta}_n)/h(\bar{X}_n) \leq x\right),$$

and

$$\hat{K}_m(x) = \mathrm{P}^*\left(n^{1/2}(\hat{\theta}_m^* - \hat{\theta}_n)/h(\bar{X}_m^*) \leq x\right).$$

The lemmas stated in Chung and Lee (2001) are simple extensions of the expansions given in Hall (1992) which were discussed in the previous sections.

**Lemma 1.** ***m-out-of-n bootstrap Edgeworth expansion:***
*Suppose $m = O(n)$ and $m \to \infty$ as $n \to \infty$. Assume that the functions $g$ and $h$ (defined in Section A.3) are continuously differentiable up to a sufficiently high order in an open neighbourhood of $\mu$, that the characteristic function $\chi$ of $X$ satisfies Cramer's condition, and that there exist sufficiently many moments of $X$. Then, for some large positive integer $j$,*

$$\begin{aligned}\hat{H}_m(x) &= \Phi(x) + m^{-1/2}\hat{p}_1(x)\phi(x) + \dots \\ &\quad + m^{-j/2}\hat{p}_j(x)\phi(x) + O_p(m^{-(j+1)/2}),\end{aligned}$$

*and*

$$\begin{aligned}\hat{K}_m(x) &= \Phi(x) + m^{-1/2}\hat{q}_1(x)\phi(x) + \dots \\ &\quad + m^{-j/2}\hat{q}_j(x)\phi(x) + O_p(m^{-(j+1)/2}),\end{aligned}$$

*uniformly in $x$, where $\hat{p}_i$ and $\hat{q}_i$ are obtained by applying the plug-in principle to population elements in the definitions of $p_i$ and $q_i$ respectively, and $O_p$ refers to an unconditional probability.*

**Lemma 2.** ***m-out-of-n bootstrap Cornish-Fisher expansions***
*Under the conditions of Lemma 1, we have that*

$$\hat{w}_m(\alpha) = z(\alpha) + m^{-1/2}\hat{p}_1^{CF}(z(\alpha)) + \dots + m^{-j/2}\hat{p}_j^{CF}(z(\alpha)) + o_p(m^{-j/2}),$$

*and*

$$\hat{v}_m(\alpha) = z(\alpha) + m^{-1/2}\hat{q}_1^{CF}(z(\alpha)) + \dots + m^{-j/2}\hat{q}_j^{CF}(z(\alpha)) + o_p(m^{-j/2}),$$

*uniformly in $\varepsilon < \alpha < 1 - \varepsilon$ for any $\varepsilon \in (0, \frac{1}{2})$, where the $\hat{q}_i^{CF}$ and $\hat{q}_i^{CF}$polynomials are obtained by applying the plug-in principle on the population elements in $p_i^{CF}$ and $q_i^{CF}$ respectively, and where $o_p$ refers to an unconditional probability.*

# Appendix B

# Hermite Polynomials

Hermite polynomials are tools used to help obtain the higher order derivatives of the standard normal density or distribution function. We define the $r^{\text{th}}$ Hermite polynomial as:

$$\begin{aligned} He_r(x) &= (-1)^r \frac{\frac{d^r}{dx^r}\phi(x)}{\phi(x)} \\ &= \sum_{j=0}^{\lfloor r/2 \rfloor} (-1)^j \,\mathrm{OF}(2j) \binom{r}{2j} x^{r-2j} \\ &= xHe_{r-1}(x) - (r-1)He_{r-2}(x), \end{aligned}$$

where $\phi(x)$ is the standard normal density function, $\lfloor x \rfloor$ denotes the largest integer smaller than $x$, and $\mathrm{OF}(x)$ is called the *odd factorial*, and is defined as

$$\mathrm{OF}(x) = x!/\left[2^{\frac{x}{2}}\left(\frac{x}{2}\right)!\right].$$

The following table is a summary of the first eleven Hermite polynomials:

| **Hermite polynomial** |
|---|
| $\mathrm{He}_0(x) = 1$ |
| $\mathrm{He}_1(x) = x$ |
| $\mathrm{He}_2(x) = x^2 - 1$ |
| $\mathrm{He}_3(x) = x^3 - 3x$ |
| $\mathrm{He}_4(x) = x^4 - 6x^2 + 3$ |
| $\mathrm{He}_5(x) = x^5 - 10x^3 + 15x$ |
| $\mathrm{He}_6(x) = x^6 - 15x^4 + 45x^2 - 15$ |
| $\mathrm{He}_7(x) = x^7 - 21x^5 + 105x^3 - 105x$ |
| $\mathrm{He}_8(x) = x^8 - 28x^6 + 210x^4 - 420x^2 + 105$ |
| $\mathrm{He}_9(x) = x^9 - 36x^7 + 378x^5 - 1260x^3 + 945x$ |
| $\mathrm{He}_{10}(x) = x^{10} - 45x^8 + 630x^6 + 4725x^2 - 945$ |

**Table B.1:** *Table of the first eleven Hermite polynomials.*

# Appendix C

# Big-$O$ and little-$o$ notation

## C.1 Stochastic convergence in the real and bootstrap world

We begin by first defining *convergence almost surely*, *convergence in probability*, *$\mathcal{L}^p$ convergence* and *convergence in distribution* in both the real world and the bootstrap world:
Let $X_1, X_2, \ldots$ and $X$ be random variables. Let $R_1^*, R_2^*, \ldots,$ and $R^*$ be bootstrap random variables.

- ⋆ **Convergence almost surely**: We say $X_n$ converges almost surely to $X$ (or $X_n \stackrel{a.s.}{\to} X$) if

$$\mathrm{P}(\lim_{n\to\infty} X_n = X) = 1.$$

- ⋆ **$\mathcal{L}^p$ convergence**: We say that $X_n$ converges in $\mathcal{L}^p$ to $X$ (or $X_n \stackrel{\mathcal{L}^p}{\to} X$) if

$$\lim_{n\to\infty} \mathrm{E}(|X_n - X|^p) = 0.$$

- ⋆ **Convergence in probability**: We say $X_n$ converges in probability to $X$ (or $X_n \stackrel{p}{\to} X$) if for every $\epsilon > 0$,

$$\lim_{n\to\infty} \mathrm{P}(|X_n - X| \geq \epsilon) = 0.$$

- ⋆ **Convergence in distribution (or convergence in law)**: Let $F_{X_n}$ be the distribution function of $X_n$, $n = 1, 2, \ldots,$ and $F_X$ be the distribution function of $X$. We say $X_n$ converges in distribution to $X$ (or $X_n \stackrel{\mathcal{D}}{\to} X$) if

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x),$$

for all $x$ at which $F_X(x)$ is continuous. Convergence in distribution is also known as *weak convergence*.

As far as the bootstrap world is concerned, we define the following two concepts:

- ⋆ **Bootstrap convergence in probability**: We say $R_n^*$, a statistic in the bootstrap world, converges in conditional (given $X_1, X_2, \ldots, X_n$) probability to $R^*$ (or $R_n^* \stackrel{p^*}{\to} R^*$ *a.s.*) if for every $\epsilon > 0$,

$$\lim_{n\to\infty} \mathrm{P}^*(|R_n^* - R^*| \geq \epsilon) = 0, \qquad a.s.$$

⋆ **Bootstrap convergence in distribution**: Let $F_{R_n^*}$ be the distribution function of $R_n^*$, $n = 1, 2, \ldots$, and $F_{R^*}$ be the distribution function of $R^*$. We say $R_n^*$ converges in conditional (given $X_1, X_2, \ldots, X_n$) distribution to $R^*$ (or $R_n^* \stackrel{\mathcal{D}^*}{\rightarrow} R^*$ *a.s.*) if

$$\lim_{n\to\infty} F_{R_n^*}(x) = F_{R^*}(x) \qquad a.s.,$$

for all $x$ at which $F_{R^*}(x)$ is continuous.

## C.2 Big-$O$ and little-$o$ notation

### Deterministic Big-$O$ and little-$o$

Let $\{a_n\}$ and $\{b_n\}$ be two sequences of numbers.

⋆ **Big-$O$**: We say that the sequence $\{a_n\}$ is "Big-$O$" $\{b_n\}$, or $a_n = O(b_n)$, if

$$\limsup_{n\to\infty} |a_n/b_n| \le K,$$

or alternatively, if

$$|a_n| \le K|b_n|, \qquad \text{for all sufficiently large } n,$$

where K is some finite constant.

⋆ **little-$o$**: We say that the term $a_n$ is "little-$o$" $b_n$, or $a_n = o(b_n)$, if

$$\lim_{n\to\infty} a_n/b_n = 0.$$

### Stochastic Big-$O$ and little-$o$

Let $\{X_n\}$ and $\{Y_n\}$ be two sequences of random variables.

⋆ **Big-$O$ a.s.**: We say that the sequence $\{X_n\}$ is "Big-$O$ *almost surely*" $\{Y_n\}$, or $X_n = O(Y_n)$ *a.s.*, if $X_n = O(Y_n)$ for almost all sequences $\{X_1, X_2, \ldots\}$ and $\{Y_1, Y_2, \ldots\}$.

⋆ **little-$o$ a.s.**: We say that the term $X_n$ is "little-$o$ *almost surely*" $Y_n$, or $X_n = o(Y_n)$ *a.s.*, if

$$\frac{X_n}{Y_n} \xrightarrow{a.s.} 0.$$

⋆ **Big-$O_p$**: We say that the sequence $\{X_n\}$ is "Big-$O_p$" $\{Y_n\}$, or $X_n = O_p(Y_n)$, if for any $\epsilon > 0$ there exists a constant $K_\epsilon > 0$, and an integer $n_\epsilon$, such that

$$\mathrm{P}\left(\frac{|X_n|}{|Y_n|} \le K_\epsilon\right) \ge 1 - \epsilon, \qquad \forall\, n \ge n_\epsilon.$$

⋆ **little-$o_p$**: We say that the term $X_n$ is "little-$o_p$" $Y_n$, or $X_n = o_p(Y_n)$, if

$$\frac{|X_n|}{|Y_n|} \xrightarrow{p} 0.$$

In the bootstrap world we can also define the following:
Let $\{X_n^*\}$ and $\{Y_n^*\}$ be two sequences of random variables in the bootstrap world.

⋆ **Big-$O_{p^*}$**: We say that the sequence $\{X_n^*\}$ is "Big-$O_{p^*}$" $\{Y_n^*\}$, or $X_n^* = O_{p^*}(Y_n^*)$, if for any $\epsilon > 0$ there exists a constant in the bootstrap world (i.e., a random variable in the real world) $\hat{K}_\epsilon > 0$, and an integer $n_\epsilon$, such that

$$\mathrm{P}^*\left(\frac{|X_n^*|}{|Y_n^*|} \leq \hat{K}_\epsilon\right) \geq 1 - \epsilon, \quad a.s., \qquad \forall\, n \geq n_\epsilon.$$

⋆ **little-$o_{p^*}$**: We say that the term $X_n^*$ is "little-$o_{p^*}$" $Y_n^*$, or $X_n^* = o_{p^*}(Y_n^*)$, if

$$\frac{|X_n^*|}{|Y_n^*|} \xrightarrow{p^*} 0, \quad a.s.$$

# Appendix D

# Expected values of functions of sample moments

The following expected values presented in this appendix are used in the discussions in Chapter 5. We will attempt to express all of the expected values in terms of population moments. Where possible intermediate steps are also provided.

We will use the variable $Y_i = X_i - \mathrm{E}(X_i)$ so that the variable $Y_i$ is centered on zero. This addition has no real effect on the results since the sample moments of order 2 and higher are location invariant, but it helps to have this condition in some of the expressions.

## D.1 Expressions for the expected values of sample moments

We will now provide expressions for the expected values of the following products of sample moments: $\hat{\mu}_2$, $\hat{\mu}_2^2$, $\hat{\mu}_2^5$, $\hat{\mu}_3$, $\hat{\mu}_3^2$, $\hat{\mu}_2\hat{\mu}_3$, $\hat{\mu}_2\hat{\mu}_4^2$, $\hat{\mu}_2^2\hat{\mu}_6$, $\hat{\mu}_2^3\hat{\mu}_4$, $\hat{\mu}_2^2\hat{\mu}_3^2$, $\hat{\mu}_2\hat{\mu}_3\hat{\mu}_5$, $\hat{\mu}_3^2\hat{\mu}_4$, $\bar{Y}_n(\hat{\mu}_3-\mu_3)$, $\bar{Y}_n(\hat{\mu}_2-\mu_2)$, $\bar{Y}_n(\hat{\mu}_3-\mu_3)(\hat{\mu}_2-\mu_2)$, $\bar{Y}_n(\hat{\mu}_2-\mu_2)^2$, $\bar{Y}_n(\hat{\mu}_2-\mu_2)^2(\hat{\mu}_3-\mu_3)$, and $\bar{Y}_n(\hat{\mu}_2-\mu_2)^3$. These expected values of products of sample moments are expressed in terms of the expected values of sums of random variables. The expected values of the sums of random variables are then resolved and the terms are collected. The expressions of the expected values of the sums of random variables are given in the next section of this appendix. Clearly, many steps have been ommitted, but these steps largely involve the collection of terms, and as such are considered trivially easy (but incredibly tedious). These calculations are available on request from the candidate.

1.

$$\begin{aligned}
\mathrm{E}(\hat{\mu}_2) &= \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n}\left(Y_i-\bar{Y}_n\right)^2\right] \\
&= \frac{1}{n}\mathrm{E}\left(\sum_{i=1}^{n}Y_i^2\right)-\frac{1}{n^2}\mathrm{E}\left(\left(\sum_{i=1}^{n}Y_i\right)^2\right) \\
&= \mu_2+\frac{1}{n}\left\{-\mu_2\right\}.
\end{aligned}$$

---

2.

$$\mathrm{E}\left(\hat{\mu}_2^2\right) = \mathrm{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n}\left(Y_i-\bar{Y}_n\right)^2\right)^2\right]$$

$$
\begin{aligned}
&= \frac{1}{n^2}\mathrm{E}\left(\left(\sum_{i=1}^{n} Y_i^2\right)^2\right) - \frac{2}{n^3}\mathrm{E}\left(\sum_{i=1}^{n} Y_i^2\left(\sum_{i=1}^{n} Y_i\right)^2\right) + \frac{1}{n^4}\mathrm{E}\left(\left(\sum_{i=1}^{n} Y_i\right)^4\right)\\
&= \mu_2^2 + \frac{1}{n}\left\{\mu_4 - 3\mu_2^2\right\} + \frac{1}{n^2}\left\{-2\mu_4 + 5\mu_2^2\right\} + \frac{1}{n^3}\left\{\mu_4 - 3\mu_2^2\right\}.
\end{aligned}
$$

---

3.

$$
\begin{aligned}
\mathrm{E}\left(\hat{\mu}_2^5\right) &= \mathrm{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \bar{Y}_n\right)^2\right)^5\right]\\
&= \frac{1}{n^5}\mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i^2\right)^5\right] - \frac{5}{n^6}\mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^2\left(\sum_{i=1}^{n} Y_i^2\right)^4\right] + \frac{10}{n^7}\mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^4\left(\sum_{i=1}^{n} Y_i^2\right)^3\right]\\
&\quad -\frac{10}{n^8}\mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^6\left(\sum_{i=1}^{n} Y_i^2\right)^2\right] + \frac{5}{n^9}\mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^8\sum_{i=1}^{n} Y_i^2\right] - \frac{1}{n^{10}}\mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^{10}\right]\\
&= \mu_2^5 + \frac{1}{n}\{-15\mu_2^5 + 10\mu_2^3\mu_4\} + \frac{1}{n^2}\{115\mu_2^5 - 110\mu_2^3\mu_4 + 15\mu_2\mu_4^2\\
&\quad +10\mu_2^2\mu_6 - 60\mu_2^2\mu_3^2\} + O(n^{-3}).
\end{aligned}
$$

---

4.

$$
\begin{aligned}
\mathrm{E}\left(\hat{\mu}_3\right) &= \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \bar{Y}_n\right)^3\right]\\
&= \frac{1}{n}\mathrm{E}\left(\sum_{i=1}^{n} Y_i^3\right) - \frac{3}{n^2}\mathrm{E}\left(\sum_{i=1}^{n} Y_i^2\sum_{i=1}^{n} Y_i\right) + \frac{2}{n^3}\mathrm{E}\left(\left(\sum_{i=1}^{n} Y_i\right)^3\right)\\
&= \mu_3 + \frac{1}{n}\left\{-3\mu_3\right\} + \frac{1}{n^2}\left\{2\mu_3\right\}.
\end{aligned}
$$

---

5.

$$
\begin{aligned}
\mathrm{E}\left(\hat{\mu}_3^2\right) &= \mathrm{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \bar{Y}_n\right)^3\right)^2\right]\\
&= \frac{1}{n^2}\mathrm{E}\left(\left(\sum_{i=1}^{n} Y_i^3\right)^2\right) - \frac{6}{n^3}\mathrm{E}\left(\sum_{i=1}^{n} Y_i\sum_{i=1}^{n} Y_i^2\sum_{i=1}^{n} Y_i^3\right) + \frac{4}{n^4}\mathrm{E}\left(\sum_{i=1}^{n} Y_i^3\left(\sum_{i=1}^{n} Y_i\right)^3\right)\\
&\quad +\frac{9}{n^4}\mathrm{E}\left(\left(\sum_{i=1}^{n} Y_i\right)^2\left(\sum_{i=1}^{n} Y_i^2\right)^2\right) - \frac{12}{n^5}\mathrm{E}\left(\sum_{i=1}^{n} Y_i^2\left(\sum_{i=1}^{n} Y_i\right)^4\right) + \frac{4}{n^6}\mathrm{E}\left(\left(\sum_{i=1}^{n} Y_i\right)^6\right)\\
&= \mu_3^2 + \frac{1}{n}\left\{\mu_6 + 9\mu_2^3 - 7\mu_3^2 - 6\mu_2\mu_4\right\} +\\
&\quad +\frac{1}{n^2}\{-6\mu_6 + 45\mu_2\mu_4 + 28\mu_3^2 - 63\mu_2^3\} +
\end{aligned}
$$

$$+\frac{1}{n^3}\left\{13\mu_6+26\mu_3^2-123\mu_2\mu_4+186\mu_2^3\right\}+$$
$$+\frac{1}{n^4}\left\{-12\mu_6+144\mu_2\mu_4+88\mu_3^2-252\mu_2^3\right\}+$$
$$+\frac{1}{n^5}\{4\mu_6-60\mu_2\mu_4-40\mu_3^2+120\mu_2^3\}.$$

---

6.

$$\begin{aligned}
\mathrm{E}\left(\hat{\mu}_2\hat{\mu}_3\right) &= \mathrm{E}\left[\frac{1}{n^2}\sum_{i=1}^{n}\left(Y_i-\bar{Y}_n\right)^2\sum_{i=1}^{n}\left(Y_i-\bar{Y}_n\right)^3\right] \\
&= \frac{1}{n^2}\mathrm{E}\left(\sum_{i=1}^{n}Y_i^2\sum_{i=1}^{n}Y_i^3\right)-\frac{1}{n^3}\mathrm{E}\left(\left(\sum_{i=1}^{n}Y_i\right)^2\sum_{i=1}^{n}Y_i^3\right)-\frac{3}{n^3}\mathrm{E}\left(\sum_{i=1}^{n}Y_i\left(\sum_{i=1}^{n}Y_i^2\right)^2\right)+ \\
&\quad +\frac{3}{n^4}\mathrm{E}\left(\left(\sum_{i=1}^{n}Y_i\right)^3\sum_{i=1}^{n}Y_i^2\right)+\frac{2}{n^4}\mathrm{E}\left(\sum_{i=1}^{n}Y_i^2\left(\sum_{i=1}^{n}Y_i\right)^3\right)-\frac{2}{n^5}\mathrm{E}\left(\left(\sum_{i=1}^{n}Y_i\right)^5\right) \\
&= \mu_2\mu_3+\frac{1}{n}\left\{\mu_5-8\mu_2\mu_3\right\}+\frac{1}{n^2}\left\{-4\mu_5+27\mu_2\mu_3\right\}+\frac{1}{n^3}\left\{5\mu_5-40\mu_2\mu_3\right\} \\
&\quad +\frac{1}{n^4}\left\{-2\mu_5+20\mu_2\mu_3\right\}.
\end{aligned}$$

---

7.

$$\begin{aligned}
\mathrm{E}\left(\hat{\mu}_2\hat{\mu}_4^2\right) &= \mathrm{E}\left[\frac{1}{n^3}\sum_{i=1}^{n}\left(Y_i-\bar{Y}_n\right)^2\left(\sum_{i=1}^{n}\left(Y_i-\bar{Y}_n\right)^4\right)^2\right] \\
&= \frac{1}{n^3}\mathrm{E}\left[\sum_{i=1}^{n}Y_i^2\left(\sum_{i=1}^{n}Y_i^4\right)^2\right]-\frac{8}{n^4}\mathrm{E}\left[\sum_{i=1}^{n}Y_i\sum_{i=1}^{n}Y_i^2\sum_{i=1}^{n}Y_i^3\sum_{i=1}^{n}Y_i^4\right] \\
&\quad +\frac{12}{n^5}\mathrm{E}\left[\left(\sum_{i=1}^{n}Y_i\right)^2\left(\sum_{i=1}^{n}Y_i^2\right)^2\sum_{i=1}^{n}Y_i^4\right]+\frac{16}{n^5}\mathrm{E}\left[\left(\sum_{i=1}^{n}Y_i\right)^2\sum_{i=1}^{n}Y_i^2\left(\sum_{i=1}^{n}Y_i^3\right)^2\right] \\
&\quad -\frac{48}{n^6}\mathrm{E}\left[\left(\sum_{i=1}^{n}Y_i\right)^3\left(\sum_{i=1}^{n}Y_i^2\right)^2\sum_{i=1}^{n}Y_i^3\right]+\frac{72}{n^7}\mathrm{E}\left[\left(\sum_{i=1}^{n}Y_i\right)^5\sum_{i=1}^{n}Y_i^2\sum_{i=1}^{n}Y_i^3\right] \\
&\quad +\frac{36}{n^7}\mathrm{E}\left[\left(\sum_{i=1}^{n}Y_i\right)^4\left(\sum_{i=1}^{n}Y_i^2\right)^3\right]-\frac{1}{n^4}\mathrm{E}\left[\left(\sum_{i=1}^{n}Y_i\right)^2\left(\sum_{i=1}^{n}Y_i^4\right)^2\right] \\
&\quad +\frac{8}{n^5}\mathrm{E}\left[\left(\sum_{i=1}^{n}Y_i\right)^3\sum_{i=1}^{n}Y_i^3\sum_{i=1}^{n}Y_i^4\right]-\frac{18}{n^6}\mathrm{E}\left[\left(\sum_{i=1}^{n}Y_i\right)^4\sum_{i=1}^{n}Y_i^2\sum_{i=1}^{n}Y_i^4\right] \\
&\quad +\frac{6}{n^7}\mathrm{E}\left[\left(\sum_{i=1}^{n}Y_i\right)^6\sum_{i=1}^{n}Y_i^4\right]-\frac{16}{n^6}\mathrm{E}\left[\left(\sum_{i=1}^{n}Y_i\right)^4\left(\sum_{i=1}^{n}Y_i^3\right)^2\right]
\end{aligned}$$

$$
\begin{aligned}
& -\frac{24}{n^8}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^7 \sum_{i=1}^n Y_i^3\right] - \frac{72}{n^8}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i^2\right)^2 \sum_{i=1}^n Y_i^6\right] \\
& +\frac{45}{n^9}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^8 \sum_{i=1}^n Y_i^2\right] - \frac{9}{n^{10}}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^{10}\right] \\
= \;& \mu_2\mu_4^2 + \frac{1}{n}\{-12\mu_2\mu_4^2 + 2\mu_4\mu_6 + \mu_2\mu_8 - 8\mu_3^2\mu_4 - 8\mu_2\mu_3\mu_5 + 12\mu_2^3\mu_4 \\
& +16\mu_2^2\mu_3^2\} + O(n^{-2}).
\end{aligned}
$$

---

8.

$$
\begin{aligned}
\mathrm{E}\left(\hat{\mu}_2^2\hat{\mu}_6\right) \;=\; & \mathrm{E}\left[\frac{1}{n^3}\left(\sum_{i=1}^n (Y_i-\bar{Y}_n)^2\right)^2 \sum_{i=1}^n (Y_i-\bar{Y}_n)^6\right] \\
= \;& \frac{1}{n^3}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i^2\right)^2 \sum_{i=1}^n Y_i^6\right] - \frac{6}{n^4}\,\mathrm{E}\left[\sum_{i=1}^n Y_i \left(\sum_{i=1}^n Y_i^2\right)^2 \sum_{i=1}^n Y_i^5\right] \\
& -\frac{2}{n^4}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^2 \sum_{i=1}^n Y_i^2 \sum_{i=1}^n Y_i^6\right] + \frac{15}{n^5}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^2 \left(\sum_{i=1}^n Y_i^2\right)^2 \sum_{i=1}^n Y_i^4\right] \\
& +\frac{12}{n^5}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^3 \sum_{i=1}^n Y_i^2 \sum_{i=1}^n Y_i^5\right] + \frac{1}{n^5}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^4 \sum_{i=1}^n Y_i^6\right] \\
& -\frac{20}{n^6}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^3 \left(\sum_{i=1}^n Y_i^2\right)^2 \sum_{i=1}^n Y_i^3\right] - \frac{30}{n^6}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^4 \sum_{i=1}^n Y_i^2 \sum_{i=1}^n Y_i^4\right] \\
& -\frac{6}{n^6}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^5 \sum_{i=1}^n Y_i^5\right] + \frac{15}{n^7}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^4 \left(\sum_{i=1}^n Y_i^2\right)^3\right] \\
& +\frac{40}{n^7}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^5 \sum_{i=1}^n Y_i^2 \sum_{i=1}^n Y_i^3\right] + \frac{15}{n^7}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^6 \sum_{i=1}^n Y_i^4\right] \\
& -\frac{35}{n^8}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^6 \left(\sum_{i=1}^n Y_i^2\right)^2\right] - \frac{20}{n^8}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^7 \sum_{i=1}^n Y_i^3\right] \\
& +\frac{25}{n^9}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^8 \sum_{i=1}^n Y_i^2\right] - \frac{5}{n^{10}}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^{10}\right] \\
= \;& \mu_2^2\mu_6 + \frac{1}{n}\{\mu_4\mu_6 + 2\mu_2\mu_8 - 11\mu_2^2\mu_6 - 12\mu_2\mu_3\mu_5 + 15\mu_2^3\mu_4\} + O(n^{-2}).
\end{aligned}
$$

---

9.

$$
\mathrm{E}\left(\hat{\mu}_2^3\hat{\mu}_4\right) \;=\; \mathrm{E}\left[\frac{1}{n^4}\left(\sum_{i=1}^n (Y_i-\bar{Y}_n)^2\right)^3 \sum_{i=1}^n (Y_i-\bar{Y}_n)^4\right]
$$

$$
\begin{aligned}
&= \frac{1}{n^4}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i^2\right)^3 \sum_{i=1}^n Y_i^4\right] - \frac{4}{n^5}\,\mathrm{E}\left[\sum_{i=1}^n Y_i \left(\sum_{i=1}^n Y_i^2\right)^3 \sum_{i=1}^n Y_i^3\right] \\
&\quad + \frac{6}{n^6}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^2 \left(\sum_{i=1}^n Y_i^2\right)^4\right] - \frac{21}{n^7}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^4 \left(\sum_{i=1}^n Y_i^2\right)^3\right] \\
&\quad - \frac{3}{n^5}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^2 \left(\sum_{i=1}^n Y_i^2\right)^2 \sum_{i=1}^n Y_i^4\right] + \frac{12}{n^6}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^3 \left(\sum_{i=1}^n Y_i^2\right)^2 \sum_{i=1}^n Y_i^3\right] \\
&\quad + \frac{27}{n^8}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^6 \left(\sum_{i=1}^n Y_i^2\right)^2\right] + \frac{3}{n^6}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^4 \sum_{i=1}^n Y_i^2 \sum_{i=1}^n Y_i^4\right] \\
&\quad - \frac{12}{n^7}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^5 \sum_{i=1}^n Y_i^2 \sum_{i=1}^n Y_i^3\right] - \frac{15}{n^9}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^8 \sum_{i=1}^n Y_i^2\right] \\
&\quad - \frac{1}{n^7}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^6 \sum_{i=1}^n Y_i^4\right] + \frac{4}{n^8}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^7 \sum_{i=1}^n Y_i^3\right] \\
&\quad + \frac{3}{n^{10}}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^{10}\right] \\
&= \mu_2^3\mu_4 + \frac{1}{n}\{-13\mu_2^3\mu_4 + 3\mu_2^2\mu_6 + 3\mu_2\mu_4^2 - 12\mu_2^2\mu_3^2 + 6\mu_2^5\} \\
&\quad + \frac{1}{n^2}\{228\mu_2^2\mu_3^2 + 158\mu_2^3\mu_4 - 18\mu_3^2\mu_4 - 30\mu_2\mu_4^2 - 48\mu_2\mu_3\mu_5 \\
&\quad - 30\mu_2^2\mu_6 + 3\mu_2\mu_8 + 4\mu_4\mu_6 - 123\mu_2^5\} + O(n^{-3}).
\end{aligned}
$$

---

10.

$$
\begin{aligned}
\mathrm{E}\left(\hat{\mu}_2^2\hat{\mu}_3^2\right) &= \mathrm{E}\left[\frac{1}{n^4}\left(\sum_{i=1}^n (Y_i-\bar{Y}_n)^2\right)^2 \left(\sum_{i=1}^n (Y_i-\bar{Y}_n)^3\right)^2\right] \\
&= \frac{1}{n^4}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i^2\right)^2 \left(\sum_{i=1}^n Y_i^3\right)^2\right] - \frac{2}{n^5}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^2 \sum_{i=1}^n Y_i^2 \left(\sum_{i=1}^n Y_i^3\right)^2\right] \\
&\quad - \frac{6}{n^5}\,\mathrm{E}\left[\sum_{i=1}^n Y_i \left(\sum_{i=1}^n Y_i^2\right)^3 \sum_{i=1}^n Y_i^3\right] + \frac{1}{n^6}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^4 \left(\sum_{i=1}^n Y_i^3\right)^2\right] \\
&\quad + \frac{16}{n^6}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^3 \left(\sum_{i=1}^n Y_i^2\right)^2 \sum_{i=1}^n Y_i^3\right] + \frac{9}{n^6}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^2 \left(\sum_{i=1}^n Y_i^2\right)^4\right] \\
&\quad - \frac{14}{n^7}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^5 \sum_{i=1}^n Y_i^2 \sum_{i=1}^n Y_i^3\right] - \frac{30}{n^7}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^4 \left(\sum_{i=1}^n Y_i^2\right)^3\right] \\
&\quad + \frac{4}{n^8}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^7 \sum_{i=1}^n Y_i^3\right] + \frac{37}{n^8}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^6 \left(\sum_{i=1}^n Y_i^2\right)^2\right]
\end{aligned}
$$

$$
\begin{aligned}
& -\frac{20}{n^9}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^8 \sum_{i=1}^n Y_i^2\right] + \frac{4}{n^{10}}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^{10}\right] \\
= \;& \mu_2^2\mu_3^2 + \frac{1}{n}\{\mu_2^2\mu_6 + 4\mu_2\mu_3\mu_5 + \mu_3^2\mu_4 - 26\mu_2^2\mu_3^2 - 6\mu_2^3\mu_4 + 9\mu_2^5\} \\
& + \frac{1}{n^2}\{2\mu_2\mu_8 + \mu_4\mu_6 - 23\mu_2^2\mu_6 + 2\mu_3\mu_7 + 2\mu_5^2 - 74\mu_2\mu_3\mu_5 \\
& -31\mu_3^2\mu_4 + 354\mu_2^2\mu_3^2 + 174\mu_2^3\mu_4 - 22\mu_2\mu_4^2 - 180\mu_2^5\} + O(n^{-3}).
\end{aligned}
$$

---

11.

$$
\begin{aligned}
\mathrm{E}\left(\hat\mu_2\hat\mu_3\hat\mu_5\right) = \;& \mathrm{E}\left[\frac{1}{n^3}\sum_{i=1}^n (Y_i-\bar Y_n)^2 \sum_{i=1}^n (Y_i-\bar Y_n)^3 \sum_{i=1}^n (Y_i-\bar Y_n)^5\right] \\
= \;& \frac{1}{n^3}\,\mathrm{E}\left[\sum_{i=1}^n Y_i^2 \sum_{i=1}^n Y_i^3 \sum_{i=1}^n Y_i^5\right] - \frac{5}{n^4}\,\mathrm{E}\left[\sum_{i=1}^n Y_i \sum_{i=1}^n Y_i^2 \sum_{i=1}^n Y_i^3 \sum_{i=1}^n Y_i^4\right] \\
& + \frac{10}{n^5}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^2 \sum_{i=1}^n Y_i^2 \left(\sum_{i=1}^n Y_i^3\right)^2\right] - \frac{40}{n^6}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^3 \left(\sum_{i=1}^n Y_i^2\right)^2 \sum_{i=1}^n Y_i^3\right] \\
& - \frac{3}{n^4}\,\mathrm{E}\left[\sum_{i=1}^n Y_i \left(\sum_{i=1}^n Y_i^2\right)^2 \sum_{i=1}^n Y_i^5\right] + \frac{15}{n^5}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^2 \left(\sum_{i=1}^n Y_i^2\right)^2 \sum_{i=1}^n Y_i^4\right] \\
& + \frac{30}{n^7}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^4 \left(\sum_{i=1}^n Y_i^2\right)^3\right] + \frac{5}{n^5}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^3 \sum_{i=1}^n Y_i^2 \sum_{i=1}^n Y_i^5\right] \\
& + \frac{64}{n^7}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^5 \sum_{i=1}^n Y_i^2 \sum_{i=1}^n Y_i^3\right] - \frac{62}{n^8}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^6 \left(\sum_{i=1}^n Y_i^2\right)^2\right] \\
& + \frac{40}{n^9}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^8 \sum_{i=1}^n Y_i^2\right] - \frac{1}{n^4}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^2 \sum_{i=1}^n Y_i^3 \sum_{i=1}^n Y_i^5\right] \\
& + \frac{5}{n^5}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^3 \sum_{i=1}^n Y_i^3 \sum_{i=1}^n Y_i^4\right] - \frac{10}{n^6}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^4 \left(\sum_{i=1}^n Y_i^3\right)^2\right] \\
& - \frac{25}{n^6}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^4 \sum_{i=1}^n Y_i^2 \sum_{i=1}^n Y_i^4\right] - \frac{2}{n^6}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^5 \sum_{i=1}^n Y_i^5\right] \\
& + \frac{10}{n^7}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^6 \sum_{i=1}^n Y_i^4\right] - \frac{24}{n^8}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^7 \sum_{i=1}^n Y_i^3\right] \\
& - \frac{8}{n^{10}}\,\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^{10}\right] \\
= \;& \mu_2\mu_3\mu_5 + \frac{1}{n}\{\mu_3\mu_7 + \mu_2\mu_8 - 15\mu_2\mu_3\mu_5 + 15\mu_2^3\mu_4 + \mu_5^2 - 5\mu_3^2\mu_4 \\
& -5\mu_2\mu_4^2 + 10\mu_2^2\mu_3^2 - 3\mu_2^2\mu_6\} + O(n^{-2}).
\end{aligned}
$$

---

12.

$$\begin{aligned}
\mathrm{E}\left(\hat{\mu}_3^2\hat{\mu}_4\right) &= \mathrm{E}\left[\frac{1}{n^3}\left(\sum_{i=1}^{n}(Y_i-\bar{Y}_n)^3\right)^2\sum_{i=1}^{n}(Y_i-\bar{Y}_n)^4\right] \\
&= \frac{1}{n^3}\mathrm{E}\left[\left(\sum_{i=1}^{n}Y_i^3\right)^2\sum_{i=1}^{n}Y_i^4\right]-\frac{4}{n^4}\mathrm{E}\left[\sum_{i=1}^{n}Y_i\left(\sum_{i=1}^{n}Y_i^3\right)^3\right] \\
&\quad+\frac{30}{n^5}\mathrm{E}\left[\left(\sum_{i=1}^{n}Y_i\right)^2\sum_{i=1}^{n}Y_i^2\left(\sum_{i=1}^{n}Y_i^3\right)^2\right]-\frac{19}{n^6}\mathrm{E}\left[\left(\sum_{i=1}^{n}Y_i\right)^4\left(\sum_{i=1}^{n}Y_i^3\right)^2\right] \\
&\quad-\frac{6}{n^4}\mathrm{E}\left[\sum_{i=1}^{n}Y_i\sum_{i=1}^{n}Y_i^2\sum_{i=1}^{n}Y_i^3\sum_{i=1}^{n}Y_i^4\right]-\frac{72}{n^6}\mathrm{E}\left[\left(\sum_{i=1}^{n}Y_i\right)^3\left(\sum_{i=1}^{n}Y_i^2\right)^2\sum_{i=1}^{n}Y_i^3\right] \\
&\quad+\frac{90}{n^7}\mathrm{E}\left[\left(\sum_{i=1}^{n}Y_i\right)^5\sum_{i=1}^{n}Y_i^2\sum_{i=1}^{n}Y_i^3\right]+\frac{4}{n^5}\mathrm{E}\left[\left(\sum_{i=1}^{n}Y_i\right)^3\sum_{i=1}^{n}Y_i^3\sum_{i=1}^{n}Y_i^4\right] \\
&\quad-\frac{28}{n^8}\mathrm{E}\left[\left(\sum_{i=1}^{n}Y_i\right)^7\sum_{i=1}^{n}Y_i^3\right]+\frac{9}{n^5}\mathrm{E}\left[\left(\sum_{i=1}^{n}Y_i\right)^2\left(\sum_{i=1}^{n}Y_i^2\right)^2\sum_{i=1}^{n}Y_i^4\right] \\
&\quad+\frac{54}{n^7}\mathrm{E}\left[\left(\sum_{i=1}^{n}Y_i\right)^4\left(\sum_{i=1}^{n}Y_i^2\right)^3\right]-\frac{99}{n^8}\mathrm{E}\left[\left(\sum_{i=1}^{n}Y_i\right)^6\left(\sum_{i=1}^{n}Y_i^2\right)^2\right] \\
&\quad-\frac{12}{n^6}\mathrm{E}\left[\left(\sum_{i=1}^{n}Y_i\right)^4\sum_{i=1}^{n}Y_i^2\sum_{i=1}^{n}Y_i^4\right]+\frac{60}{n^9}\mathrm{E}\left[\left(\sum_{i=1}^{n}Y_i\right)^8\sum_{i=1}^{n}Y_i^2\right] \\
&\quad+\frac{4}{n^7}\mathrm{E}\left[\left(\sum_{i=1}^{n}Y_i\right)^6\sum_{i=1}^{n}Y_i^4\right]-\frac{12}{n^{10}}\mathrm{E}\left[\left(\sum_{i=1}^{n}Y_i\right)^{10}\right] \\
&= \mu_3^2\mu_4+\frac{1}{n}\{-21\mu_3^2\mu_4+2\mu_3\mu_7+\mu_4\mu_6+30\mu_2^2\mu_3^2-18\mu_2\mu_4^2 \\
&\quad-18\mu_2\mu_3\mu_5+9\mu_2^3\mu_4\}+O(n^{-2}).
\end{aligned}$$

---

13.

$$\begin{aligned}
\mathrm{E}\left(\bar{Y}_n(\hat{\mu}_3-\mu_3)\right) &= \mathrm{E}\left(\bar{Y}_n\hat{\mu}_3\right) \\
&= \frac{1}{n^2}\mathrm{E}\left[\sum_{i=1}^{n}Y_i\sum_{i=1}^{n}Y_i^3\right]-\frac{3}{n^3}\mathrm{E}\left[\left(\sum_{i=1}^{n}Y_i\right)^2\sum_{i=1}^{n}Y_i^2\right]+\frac{2}{n^4}\mathrm{E}\left[\left(\sum_{i=1}^{n}Y_i\right)^4\right] \\
&= \frac{1}{n}\left\{\mu_4-3\mu_2^2\right\}+\frac{1}{n^2}\left\{-3\mu_4+9\mu_2^2\right\}+O(n^{-3}).
\end{aligned}$$

---

14.

$$\mathrm{E}\left(\bar{Y}_n(\hat{\mu}_2-\mu_2)\right) = \mathrm{E}(\bar{Y}_n\hat{\mu}_2)$$

$$
\begin{aligned}
&= \frac{1}{n^2}\,\mathrm{E}\left[\sum_{i=1}^{n} Y_i \sum_{i=1}^{n} Y_i^2\right] - \frac{1}{n^3}\,\mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^3\right] \\
&= \frac{1}{n}\left\{\mu_3\right\} + \frac{1}{n^2}\left\{-\mu_3\right\}.
\end{aligned}
$$

---

15.

$$
\begin{aligned}
\mathrm{E}\left(\bar{Y}_n(\hat{\mu}_3-\mu_3)(\hat{\mu}_2-\mu_2)\right) &= \mathrm{E}\left(\bar{Y}_n\hat{\mu}_3\hat{\mu}_2\right) - \mu_2\,\mathrm{E}\left(\bar{Y}_n\hat{\mu}_3\right) - \mu_3\,\mathrm{E}\left(\bar{Y}_n\hat{\mu}_2\right) \\
&= \frac{1}{n^3}\,\mathrm{E}\left[\sum_{i=1}^{n} Y_i^3 \sum_{i=1}^{n} Y_i^2 \sum_{i=1}^{n} Y_i\right] - \frac{1}{n^4}\,\mathrm{E}\left[\sum_{i=1}^{n} Y_i^3 \left(\sum_{i=1}^{n} Y_i\right)^3\right] \\
&\quad - \frac{3}{n^4}\,\mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^2 \left(\sum_{i=1}^{n} Y_i^2\right)^2\right] + \frac{5}{n^5}\,\mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^4 \sum_{i=1}^{n} Y_i^2\right] \\
&\quad - \frac{2}{n^6}\,\mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^6\right] - \frac{\mu_2}{n^2}\,\mathrm{E}\left[\sum_{i=1}^{n} Y_i \sum_{i=1}^{n} Y_i^3\right] \\
&\quad + \frac{3\mu_2}{n^3}\,\mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^2 \sum_{i=1}^{n} Y_i^2\right] - \frac{2\mu_2}{n^4}\,\mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^4\right] \\
&\quad - \frac{\mu_3}{n^2}\,\mathrm{E}\left[\sum_{i=1}^{n} Y_i \sum_{i=1}^{n} Y_i^2\right] + \frac{\mu_3}{n^3}\,\mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^3\right] \\
&= \frac{1}{n^2}\left\{\mu_6 - 10\mu_2\mu_4 - 7\mu_3^2 + 15\mu_2^3\right\} + O(n^{-3}).
\end{aligned}
$$

---

16.

$$
\begin{aligned}
\mathrm{E}\left(\bar{Y}_n(\hat{\mu}_2-\mu_2)^2\right) &= \mathrm{E}\left(\bar{Y}_n\hat{\mu}_2^2\right) - 2\mu_2\,\mathrm{E}\left(\bar{Y}_n\hat{\mu}_2\right) \\
&= \frac{1}{n^3}\,\mathrm{E}\left[\sum_{i=1}^{n} Y_i \left(\sum_{i=1}^{n} Y_i^2\right)^2\right] - \frac{2}{n^4}\,\mathrm{E}\left[\sum_{i=1}^{n} Y_i^2 \left(\sum_{i=1}^{n} Y_i\right)^3\right] \\
&\quad + \frac{1}{n^5}\,\mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^5\right] - \frac{2\mu_2}{n^2}\,\mathrm{E}\left[\sum_{i=1}^{n} Y_i \sum_{i=1}^{n} Y_i^2\right] \\
&\quad + \frac{2\mu_2}{n^3}\,\mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^3\right] \\
&= \frac{1}{n^2}\left\{\mu_5 - 8\mu_2\mu_3\right\} + O(n^{-3}).
\end{aligned}
$$

---

17.

$$
\mathrm{E}\left(\bar{Y}_n(\hat{\mu}_2-\mu_2)^2(\hat{\mu}_3-\mu_3)\right) = \mathrm{E}\left(\bar{Y}_n\hat{\mu}_2^2\hat{\mu}_3\right) - 2\mu_2\,\mathrm{E}\left(\bar{Y}_n\hat{\mu}_2\hat{\mu}_3\right) + \mu_2^2\,\mathrm{E}\left(\bar{Y}_n\hat{\mu}_3\right)
$$

$$
\begin{aligned}
& -\mu_3 \,\mathrm{E}\left(\bar{Y}_n\hat{\mu}_2^2\right) + 2\mu_2\mu_3 \,\mathrm{E}\left(\bar{Y}_n\hat{\mu}_2\right) \\
= \; & \frac{1}{n^4}\mathrm{E}\left[\sum_{i=1}^n Y_i^3\left(\sum_{i=1}^n Y_i^2\right)^2\sum_{i=1}^n Y_i\right] - \frac{2}{n^5}\mathrm{E}\left[\sum_{i=1}^n Y_i^3\sum_{i=1}^n Y_i^2\left(\sum_{i=1}^n Y_i\right)^3\right] \\
& + \frac{1}{n^6}\mathrm{E}\left[\sum_{i=1}^n Y_i^3\left(\sum_{i=1}^n Y_i\right)^5\right] - \frac{3}{n^5}\mathrm{E}\left[\left(\sum_{i=1}^n Y_i^2\right)^3\left(\sum_{i=1}^n Y_i\right)^2\right] \\
& + \frac{8}{n^6}\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^4\left(\sum_{i=1}^n Y_i^2\right)^2\right] - \frac{7}{n^7}\mathrm{E}\left[\sum_{i=1}^n Y_i^2\left(\sum_{i=1}^n Y_i\right)^6\right] \\
& + \frac{2}{n^8}\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^8\right] - \frac{2\mu_2}{n^3}\mathrm{E}\left[\sum_{i=1}^n Y_i^3\sum_{i=1}^n Y_i^2\sum_{i=1}^n Y_i\right] \\
& + \frac{2\mu_2}{n^4}\mathrm{E}\left[\sum_{i=1}^n Y_i^3\left(\sum_{i=1}^n Y_i\right)^3\right] + \frac{6\mu_2}{n^4}\mathrm{E}\left[\left(\sum_{i=1}^n Y_i^2\right)^2\left(\sum_{i=1}^n Y_i\right)^2\right] \\
& - \frac{10\mu_2}{n^5}\mathrm{E}\left[\sum_{i=1}^n Y_i^2\left(\sum_{i=1}^n Y_i\right)^4\right] + \frac{4\mu_2}{n^6}\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^6\right] \\
& + \frac{\mu_2^2}{n^2}\mathrm{E}\left[\sum_{i=1}^n Y_i\sum_{i=1}^n Y_i^3\right] - \frac{3\mu_2^2}{n^3}\mathrm{E}\left[\sum_{i=1}^n Y_i^2\left(\sum_{i=1}^n Y_i\right)^2\right] \\
& + \frac{2\mu_2^2}{n^4}\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^4\right] - \frac{\mu_3}{n^3}\mathrm{E}\left[\sum_{i=1}^n Y_i\left(\sum_{i=1}^n Y_i^2\right)^2\right] \\
& + \frac{2\mu_3}{n^4}\mathrm{E}\left[\sum_{i=1}^n Y_i^2\left(\sum_{i=1}^n Y_i\right)^3\right] - \frac{\mu_3}{n^5}\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^5\right] \\
& + \frac{2\mu_2\mu_3}{n^2}\mathrm{E}\left[\sum_{i=1}^n Y_i\sum_{i=1}^n Y_i^2\right] - \frac{2\mu_2\mu_3}{n^3}\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^3\right] \\
= \; & \frac{1}{n^2}\left\{2\mu_3\mu_5 + \mu_4^2 - 4\mu_2^2\mu_4 - 8\mu_2\mu_3^2 \right. \\
& \left. +3\mu_2^4\right\} + O(n^{-3}).
\end{aligned}
$$

---

18.

$$
\begin{aligned}
\mathrm{E}\left(\bar{Y}_n(\hat{\mu}_2-\mu_2)^3\right) &= \mathrm{E}\left(\bar{Y}_n\hat{\mu}_2^3\right) - 3\mu_2\,\mathrm{E}\left(\bar{Y}_n\hat{\mu}_2^2\right) + 3\mu_2^2\,\mathrm{E}\left(\bar{Y}_n\hat{\mu}_2\right) \\
&= \frac{1}{n^4}\mathrm{E}\left[\left(\sum_{i=1}^n Y_i^2\right)^3\sum_{i=1}^n Y_i\right] - \frac{3}{n^5}\mathrm{E}\left[\left(\sum_{i=1}^n Y_i^2\right)^2\left(\sum_{i=1}^n Y_i\right)^3\right] \\
&\quad + \frac{3}{n^6}\mathrm{E}\left[\sum_{i=1}^n Y_i^2\left(\sum_{i=1}^n Y_i\right)^5\right] - \frac{1}{n^7}\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^7\right].
\end{aligned}
$$

$$
\begin{aligned}
&- \frac{3\mu_2}{n^3} \mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i^2\right)^2 \sum_{i=1}^{n} Y_i\right] + \frac{6\mu_2}{n^4} \mathrm{E}\left[\sum_{i=1}^{n} Y_i^2 \left(\sum_{i=1}^{n} Y_i\right)^3\right] \\
&- \frac{3\mu_2}{n^5} \mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^5\right] + \frac{3\mu_2^2}{n^2} \mathrm{E}\left[\sum_{i=1}^{n} Y_i^2 \sum_{i=1}^{n} Y_i\right] \\
&- \frac{3\mu_2^2}{n^3} \mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^3\right] \\
= \;& \frac{1}{n^2}\left\{3\mu_3\mu_4 - 3\mu_2^2\mu_3\right\} + O(n^{-3}).
\end{aligned}
$$

---

## D.2 Expressions for the expected value of sums of random variables

This section will present the expressions for the expected value of the sums of random variables which were utilised in the previous section.

Let $Y_1, Y_2, \ldots, Y_n$ be i.i.d random variables with $\mathrm{E}(Y_i) = 0$ for $i = 1, 2, \ldots, n$. Further, define $\mu_k = \mathrm{E}(Y_i^k)$ for $i = 1, 2, \ldots, n$ and $k = 2, 3, \ldots$.

1.

$$
\begin{aligned}
\mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^2\right] &= \mathrm{E}\left(\sum_{i=1}^{n} Y_i^2 + \sum_{i=1}^{n}\sum_{j\neq i} Y_i Y_j^2\right) \\
&= n\mu_2
\end{aligned}
$$

---

2.

$$
\begin{aligned}
\mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^3\right] &= \mathrm{E}\left(\sum_{i=1}^{n} Y_i^3 + 3\sum_{i=1}^{n}\sum_{j\neq i} Y_i Y_j^2 + \sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i \\ k\neq j}} Y_i Y_j Y_k \cdot\right) \\
&= n\mu_3
\end{aligned}
$$

---

3.

$$
\begin{aligned}
\mathrm{E}\left[\sum_{i=1}^{n} Y_i \sum_{i=1}^{n} Y_i^2\right] &= \mathrm{E}\left(\sum_{i=1}^{n} Y_i^3 + \sum_{i=1}^{n}\sum_{j\neq i} Y_i Y_j^2\right) \\
&= n\mu_3
\end{aligned}
$$

---

4.

$$\begin{aligned} \mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^4\right] &= \mathrm{E}\left(\sum_{i=1}^{n} Y_i^4 + 3\sum_{i=1}^{n}\sum_{j\neq i} Y_i^2 Y_j^2 + 6\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_i Y_j Y_k^2 \right. \\ &\quad \left. + 4\sum_{i=1}^{n}\sum_{j\neq i} Y_i Y_j^3 + \sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}} Y_i Y_j Y_k Y_l\right) \\ &= n\mu_4 + 3n^2\mu_2^2 - 3n\mu_2^2 \end{aligned}$$

---

5.

$$\begin{aligned} \mathrm{E}\left[\sum_{i=1}^{n} Y_i \sum_{i=1}^{n} Y_i^3\right] &= \mathrm{E}\left(\sum_{i=1}^{n} Y_i^4 + \sum_{i=1}^{n}\sum_{j\neq i} Y_i Y_j^3\right) \\ &= n\mu_4 \end{aligned}$$

---

6.

$$\begin{aligned} \mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^2 \sum_{i=1}^{n} Y_i^2\right] &= \mathrm{E}\left(\sum_{i=1}^{n} Y_i^4 + \sum_{i=1}^{n}\sum_{j\neq i} Y_i^2 Y_j^2 + 2\sum_{i=1}^{n}\sum_{j\neq i} Y_i^3 Y_j + \sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_i Y_j Y_k^2\right) \\ &= n\mu_4 + n^2\mu_2^2 - n\mu_2^2 \end{aligned}$$

---

7.

$$\begin{aligned} \mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i^2\right)^2\right] &= \mathrm{E}\left(\sum_{i=1}^{n} Y_i^4 + \sum_{i=1}^{n}\sum_{j\neq i} Y_i^2 Y_j^2\right) \\ &= n\mu_4 + n^2\mu_2^2 - n\mu_2^2 \end{aligned}$$

---

8.

$$\begin{aligned} \mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^5\right] &= \mathrm{E}\left(\sum_{i=1}^{n} Y_i^5 + 5\sum_{i=1}^{n}\sum_{j\neq i} Y_i Y_j^4 + 10\sum_{i=1}^{n}\sum_{j\neq i} Y_i^2 Y_j^3 \right. \\ &\quad + 10\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_i Y_j Y_k^3 + 15\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_i Y_j^2 Y_k^2 \end{aligned}$$

$$\left. + 10\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}} Y_iY_jY_kY_l^2 + \sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}}\sum_{\substack{m\neq i\\m\neq j\\m\neq k\\m\neq l}} Y_iY_jY_kY_lY_m \right)$$

$$= n\mu_5 + 10n^2\mu_2\mu_3 - 10n\mu_2\mu_3$$

---

9.

$$\begin{aligned} \mathrm{E}\left[\left(\sum_{i=1}^{n}Y_i\right)^2\sum_{i=1}^{n}Y_i^3\right] &= \mathrm{E}\left(\sum_{i=1}^{n}Y_i^5 + \sum_{i=1}^{n}\sum_{j\neq i}Y_i^2Y_j^3 + 2\sum_{i=1}^{n}\sum_{j\neq i}Y_iY_j^4\right. \\ &\quad \left. + 2\sum_{i=1}^{n}\sum_{j\neq i}Y_iY_j^4 + \sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_iY_jY_k^3\right) \\ &= n\mu_5 + n^2\mu_2\mu_3 - n\mu_2\mu_3 \end{aligned}$$

---

10.

$$\begin{aligned} \mathrm{E}\left[\left(\sum_{i=1}^{n}Y_i\right)^3\sum_{i=1}^{n}Y_i^2\right] &= \mathrm{E}\left(\sum_{i=1}^{n}Y_i^5 + 3\sum_{i=1}^{n}\sum_{j\neq i}Y_iY_j^4 + 4\sum_{i=1}^{n}\sum_{j\neq i}Y_i^2Y_j^3\right. \\ &\quad + 3\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_iY_j^2Y_k^2 + 3\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_iY_jY_k^3 \\ &\quad \left. + \sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}} Y_iY_jY_kY_l^2\right) \\ &= n\mu_5 + 4n^2\mu_2\mu_3 - 4n\mu_2\mu_3 \end{aligned}$$

---

11.

$$\begin{aligned} \mathrm{E}\left[\sum_{i=1}^{n}Y_i\left(\sum_{i=1}^{n}Y_i^2\right)^2\right] &= \mathrm{E}\left(\sum_{i=1}^{n}Y_i^5 + \sum_{i=1}^{n}\sum_{j\neq i}Y_iY_j^4 + 2\sum_{i=1}^{n}\sum_{j\neq i}Y_i^2Y_j^3 + \sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_iY_j^2Y_k^2\right) \\ &= n\mu_5 + 2n^2\mu_2\mu_3 - 2n\mu_2\mu_3 \end{aligned}$$

---

12.

$$
\begin{aligned}
\mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^6\right] &= \mathrm{E}\left(\sum_{i=1}^{n} Y_i^6 + 6\sum_{i=1}^{n}\sum_{j\neq i} Y_iY_j^5 + 15\sum_{i=1}^{n}\sum_{j\neq i} Y_i^2Y_j^4\right. \\
&\quad + 10\sum_{i=1}^{n}\sum_{j\neq i} Y_i^3Y_j^3 + 15\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_iY_jY_k^4 + 60\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_iY_j^2Y_k^3 \\
&\quad + 20\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}} Y_iY_jY_kY_l^3 + 15\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_i^2Y_j^2Y_k^2 \\
&\quad + 45\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}} Y_iY_jY_k^2Y_l^2 + 15\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}}\sum_{\substack{m\neq i\\m\neq j\\m\neq k\\m\neq l}} Y_iY_jY_kY_lY_m^2 \\
&\quad \left. + \sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}}\sum_{\substack{m\neq i\\m\neq j\\m\neq k\\m\neq l}}\sum_{\substack{o\neq i\\o\neq j\\o\neq k\\o\neq l\\o\neq m}} Y_iY_jY_kY_lY_mY_o\right) \\
&= n\mu_6 + 15n^2\mu_2\mu_4 - 15n\mu_2\mu_4 + 10n^2\mu_3^2 - 10n\mu_3^2 + 15n^3\mu_2^3 \\
&\quad -45n^2\mu_2^3 + 30n\mu_2^3
\end{aligned}
$$

---

13.

$$
\begin{aligned}
\mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^2 \sum_{i=1}^{n} Y_i^4\right] &= \mathrm{E}\left(\sum_{i=1}^{n} Y_i^6 + \sum_{i=1}^{n}\sum_{j\neq i} Y_i^2Y_j^4 + 2\sum_{i=1}^{n}\sum_{j\neq i} Y_iY_j^5 + \sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_iY_jY_k^4\right) \\
&= n\mu_6 + n^2\mu_2\mu_4 - n\mu_2\mu_4
\end{aligned}
$$

---

14.

$$
\begin{aligned}
\mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^4 \sum_{i=1}^{n} Y_i^2\right] &= \mathrm{E}\left(\sum_{i=1}^{n} Y_i^6 + 7\sum_{i=1}^{n}\sum_{j\neq i} Y_i^2Y_j^4 + 4\sum_{i=1}^{n}\sum_{j\neq i} Y_iY_j^5\right. \\
&\quad + 4\sum_{i=1}^{n}\sum_{j\neq i} Y_i^3Y_j^3 + 16\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_iY_j^2Y_k^3 + 3\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_i^2Y_j^2Y_k^2 \\
&\quad + 6\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_iY_jY_k^4 + 6\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}} Y_iY_jY_k^2Y_l^2
\end{aligned}
$$

$$
\begin{aligned}
& + 4\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}} Y_iY_jY_kY_l^3 + \sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}}\sum_{\substack{m\neq i\\m\neq j\\m\neq k\\m\neq l}} Y_iY_jY_kY_lY_mY_o^2\Bigg) \\
= \; & n\mu_6 + 7n^2\mu_2\mu_4 - 7n\mu_2\mu_4 + 4n^2\mu_3^2 - 4n\mu_3^2 + 3n^3\mu_2^3 - 9n^2\mu_2^3 + 6n\mu_2^3
\end{aligned}
$$

---

15.

$$
\begin{aligned}
\mathrm{E}\left[\sum_{i=1}^{n} Y_i \sum_{i=1}^{n} Y_i^2 \sum_{i=1}^{n} Y_i^3\right] = \; & \mathrm{E}\Bigg(\sum_{i=1}^{n} Y_i^6 + \sum_{i=1}^{n}\sum_{j\neq i} Y_iY_j^5 + \sum_{i=1}^{n}\sum_{j\neq i} Y_i^2Y_j^4 \\
& + \sum_{i=1}^{n}\sum_{j\neq i} Y_i^3Y_j^3 + \sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_iY_j^2Y_k^3\Bigg) \\
= \; & n\mu_6 + n^2\mu_2\mu_4 - n\mu_2\mu_4 + n^2\mu_3^2 - n\mu_3^2
\end{aligned}
$$

---

16.

$$
\begin{aligned}
\mathrm{E}\left[\sum_{i=1}^{n} Y_i \sum_{i=1}^{n} Y_i^5\right] = \; & \mathrm{E}\left(\sum_{i=1}^{n} Y_i^6 + \sum_{i=1}^{n}\sum_{j\neq i} Y_iY_j^5\right) \\
= \; & n\mu_6
\end{aligned}
$$

---

17.

$$
\begin{aligned}
\mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^2 \left(\sum_{i=1}^{n} Y_i^2\right)^2\right] = \; & \mathrm{E}\Bigg(\sum_{i=1}^{n} Y_i^6 + 3\sum_{i=1}^{n}\sum_{j\neq i} Y_i^2Y_j^4 + 2\sum_{i=1}^{n}\sum_{j\neq i} Y_iY_j^5 \\
& + \sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_iY_jY_k^4 + \sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_i^2Y_j^2Y_k^2 \\
& + 2\sum_{i=1}^{n}\sum_{j\neq i} Y_i^3Y_j^3 + 4\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_iY_j^2Y_k^3 \\
& + \sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}} Y_iY_jY_k^2Y_l^2\Bigg) \\
= \; & n\mu_6 + 3n^2\mu_2\mu_4 - 3n\mu_2\mu_4 + n^3\mu_2^3 - 3n^2\mu_2^3 + 2n\mu_2^3 + 2n^2\mu_3^2 - 2n\mu_3^2
\end{aligned}
$$

---

18.

$$\begin{aligned}
\mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^3 \sum_{i=1}^{n} Y_i^3\right] &= \mathrm{E}\left(\sum_{i=1}^{n} Y_i^6 + \sum_{i=1}^{n}\sum_{j\neq i} Y_i^3 Y_j^3 + 3\sum_{i=1}^{n}\sum_{j\neq i} Y_i Y_j^5\right. \\
&\quad + 3\sum_{i=1}^{n}\sum_{j\neq i} Y_i^2 Y_j^4 + 3\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_i Y_j^2 Y_k^3 + 3\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_i Y_j Y_k^4 \\
&\quad \left. + \sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}} Y_i Y_j Y_k Y_l^3\right) \\
&= n\mu_6 + 3n^2\mu_2\mu_4 - 3n\mu_2\mu_4 + n^2\mu_3^2 - n\mu_3^2
\end{aligned}$$

---

19.

$$\begin{aligned}
\mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^4 \sum_{i=1}^{n} Y_i^2\right] &= \mathrm{E}\left(\sum_{i=1}^{n} Y_i^6 + 7\sum_{i=1}^{n}\sum_{j\neq i} Y_i^2 Y_j^4 + 4\sum_{i=1}^{n}\sum_{j\neq i} Y_i Y_j^5\right. \\
&\quad + 4\sum_{i=1}^{n}\sum_{j\neq i} Y_i^3 Y_j^3 + 16\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_i Y_j^2 Y_k^3 + 3\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_i^2 Y_j^2 Y_k^2 \\
&\quad + 6\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_i Y_j Y_k^4 + 6\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_i Y_j^2 Y_k^2 \\
&\quad + 4\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}} Y_i Y_j Y_k Y_l^3 \\
&\quad \left. + \sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}}\sum_{\substack{m\neq i\\m\neq j\\m\neq k\\m\neq l}} Y_i Y_j Y_k Y_l Y_m^2\right) \\
&= n\mu_6 + 7n^2\mu_2\mu_4 - 7n\mu_2\mu_4 + 4n^2\mu_3^2 - 4n\mu_3^2 + 3n^3\mu_2^3 - 9n^2\mu_2^3 + 6n\mu_2^3
\end{aligned}$$

---

20.

$$\begin{aligned}
\mathrm{E}\left[\sum_{i=1}^{n} Y_i^2 \sum_{i=1}^{n} Y_i^4\right] &= \mathrm{E}\left(\sum_{i=1}^{n} Y_i^6 + \sum_{i=1}^{n}\sum_{j\neq i} Y_i^2 Y_j^4\right) \\
&= n\mu_6 + n^2\mu_2\mu_4 - n\mu_2\mu_4
\end{aligned}$$

---

21.

$$\begin{aligned} E\left[\left(\sum_{i=1}^{n} Y_i^2\right)^3\right] &= E\left(\sum_{i=1}^{n} Y_i^6 + 3\sum_{i=1}^{n}\sum_{j\neq i} Y_i^2 Y_j^4 + \sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\ k\neq j}} Y_i^2 Y_j^2 Y_k^2\right) \\ &= n\mu_6 + 3n^2\mu_2\mu_4 - 3n\mu_2\mu_4 + n^3\mu_2^3 - 3n^2\mu_2^3 + 2n\mu_2^3 \end{aligned}$$

---

22.

$$\begin{aligned} E\left[\left(\sum_{i=1}^{n} Y_i^3\right)^2\right] &= E\left(\sum_{i=1}^{n} Y_i^6 + \sum_{i=1}^{n}\sum_{j\neq i} Y_i^3 Y_j^3\right) \\ &= n\mu_6 + n^2\mu_3^2 - n\mu_3^2 \end{aligned}$$

---

23.

$$\begin{aligned} E\left[\left(\sum_{i=1}^{n} Y_i\right)^7\right] &= E\left(\sum_{i=1}^{n} Y_i^7 + 7\sum_{i=1}^{n}\sum_{j\neq i} Y_i Y_j^6 + 21\sum_{i=1}^{n}\sum_{j\neq i} Y_i^2 Y_j^5\right. \\ &+ 21\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\ k\neq j}} Y_i Y_j Y_k^5 + 35\sum_{i=1}^{n}\sum_{j\neq i} Y_i^3 Y_j^4 + 105\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\ k\neq j}} Y_i Y_j^2 Y_k^4 \\ &+ 35\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\ k\neq j}}\sum_{\substack{l\neq i\\ l\neq j\\ l\neq k}} Y_i Y_j Y_k Y_l^4 + 70\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\ k\neq j}} Y_i Y_j^3 Y_k^3 \\ &+ 105\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\ k\neq j}} Y_i^2 Y_j^2 Y_k^3 + 210\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\ k\neq j}}\sum_{\substack{l\neq i\\ l\neq j\\ l\neq k}} Y_i Y_j Y_k^2 Y_l^3 \\ &+ 35\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\ k\neq j}}\sum_{\substack{l\neq i\\ l\neq j\\ l\neq k}}\sum_{\substack{m\neq i\\ m\neq j\\ m\neq k\\ m\neq l}} Y_i Y_j Y_k Y_l Y_m^3 + 105\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\ k\neq j}}\sum_{\substack{l\neq i\\ l\neq j\\ l\neq k}} Y_i Y_j^2 Y_k^2 Y_l^2 \\ &+ 105\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\ k\neq j}}\sum_{\substack{l\neq i\\ l\neq j\\ l\neq k}}\sum_{\substack{m\neq i\\ m\neq j\\ m\neq k\\ m\neq l}} Y_i Y_j Y_k Y_l^2 Y_m^2 \\ &+ 21\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\ k\neq j}}\sum_{\substack{l\neq i\\ l\neq j\\ l\neq k}}\sum_{\substack{m\neq i\\ m\neq j\\ m\neq k\\ m\neq l}}\sum_{\substack{o\neq i\\ o\neq j\\ o\neq k\\ o\neq l\\ o\neq m}} Y_i Y_j Y_k Y_l Y_m Y_o^2 \end{aligned}$$

$$+\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}}\sum_{\substack{m\neq i\\m\neq j\\m\neq k\\m\neq l}}\sum_{\substack{o\neq i\\o\neq j\\o\neq k\\o\neq l\\o\neq m}}\sum_{\substack{p\neq i\\p\neq j\\p\neq k\\p\neq l\\p\neq m\\p\neq o}} Y_iY_jY_kY_lY_mY_oY_p\Bigg)$$

$$= n\mu_7 + 21n^2\mu_2\mu_5 - 21n\mu_2\mu_5 + 35n^2\mu_3\mu_4 - 35n\mu_3\mu_4 + 105n^3\mu_2^2\mu_3 - 315n^2\mu_2^2\mu_3 + 210n\mu_2^2\mu_3$$

---

24.

$$\mathrm{E}\left[\sum_{i=1}^{n} Y_i\left(\sum_{i=1}^{n} Y_i^2\right)^3\right] = \mathrm{E}\Bigg(\sum_{i=1}^{n} Y_i^7 + \sum_{i=1}^{n}\sum_{j\neq i} Y_iY_j^6 + 3\sum_{i=1}^{n}\sum_{j\neq i} Y_i^2Y_j^5 + 3\sum_{i=1}^{n}\sum_{j\neq i} Y_i^3Y_j^4 + 3\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_iY_j^2Y_k^4 + 3\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_i^2Y_j^2Y_k^3 + \sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}} Y_iY_j^2Y_k^2Y_l^2\Bigg)$$

$$= n\mu_7 + 3n^2\mu_2\mu_5 - 3n\mu_2\mu_5 + 3n^2\mu_3\mu_4 - 3n\mu_3\mu_4 + 3n^3\mu_2^2\mu_3 - 9n^2\mu_2^2\mu_3 + 6n\mu_2^2\mu_3$$

---

25.

$$\mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^3\left(\sum_{i=1}^{n} Y_i^2\right)^2\right] = \mathrm{E}\Bigg(\sum_{i=1}^{n} Y_i^7 + 3\sum_{i=1}^{n}\sum_{j\neq i} Y_iY_j^6 + 5\sum_{i=1}^{n}\sum_{j\neq i} Y_i^2Y_j^5 + 3\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_iY_jY_k^5 + 7\sum_{i=1}^{n}\sum_{j\neq i} Y_i^3Y_j^4 + 9\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_iY_j^2Y_k^4 + \sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}} Y_iY_jY_kY_l^4 + 6\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_iY_j^3Y_k^3 + 7\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_i^2Y_j^2Y_k^3 + 6\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}} Y_iY_jY_k^2Y_l^3 + 3\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}} Y_iY_j^2Y_k^2Y_l^2$$

$$+ \sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}}\sum_{\substack{m\neq i\\m\neq j\\m\neq k\\m\neq l}} Y_iY_jY_kY_l^2Y_m^2\Bigg)$$

$$= n\mu_7 + 5n^2\mu_2\mu_5 - 5n\mu_2\mu_5 + 7n^2\mu_3\mu_4 - 7n\mu_3\mu_4 + 7n^3\mu_2^2\mu_3$$
$$-21n^2\mu_2^2\mu_3 + 14n\mu_2^2\mu_3$$

---

26.

$$\mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^5 \sum_{i=1}^{n} Y_i^2\right] = \mathrm{E}\Bigg(\sum_{i=1}^{n} Y_i^7 + 5\sum_{i=1}^{n}\sum_{j\neq i} Y_iY_j^6 + 11\sum_{i=1}^{n}\sum_{j\neq i} Y_i^2Y_j^5$$

$$+ 15\sum_{i=1}^{n}\sum_{j\neq i} Y_i^3Y_j^4 + 35\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_iY_j^2Y_k^4 + 25\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_i^2Y_j^2Y_k^3$$

$$+ 10\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_iY_jY_k^5 + 20\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_iY_j^3Y_k^3$$

$$+ 40\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}} Y_iY_jY_k^2Y_l^3$$

$$+ 15\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}} Y_iY_j^2Y_k^2Y_l^2 + 10\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}} Y_iY_jY_kY_l^4$$

$$+ 10\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}}\sum_{\substack{m\neq i\\m\neq j\\m\neq k\\m\neq l}} Y_iY_jY_kY_l^2Y_m^2$$

$$+ 5\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}}\sum_{\substack{m\neq i\\m\neq j\\m\neq k\\m\neq l}} Y_iY_jY_kY_lY_m^3$$

$$+ \sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}}\sum_{\substack{m\neq i\\m\neq j\\m\neq k\\m\neq l}}\sum_{\substack{o\neq i\\o\neq j\\o\neq k\\o\neq l\\o\neq m}} Y_iY_jY_kY_lY_mY_o^2\Bigg)$$

$$= n\mu_7 + 11n^2\mu_2\mu_5 - 11n\mu_2\mu_5 + 15n^2\mu_3\mu_4 - 15n\mu_3\mu_4 + 25n^3\mu_2^2\mu_3$$
$$-75n^2\mu_2^2\mu_3 + 50n\mu_2^2\mu_3$$

---

27.

$$
\begin{aligned}
\mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^{8}\right] = {} & \mathrm{E}\left(\sum_{i=1}^{n} Y_i^8 + 8\sum_{i=1}^{n}\sum_{j\neq i} Y_i Y_j^7 + 28\sum_{i=1}^{n}\sum_{j\neq i} Y_i^2 Y_j^6\right. \\
& + 56\sum_{i=1}^{n}\sum_{j\neq i} Y_i^3 Y_j^5 + 168\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\ k\neq j}} Y_i Y_j^2 Y_k^5 + 56\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\ k\neq j}}\sum_{\substack{l\neq i\\ l\neq j\\ l\neq k}} Y_i Y_j Y_k Y_l^5 \\
& + 35\sum_{i=1}^{n}\sum_{j\neq i} Y_i^4 Y_j^4 + 280\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\ k\neq j}} Y_i Y_j^3 Y_k^4 + 28\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\ k\neq j}} Y_i Y_j Y_k^6 \\
& + 210\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\ k\neq j}} Y_i^2 Y_j^2 Y_k^4 + 420\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\ k\neq j}}\sum_{\substack{l\neq i\\ l\neq j\\ l\neq k}} Y_i Y_j Y_k^2 Y_l^4 \\
& + 70\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\ k\neq j}}\sum_{\substack{l\neq i\\ l\neq j\\ l\neq k}}\sum_{\substack{m\neq i\\ m\neq j\\ m\neq k\\ m\neq l}} Y_i Y_j Y_k Y_l Y_m^4 + 280\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\ k\neq j}} Y_i^2 Y_j^3 Y_k^3 \\
& + 280\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\ k\neq j}}\sum_{\substack{l\neq i\\ l\neq j\\ l\neq k}} Y_i Y_j Y_k^3 Y_l^3 + 840\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\ k\neq j}}\sum_{\substack{l\neq i\\ l\neq j\\ l\neq k}} Y_i Y_j^2 Y_k^2 Y_l^3 \\
& + 560\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\ k\neq j}}\sum_{\substack{l\neq i\\ l\neq j\\ l\neq k}}\sum_{\substack{m\neq i\\ m\neq j\\ m\neq k\\ m\neq l}} Y_i Y_j Y_k Y_l^2 Y_m^3 \\
& + 56\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\ k\neq j}}\sum_{\substack{l\neq i\\ l\neq j\\ l\neq k}}\sum_{\substack{m\neq i\\ m\neq j\\ m\neq k\\ m\neq l}}\sum_{\substack{o\neq i\\ o\neq j\\ o\neq k\\ o\neq l\\ o\neq m}} Y_i Y_j Y_k Y_l Y_m Y_o^3 \\
& + 105\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\ k\neq j}}\sum_{\substack{l\neq i\\ l\neq j\\ l\neq k}} Y_i^2 Y_j^2 Y_k^2 Y_l^2 + 420\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\ k\neq j}}\sum_{\substack{l\neq i\\ l\neq j\\ l\neq k}}\sum_{\substack{m\neq i\\ m\neq j\\ m\neq k\\ m\neq l}} Y_i Y_j Y_k^2 Y_l^2 Y_m^2 \\
& + 210\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\ k\neq j}}\sum_{\substack{l\neq i\\ l\neq j\\ l\neq k}}\sum_{\substack{m\neq i\\ m\neq j\\ m\neq k\\ m\neq l}}\sum_{\substack{o\neq i\\ o\neq j\\ o\neq k\\ o\neq l\\ o\neq m}} Y_i Y_j Y_k Y_l Y_m^2 Y_o^2 \\
& + 28\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\ k\neq j}}\sum_{\substack{l\neq i\\ l\neq j\\ l\neq k}}\sum_{\substack{m\neq i\\ m\neq j\\ m\neq k\\ m\neq l}}\sum_{\substack{o\neq i\\ o\neq j\\ o\neq k\\ o\neq l\\ o\neq m}}\sum_{\substack{p\neq i\\ p\neq j\\ p\neq k\\ p\neq l\\ p\neq m\\ p\neq o}} Y_i Y_j Y_k Y_l Y_m Y_o Y_p^2
\end{aligned}
$$

$$
\left. + \sum_{i=1}^{n} \sum_{j\neq i} \sum_{\substack{k\neq i\\k\neq j}} \sum_{\substack{l\neq i\\l\neq j\\l\neq k}} \sum_{\substack{m\neq i\\m\neq j\\m\neq k\\m\neq l}} \sum_{\substack{o\neq i\\o\neq j\\o\neq k\\o\neq l\\o\neq m}} \sum_{\substack{p\neq i\\p\neq j\\p\neq k\\p\neq l\\p\neq m\\p\neq o}} \sum_{\substack{q\neq i\\q\neq j\\q\neq k\\q\neq l\\q\neq m\\q\neq o\\q\neq p}} Y_iY_jY_kY_lY_mY_oY_pY_q \right)
$$

$$
\begin{aligned}
= \; & n\mu_8 + 28n^2\mu_2\mu_6 - 28n\mu_2\mu_6 + 56n^2\mu_3\mu_5 - 56n\mu_3\mu_5 \\
& +35n^2\mu_4^2 - 35n\mu_4^2 + 210n^3\mu_2^2\mu_4 - 630n^2\mu_2^2\mu_4 + 420n\mu_2^2\mu_4 \\
& +280n^3\mu_2\mu_3^2 - 840n^2\mu_2\mu_3^2 + 560n\mu_2\mu_3^2 \\
& +105n^4\mu_2^4 - 630n^3\mu_2^4 + 1155n^2\mu_2^4 - 630n\mu_2^4
\end{aligned}
$$

---

28.

$$
\begin{aligned}
\mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^5 \sum_{i=1}^{n} Y_i^3\right] = \; & \mathrm{E}\left(\sum_{i=1}^{n} Y_i^8 + 11\sum_{i=1}^{n}\sum_{j\neq i} Y_i^3Y_j^5 + 5\sum_{i=1}^{n}\sum_{j\neq i} Y_iY_j^7\right. \\
& + 5\sum_{i=1}^{n}\sum_{j\neq i} Y_i^4Y_j^4 + 25\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_iY_j^3Y_k^4 + 10\sum_{i=1}^{n}\sum_{j\neq i} Y_i^2Y_j^6 \\
& + 10\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_iY_jY_k^6 + 10\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}} Y_iY_jY_k^3Y_l^3 \\
& + 30\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_iY_j^2Y_k^5 + 15\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_i^2Y_j^2Y_k^4 \\
& + 10\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_i^2Y_j^3Y_k^3 \\
& + 15\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}} Y_iY_j^2Y_k^2Y_l^3 + 10\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}} Y_iY_jY_kY_l^5 \\
& + 30\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}} Y_iY_jY_k^2Y_l^4 + 10\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}}\sum_{\substack{m\neq i\\m\neq j\\m\neq k\\m\neq l}} Y_iY_jY_kY_l^2Y_m^3 \\
& + 5\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}}\sum_{\substack{m\neq i\\m\neq j\\m\neq k\\m\neq l}} Y_iY_jY_kY_lY_m^4
\end{aligned}
$$

$$+ \sum_{i=1}^{n} \sum_{j \neq i} \sum_{\substack{k \neq i \\ k \neq j}} \sum_{\substack{l \neq i \\ l \neq j \\ l \neq k}} \sum_{\substack{m \neq i \\ m \neq j \\ m \neq k \\ m \neq l}} \sum_{\substack{o \neq i \\ o \neq j \\ o \neq k \\ o \neq l \\ o \neq m}} Y_i Y_j Y_k Y_l Y_m Y_o^3 \Bigg)$$

$$\begin{aligned} = \; & n\mu_8 + 11n^2\mu_3\mu_5 - 11n\mu_3\mu_5 + 5n^2\mu_4^2 - 5n\mu_4^2 \\ & +10n^2\mu_2\mu_6 - 10n\mu_2\mu_6 + 10n^3\mu_2\mu_3^2 - 30n^2\mu_2\mu_3^2 + 20n\mu_2\mu_3^2 \\ & +15n^3\mu_2^2\mu_4 - 45n^2\mu_2^2\mu_4 + 30n\mu_2^2\mu_4 \end{aligned}$$

---

29.

$$\begin{aligned} \mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^6 \sum_{i=1}^{n} Y_i^2\right] = \; & \mathrm{E}\Bigg(\sum_{i=1}^{n} Y_i^8 + 6\sum_{i=1}^{n}\sum_{j\neq i} Y_i Y_j^7 + 16\sum_{i=1}^{n}\sum_{j\neq i} Y_i^2 Y_j^6 \\ & + 26\sum_{i=1}^{n}\sum_{j\neq i} Y_i^3 Y_j^5 + 66\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_i Y_j^2 Y_k^5 + 60\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_i^2 Y_j^2 Y_k^4 \\ & + 15\sum_{i=1}^{n}\sum_{j\neq i} Y_i^4 Y_j^4 + 15\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_i Y_j Y_k^6 \\ & + 90\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_i Y_j^3 Y_k^4 + 105\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}} Y_i Y_j Y_k^2 Y_l^4 \\ & + 70\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_i^2 Y_j^3 Y_k^3 + 150\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}} Y_i Y_j^2 Y_k^2 Y_l^3 \\ & + 20\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}} Y_i Y_j Y_k Y_l^5 + 60\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}} Y_i Y_j Y_k^3 Y_l^3 \\ & + 80\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}}\sum_{\substack{m\neq i\\m\neq j\\m\neq k\\m\neq l}} Y_i Y_j Y_k Y_l^2 Y_m^3 \\ & + 15\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}} Y_i^2 Y_j^2 Y_k^2 Y_l^2 \\ & + 45\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}}\sum_{\substack{m\neq i\\m\neq j\\m\neq k\\m\neq l}} Y_i Y_j Y_k^2 Y_l^2 Y_m^2 \end{aligned}$$

$$+ 15\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}}\sum_{\substack{m\neq i\\m\neq j\\m\neq k\\m\neq l}} Y_iY_jY_kY_lY_m^4$$

$$+ 15\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}}\sum_{\substack{m\neq i\\m\neq j\\m\neq k\\m\neq l}}\sum_{\substack{o\neq i\\o\neq j\\o\neq k\\o\neq l\\o\neq m}} Y_iY_jY_kY_lY_m^2Y_o^2$$

$$+ 6\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}}\sum_{\substack{m\neq i\\m\neq j\\m\neq k\\m\neq l}}\sum_{\substack{o\neq i\\o\neq j\\o\neq k\\o\neq l\\o\neq m}} Y_iY_jY_kY_lY_mY_o^3$$

$$\left. + \sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}}\sum_{\substack{m\neq i\\m\neq j\\m\neq k\\m\neq l}}\sum_{\substack{o\neq i\\o\neq j\\o\neq k\\o\neq l\\o\neq m}}\sum_{\substack{p\neq i\\p\neq j\\p\neq k\\p\neq l\\p\neq m\\p\neq o}} Y_iY_jY_kY_lY_mY_oY_p^2 \right)$$

$$\begin{aligned} = \; & n\mu_8 + 16n^2\mu_2\mu_6 - 16n\mu_2\mu_6 + 26n^2\mu_3\mu_5 - 26n\mu_3\mu_5 \\ & +15n^2\mu_4^2 - 15n\mu_4^2 + 60n^3\mu_2^2\mu_4 - 180n^2\mu_2^2\mu_4 + 120n\mu_2^2\mu_4 \\ & +70n^3\mu_2\mu_3^2 - 210n^2\mu_2\mu_3^2 + 140n\mu_2\mu_3^2 \\ & +15n^4\mu_2^4 - 90n^3\mu_2^4 + 165n^2\mu_2^4 - 90n\mu_2^4 \end{aligned}$$

---

30.

$$\begin{aligned} \mathrm{E}\left[\sum_{i=1}^{n} Y_i\left(\sum_{i=1}^{n} Y_i^2\right)^2 \sum_{i=1}^{n} Y_i^3\right] = \; & \mathrm{E}\left(\sum_{i=1}^{n} Y_i^8 + \sum_{i=1}^{n}\sum_{j\neq i} Y_iY_j^7 + 2\sum_{i=1}^{n}\sum_{j\neq i} Y_i^2Y_j^6\right. \\ & + 3\sum_{i=1}^{n}\sum_{j\neq i} Y_i^3Y_j^5 + 2\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_iY_j^2Y_k^5 \\ & + \sum_{i=1}^{n}\sum_{j\neq i} Y_i^4Y_j^4 + \sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_iY_j^3Y_k^4 + \sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_i^2Y_j^2Y_k^4 \\ & \left. + 2\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_i^2Y_j^3Y_k^3 + \sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}} Y_iY_j^2Y_k^2Y_l^3\right) \\ = \; & n\mu_8 + 2n^2\mu_2\mu_6 - 2n\mu_2\mu_6 + 3n^2\mu_3\mu_5 - 3n\mu_3\mu_5 \\ & +n^2\mu_4^2 - n\mu_4^2 + n^3\mu_2^2\mu_4 - 3n^2\mu_2^2\mu_4 + 2n\mu_2^2\mu_4 \\ & +2n^3\mu_2\mu_3^2 - 6n^2\mu_2\mu_3^2 + 4n\mu_2\mu_3^2 \end{aligned}$$

31.

$$\begin{aligned} \mathrm{E}\left[\sum_{i=1}^{n} Y_i \sum_{i=1}^{n} Y_i^3 \sum_{i=1}^{n} Y_i^4\right] &= \mathrm{E}\left(\sum_{i=1}^{n} Y_i^8 + \sum_{i=1}^{n}\sum_{j\neq i} Y_i^4 Y_j^4 + \sum_{i=1}^{n}\sum_{j\neq i} Y_i Y_j^7\right. \\ &\quad \left. + \sum_{i=1}^{n}\sum_{j\neq i} Y_i^3 Y_j^5 + \sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\ k\neq j}} Y_i Y_j^3 Y_k^4\right) \\ &= n\mu_8 + n^2\mu_4^2 - n\mu_4^2 + n^2\mu_3\mu_5 - n\mu_3\mu_5 \end{aligned}$$

32.

$$\begin{aligned} \mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^2 \left(\sum_{i=1}^{n} Y_i^3\right)^2 \sum_{i=1}^{n} Y_i^3\right] &= \mathrm{E}\left(\sum_{i=1}^{n} Y_i^8 + 2\sum_{i=1}^{n}\sum_{j\neq i} Y_i^3 Y_j^5 + \sum_{i=1}^{n}\sum_{j\neq i} Y_i^2 Y_j^6\right. \\ &\quad + \sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\ k\neq j}} Y_i^2 Y_j^3 Y_k^3 + 2\sum_{i=1}^{n}\sum_{j\neq i} Y_i Y_j^7 \\ &\quad + 2\sum_{i=1}^{n}\sum_{j\neq i} Y_i^4 Y_j^4 + 4\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\ k\neq j}} Y_i Y_j^3 Y_k^4 \\ &\quad + \sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\ k\neq j}} Y_i Y_j Y_k^6 \\ &\quad \left. + 2\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\ k\neq j}}\sum_{\substack{l\neq i\\ l\neq j\\ l\neq k}} Y_i Y_j Y_k^3 Y_l^3\right) \\ &= n\mu_8 + 2n^2\mu_3\mu_5 - 2n\mu_3\mu_5 + n^2\mu_2\mu_6 - n\mu_2\mu_6 \\ &\quad + n^3\mu_2\mu_3^2 - 3n^2\mu_2\mu_3^2 + 2n\mu_2\mu_3^2 + 2n^2\mu_4^2 - 2n\mu_4^2 \end{aligned}$$

33.

$$\begin{aligned} \mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^3 \sum_{i=1}^{n} Y_i^2 \sum_{i=1}^{n} Y_i^3\right] &= \mathrm{E}\left(\sum_{i=1}^{n} Y_i^8 + 3\sum_{i=1}^{n}\sum_{j\neq i} Y_i Y_j^7 + 4\sum_{i=1}^{n}\sum_{j\neq i} Y_i^2 Y_j^6\right. \\ &\quad + 5\sum_{i=1}^{n}\sum_{j\neq i} Y_i^3 Y_j^5 + 6\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\ k\neq j}} Y_i Y_j^2 Y_k^5 \\ &\quad + \sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\ k\neq j}}\sum_{\substack{l\neq i\\ l\neq j\\ l\neq k}} Y_i Y_j Y_k Y_l^5 \end{aligned}$$

$$
\begin{aligned}
&+ 3\sum_{i=1}^{n}\sum_{j\neq i} Y_i^4Y_j^4 + 9\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_iY_j^3Y_k^4 \\
&+ 3\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_i^2Y_j^2Y_k^4 + 3\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}} Y_iY_jY_k^2Y_l^4 \\
&+ 4\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_i^2Y_j^3Y_k^3 + 3\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_iY_jY_k^6 \\
&+ 3\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}} Y_iY_jY_k^3Y_l^3 + 3\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}} Y_iY_j^2Y_k^2Y_l^3 \\
&\left.+ \sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}}\sum_{\substack{m\neq i\\m\neq j\\m\neq k\\m\neq l}} Y_iY_jY_kY_l^2Y_m^3\right) \\
= \;& n\mu_8 + 4n^2\mu_2\mu_6 - 4n\mu_2\mu_6 + 5n^2\mu_3\mu_5 - 5n\mu_3\mu_5 \\
&+3n^2\mu_4^2 - 3n\mu_4^2 + 3n^3\mu_2^2\mu_4 - 9n^2\mu_2^2\mu_4 + 6n\mu_2^2\mu_4 \\
&+4n^3\mu_2\mu_3^2 - 12n^2\mu_2\mu_3^2 + 8n\mu_2\mu_3^2
\end{aligned}
$$

---

34.

$$
\begin{aligned}
\mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^2\left(\sum_{i=1}^{n} Y_i^2\right)^3\right] = \;& \mathrm{E}\left(\sum_{i=1}^{n} Y_i^8 + 2\sum_{i=1}^{n}\sum_{j\neq i} Y_iY_j^7 + 4\sum_{i=1}^{n}\sum_{j\neq i} Y_i^2Y_j^6\right. \\
&+ 6\sum_{i=1}^{n}\sum_{j\neq i} Y_i^3Y_j^5 + 6\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_iY_j^2Y_k^5 \\
&+ 3\sum_{i=1}^{n}\sum_{j\neq i} Y_i^4Y_j^4 + 6\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_iY_j^3Y_k^4 \\
&+ 6\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_i^2Y_j^2Y_k^4 + 3\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}} Y_iY_jY_k^2Y_l^4 \\
&+ 6\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_i^2Y_j^3Y_k^3 + \sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_iY_jY_k^6 \\
&+ 6\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}} Y_iY_j^2Y_k^2Y_l^3
\end{aligned}
$$

$$+ \sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}} Y_i^2Y_j^2Y_k^2Y_l^2 + \sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}}\sum_{\substack{m\neq i\\m\neq j\\m\neq k\\m\neq l}} Y_iY_jY_k^2Y_l^2Y_m^2 \Bigg)$$

$$\begin{aligned} = \; & n\mu_8 + 4n^2\mu_2\mu_6 - 4n\mu_2\mu_6 + 6n^2\mu_3\mu_5 - 6n\mu_3\mu_5 \\ & +3n^2\mu_4^2 - 3n\mu_4^2 + 6n^3\mu_2^2\mu_4 - 18n^2\mu_2^2\mu_4 + 12n\mu_2^2\mu_4 \\ & +6n^3\mu_2\mu_3^2 - 18n^2\mu_2\mu_3^2 + 12n\mu_2\mu_3^2 \\ & +n^4\mu_2^4 - 6n^3\mu_2^4 + 11n^2\mu_2^4 - 6n\mu_2^4 \end{aligned}$$

---

35.

$$\begin{aligned} \mathrm{E}\left[\left(\sum_{i=1}^{n}Y_i\right)^2\sum_{i=1}^{n}Y_i^2\sum_{i=1}^{n}Y_i^4\right] = \; & \mathrm{E}\Bigg(\sum_{i=1}^{n}Y_i^8 + \sum_{i=1}^{n}\sum_{j\neq i}Y_i^4Y_j^4 + 2\sum_{i=1}^{n}\sum_{j\neq i}Y_i^2Y_j^6 + \sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_i^2Y_j^2Y_k^4 \\ & + \sum_{i=1}^{n}\sum_{j\neq i}Y_iY_j^7 + 2\sum_{i=1}^{n}\sum_{j\neq i}Y_i^3Y_j^5 + 2\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_iY_j^3Y_k^4 \\ & + 3\sum_{i=1}^{n}\sum_{j\neq i}Y_i^4Y_j^4 + 6\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_iY_j^3Y_k^4 \\ & + 2\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_iY_j^2Y_k^5 + \sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_iY_jY_k^6 \\ & + \sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}} Y_iY_jY_k^2Y_l^4\Bigg) \\ = \; & n\mu_8 + n^2\mu_4^2 - n\mu_4^2 + 2n^2\mu_2\mu_6 - 2n\mu_2\mu_6 \\ & +n^3\mu_2^2\mu_4 - 3n^2\mu_2^2\mu_4 + 2n\mu_2^2\mu_4 + 2n^2\mu_3\mu_5 - 2n\mu_3\mu_5 \end{aligned}$$

---

36.

$$\begin{aligned} \mathrm{E}\left[\left(\sum_{i=1}^{n}Y_i\right)^4\left(\sum_{i=1}^{n}Y_i^2\right)^2\right] = \; & \mathrm{E}\Bigg(\sum_{i=1}^{n}Y_i^8 + 4\sum_{i=1}^{n}\sum_{j\neq i}Y_iY_j^7 + 8\sum_{i=1}^{n}\sum_{j\neq i}Y_i^2Y_j^6 \\ & + 12\sum_{i=1}^{n}\sum_{j\neq i}Y_i^3Y_j^5 + 20\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_iY_j^2Y_k^5 \\ & + 4\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}} Y_iY_jY_kY_l^5 \end{aligned}$$

$$
\begin{aligned}
&+ 7\sum_{i=1}^{n}\sum_{j\neq i} Y_i^4 Y_j^4 + 28\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_i Y_j^3 Y_k^4 + 6\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_i Y_j Y_k^6 \\
&+ 16\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_i^2 Y_j^2 Y_k^4 + 18\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}} Y_i Y_j Y_k^2 Y_l^4 \\
&+ \sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}}\sum_{\substack{m\neq i\\m\neq j\\m\neq k\\m\neq l}} Y_i Y_j Y_k Y_l Y_m^4 + 20\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_i^2 Y_j^3 Y_k^3 \\
&+ 12\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}} Y_i Y_j Y_k^3 Y_l^3 + 28\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}} Y_i Y_j^2 Y_k^2 Y_l^3 \\
&+ 8\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}}\sum_{\substack{m\neq i\\m\neq j\\m\neq k\\m\neq l}} Y_i Y_j Y_k Y_l^2 Y_m^3 \\
&+ 3\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}} Y_i^2 Y_j^2 Y_k^2 Y_l^2 \\
&+ 6\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}}\sum_{\substack{m\neq i\\m\neq j\\m\neq k\\m\neq l}} Y_i Y_j Y_k^2 Y_l^2 Y_m^2 \\
&\left.+ \sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}}\sum_{\substack{m\neq i\\m\neq j\\m\neq k\\m\neq l}}\sum_{\substack{o\neq i\\o\neq j\\o\neq k\\o\neq l\\o\neq m}} Y_i Y_j Y_k Y_l Y_m^2 Y_o^2 \right) \\
= \;& n\mu_8 + 8n^2\mu_2\mu_6 - 8n\mu_2\mu_6 + 12n^2\mu_3\mu_5 - 12n\mu_3\mu_5 \\
&+7n^2\mu_4^2 - 7n\mu_4^2 + 16n^3\mu_2^2\mu_4 - 48n^2\mu_2^2\mu_4 + 32n\mu_2^2\mu_4 \\
&+20n^3\mu_2\mu_3^2 - 60n^2\mu_2\mu_3^2 + 40n\mu_2\mu_3^2 \\
&+3n^4\mu_2^4 - 18n^3\mu_2^4 + 33n^2\mu_2^4 - 18n\mu_2^4
\end{aligned}
$$

---

37.

$$
\begin{aligned}
\mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^4 \sum_{i=1}^{n} Y_i^4\right] = \;& \mathrm{E}\left(\sum_{i=1}^{n} Y_i^8 + \sum_{i=1}^{n}\sum_{j\neq i} Y_i^4 Y_j^4 + 4\sum_{i=1}^{n}\sum_{j\neq i} Y_i Y_j^7 + 4\sum_{i=1}^{n}\sum_{j\neq i} Y_i^3 Y_j^5\right. \\
&+ 4\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_i Y_j^3 Y_k^4 + 6\sum_{i=1}^{n}\sum_{j\neq i} Y_i^2 Y_j^6 + 3\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_i^2 Y_j^2 Y_k^4
\end{aligned}
$$

$$
\begin{aligned}
&+ 6\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_iY_jY_k^6 + 12\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}} Y_iY_j^2Y_k^5 \\
&+ 6\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}} Y_iY_jY_k^2Y_l^4 + 4\sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}} Y_iY_jY_kY_l^5 \\
&\left. + \sum_{i=1}^{n}\sum_{j\neq i}\sum_{\substack{k\neq i\\k\neq j}}\sum_{\substack{l\neq i\\l\neq j\\l\neq k}}\sum_{\substack{m\neq i\\m\neq j\\m\neq k\\m\neq l}} Y_iY_jY_kY_lY_m^4 \right) \\
= \; & n\mu_8 + n^2\mu_4^2 - n\mu_4^2 + 4n^2\mu_3\mu_5 - 4n\mu_3\mu_5 \\
& +6n^2\mu_2\mu_6 - 6n\mu_2\mu_6 + 3n^3\mu_2^2\mu_4 - 9n^2\mu_2^2\mu_4 + 6n\mu_2^2\mu_4
\end{aligned}
$$

---

38.

$$
\begin{aligned}
\mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i^2\right)^1\right] &= \mathrm{E}\left(\sum_{i=1}^{n} Y_i^8 + \sum_{i=1}^{n}\sum_{j\neq i} Y_i^4Y_j^4\right) \\
&= n\mu_8 + n^2\mu_4^2 - n\mu_4^2
\end{aligned}
$$

---

39. For the following expressions we will only provide the result of the expected value because of the length of some of the expressions; the intermediate steps are obtained in a similar manner to the previous expressions but they will be ommitted. These calculations can be obtained from the candidate on request.

$$
\begin{aligned}
\mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^{10}\right] = \; & n\mu_{10} + 45n^2\mu_2\mu_8 - 45n\mu_2\mu_8 + 120n^2\mu_3\mu_7 - 120n\mu_3\mu_7 \\
& +210n^2\mu_4\mu_6 - 210n\mu_4\mu_6 + 630n^3\mu_2^2\mu_6 - 1890n^2\mu_2^2\mu_6 + 1260n\mu_2^2\mu_6 \\
& +126n^2\mu_5^2 - 126n\mu_5^2 + 2520n^3\mu_2\mu_3\mu_5 - 7560n^2\mu_2\mu_3\mu_5 + 5040n\mu_2\mu_3\mu_5 \\
& +1575n^3\mu_2\mu_4^2 - 4725n^2\mu_2\mu_4^2 + 3150n\mu_2\mu_4^2 + 2100n^3\mu_3^2\mu_4 - 6300n^2\mu_3^2\mu_4 \\
& +4200n\mu_3^2\mu_4 + 3150n^4\mu_2^3\mu_4 - 18900n^3\mu_2^3\mu_4 + 34650n^2\mu_2^3\mu_4 - 18900n\mu_2^3\mu_4 \\
& +6300n^4\mu_2^2\mu_3^2 - 37800n^3\mu_2^2\mu_3^2 + 69300n^2\mu_2^2\mu_3^2 - 37800n\mu_2^2\mu_3^2 \\
& +945n^5\mu_2^5 - 9450n^4\mu_2^5 + 33075n^3\mu_2^5 - 47250n^2\mu_2^5 + 22680n\mu_2^5
\end{aligned}
$$

---

40.

$$
\mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^8 \sum_{i=1}^{n} Y_i^2\right] = n\mu_{10} + 29n^2\mu_2\mu_8 - 29n\mu_2\mu_8 + 64n^2\mu_3\mu_7 - 64n\mu_3\mu_7
$$

$$+98n^2\mu_4\mu_6 - 98n\mu_4\mu_6 + 238n^3\mu_2^2\mu_6 - 714n^2\mu_2^2\mu_6 + 476n\mu_2^2\mu_6 + 56n^2\mu_5^2 - 56n\mu_5^2 + 784n^3\mu_2\mu_3\mu_5 - 2352n^2\mu_2\mu_3\mu_5 + 1568n\mu_2\mu_3\mu_5 + 455n^3\mu_2\mu_4^2 - 1365n^2\mu_2\mu_4^2 + 910n\mu_2\mu_4^2 + 560n^3\mu_3^2\mu_4 - 1680n^2\mu_3^2\mu_4 + 1120n\mu_3^2\mu_4 + 630n^4\mu_2^3\mu_4 - 3780n^3\mu_2^3\mu_4 + 6930n^2\mu_2^3\mu_4 - 3780n\mu_2^3\mu_4 + 1120n^4\mu_2^2\mu_3^2 - 6720n^3\mu_2^2\mu_3^2 + 12320n^2\mu_2^2\mu_3^2 - 6720n\mu_2^2\mu_3^2 + 105n^5\mu_2^5 - 1050n^4\mu_2^5 + 3675n^3\mu_2^5 - 5250n^2\mu_2^5 + 2520n\mu_2^5$$

---

41.

$$\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^7 \sum_{i=1}^n Y_i^3\right] = n\mu_{10} + 21n^2\mu_2\mu_8 - 21n\mu_2\mu_8 + 36n^2\mu_3\mu_7 - 36n\mu_3\mu_7 + 42n^2\mu_4\mu_6 - 42n\mu_4\mu_6 + 105n^3\mu_2^2\mu_6 - 315n^2\mu_2^2\mu_6 + 210n\mu_2^2\mu_6 + 21n^2\mu_5^2 - 21n\mu_5^2 + 231n^3\mu_2\mu_3\mu_5 - 693n^2\mu_2\mu_3\mu_5 + 462n\mu_2\mu_3\mu_5 + 105n^3\mu_2\mu_4^2 - 315n^2\mu_2\mu_4^2 + 210n\mu_2\mu_4^2 + 105n^3\mu_3^2\mu_4 - 315n^2\mu_3^2\mu_4 + 210n\mu_3^2\mu_4 + 105n^4\mu_2^3\mu_4 - 630n^3\mu_2^3\mu_4 + 1155n^2\mu_2^3\mu_4 - 630n\mu_2^3\mu_4 + 105n^4\mu_2^2\mu_3^2 - 630n^3\mu_2^2\mu_3^2 + 1155n^2\mu_2^2\mu_3^2 - 630n\mu_2^2\mu_3^2$$

---

42.

$$\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^6 \left(\sum_{i=1}^n Y_i^2\right)^2\right] = n\mu_{10} + 17n^2\mu_2\mu_8 - 17n\mu_2\mu_8 + 32n^2\mu_3\mu_7 - 32n\mu_3\mu_7 + 46n^2\mu_4\mu_6 - 46n\mu_4\mu_6 + 76n^3\mu_2^2\mu_6 - 228n^2\mu_2^2\mu_6 + 152n\mu_2^2\mu_6 + 26n^2\mu_5^2 - 26n\mu_5^2 + 232n^3\mu_2\mu_3\mu_5 - 696n^2\mu_2\mu_3\mu_5 + 464n\mu_2\mu_3\mu_5 + 135n^3\mu_2\mu_4^2 - 405n^2\mu_2\mu_4^2 + 270n\mu_2\mu_4^2 + 160n^3\mu_3^2\mu_4 - 480n^2\mu_3^2\mu_4 + 320n\mu_3^2\mu_4 + 120n^4\mu_2^3\mu_4 - 720n^3\mu_2^3\mu_4 + 1320n^2\mu_2^3\mu_4 - 720n\mu_2^3\mu_4 + 220n^4\mu_2^2\mu_3^2 - 1320n^3\mu_2^2\mu_3^2 + 2420n^2\mu_2^2\mu_3^2 - 1320n\mu_2^2\mu_3^2 + 15n^5\mu_2^5 - 150n^4\mu_2^5 + 525n^3\mu_2^5 - 750n^2\mu_2^5 + 360n\mu_2^5$$

---

43.

$$\mathrm{E}\left[\left(\sum_{i=1}^n Y_i\right)^6 \sum_{i=1}^n Y_i^4\right] = n\mu_{10} + 16n^2\mu_4\mu_6 - 16n\mu_4\mu_6 + 6n^2\mu_5^2 - 6n\mu_5^2 + 15n^2\mu_2\mu_8$$

$$-15n\mu_2\mu_8 + 15n^3\mu_2\mu_4^2 - 45n^2\mu_2\mu_4^2 + 30n\mu_2\mu_4^2 + 20n^2\mu_3\mu_7$$
$$-20n\mu_3\mu_7 + 10n^3\mu_3^2\mu_4 - 30n^2\mu_3^2\mu_4 + 20n\mu_3^2\mu_4 + 60n^3\mu_2\mu_3\mu_5$$
$$-180n^2\mu_2\mu_3\mu_5 + 120n\mu_2\mu_3\mu_5$$

---

44.

$$\mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^5 \sum_{i=1}^{n} Y_i^2 \sum_{i=1}^{n} Y_i^3\right] = n\mu_{10} + 11n^2\mu_2\mu_8 - 11n\mu_2\mu_8 + 16n^2\mu_3\mu_7 - 16n\mu_3\mu_7$$
$$+20n^2\mu_4\mu_6 - 20n\mu_4\mu_6 + 25n^3\mu_2^2\mu_6 - 75n^2\mu_2^2\mu_6$$
$$+50n\mu_2^2\mu_6 + 11n^2\mu_5^2 - 11n\mu_5^2 + 61n^3\mu_2\mu_3\mu_5$$
$$-183n^2\mu_2\mu_3\mu_5 + 122n\mu_2\mu_3\mu_5 + 35n^3\mu_2\mu_4^2$$
$$-105n^2\mu_2\mu_4^2 + 70n\mu_2\mu_4^2 + 35n^3\mu_3^2\mu_4 - 105n^2\mu_3^2\mu_4$$
$$+70n\mu_3^2\mu_4 + 15n^4\mu_2^3\mu_4 - 90n^3\mu_2^3\mu_4 + 165n^2\mu_2^3\mu_4$$
$$-90n\mu_2^3\mu_4 + 25n^4\mu_2^2\mu_3^2 - 150n^3\mu_2^2\mu_3^2 + 275n^2\mu_2^2\mu_3^2$$
$$-150n\mu_2^2\mu_3^2$$

---

45.

$$\mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^5 \sum_{i=1}^{n} Y_i^5\right] = n\mu_{10} + n^2\mu_5^2 - n\mu_5^2 + 5n^2\mu_4\mu_6 - 5n\mu_4\mu_6$$
$$+10n^2\mu_2\mu_8 - 10n\mu_2\mu_8 + 10n^2\mu_3\mu_7 - 10n\mu_3\mu_7$$
$$+10n^3\mu_2\mu_3\mu_5 - 30n^2\mu_2\mu_3\mu_5 + 20n\mu_2\mu_3\mu_5$$
$$+15n^3\mu_2^2\mu_6 - 45n^2\mu_2^2\mu_6 + 30n\mu_2^2\mu_6$$

---

46.

$$\mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^4 \left(\sum_{i=1}^{n} Y_i^2\right)^3\right] = n\mu_{10} + 9n^2\mu_2\mu_8 - 9n\mu_2\mu_8 + 16n^2\mu_3\mu_7 - 16n\mu_3\mu_7$$
$$+22n^2\mu_4\mu_6 - 22n\mu_4\mu_6 + 24n^2\mu_2^2\mu_6 - 24n\mu_2^2\mu_6$$
$$+12n^2\mu_5^2 - 12n\mu_5^2 + 72n^3\mu_2\mu_3\mu_5 - 216n^2\mu_2\mu_3\mu_5$$
$$+144n\mu_2\mu_3\mu_5 + 39n^3\mu_2\mu_4^2 - 117n^2\mu_2\mu_4^2 + 78n\mu_2\mu_4^2$$
$$+48n^3\mu_3^2\mu_4 - 144n^2\mu_3^2\mu_4 + 96n\mu_3^2\mu_4 + 28n^4\mu_2^3\mu_4$$
$$-168n^3\mu_2^3\mu_4 + 308n^2\mu_2^3\mu_4 - 168n\mu_2^3\mu_4 + 48n^4\mu_2^2\mu_3^2$$
$$-288n^3\mu_2^2\mu_3^2 + 528n^2\mu_2^2\mu_3^2 - 288n\mu_2^2\mu_3^2$$
$$+3n^5\mu_2^5 - 30n^4\mu_2^5 + 105n^3\mu_2^5 - 150n^2\mu_2^5 + 72n\mu_2^5$$

---

47.

$$\mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^4 \left(\sum_{i=1}^{n} Y_i^3\right)^2\right] = n\mu_{10} + 6n^2\mu_2\mu_8 - 6n\mu_2\mu_8 + 6n^2\mu_3\mu_7 - 6n\mu_3\mu_7$$
$$+9n^2\mu_4\mu_6 - 9n\mu_4\mu_6 + 3n^3\mu_2^2\mu_6 - 9n^2\mu_2^2\mu_6$$
$$+6n\mu_2^2\mu_6 + 6n^2\mu_5^2 - 6n\mu_5^2 + 12n^3\mu_2\mu_3\mu_5$$
$$-36n^2\mu_2\mu_3\mu_5 + 24n\mu_2\mu_3\mu_5 + 12n^3\mu_2\mu_4^2$$
$$-36n^2\mu_2\mu_4^2 + 24n\mu_2\mu_4^2 + 9n^3\mu_3^2\mu_4$$
$$-27n^2\mu_3^2\mu_4 + 28n\mu_3^2\mu_4 + 3n^4\mu_2^2\mu_3^2$$
$$-18n^3\mu_2^2\mu_3^2 + 33n^2\mu_2^2\mu_3^2 - 18n\mu_2^2\mu_3^2$$

---

48.

$$\mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^4 \sum_{i=1}^{n} Y_i^6\right] = n\mu_{10} + n^2\mu_4\mu_6 - n\mu_4\mu_6 + 4n^2\mu_3\mu_7 - 4n\mu_3\mu_7 + 6n^2\mu_2\mu_8 - 6n\mu_2\mu_8$$
$$+3n^3\mu_2^2\mu_6 - 9n^2\mu_2^2\mu_6 + 6n\mu_2^2\mu_6$$

---

49.

$$\mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^4 \sum_{i=1}^{n} Y_i^2 \sum_{i=1}^{n} Y_i^4\right] = n\mu_{10} + 8n^2\mu_4\mu_6 - 8n\mu_4\mu_6 + 7n^2\mu_2\mu_8 - 7n\mu_2\mu_8 + 7n^3\mu_2\mu_4^2$$
$$-21n^2\mu_2\mu_4^2 + 14n\mu_2\mu_4^2 + 4n^2\mu_5^2 - 4n\mu_5^2 + 8n^2\mu_3\mu_7$$
$$-8n\mu_3\mu_7 + 4n^3\mu_3^2\mu_4 - 12n^2\mu_3^2\mu_4 + 8n\mu_3^2\mu_4$$
$$+16n^3\mu_2\mu_3\mu_5 - 48n^2\mu_2\mu_3\mu_5 + 32n\mu_2\mu_3\mu_5 + 9n^3\mu_2^2\mu_6$$
$$-27n^2\mu_2^2\mu_6 + 18n\mu_2^2\mu_6 + 3n^4\mu_2^3\mu_4 - 18n^3\mu_2^3\mu_4$$
$$+33n^2\mu_2^3\mu_4 - 18n\mu_2^3\mu_4$$

---

50.

$$\mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^3 \left(\sum_{i=1}^{n} Y_i^2\right)^2 \sum_{i=1}^{n} Y_i^3\right] = n\mu_{10} + 5n^2\mu_2\mu_8 - 5n\mu_2\mu_8 + 8n^2\mu_3\mu_7 - 8n\mu_3\mu_7$$
$$+n^2\mu_4\mu_6 - n\mu_4\mu_6 + 7n^3\mu_2^2\mu_6 - 21n^2\mu_2^2\mu_6$$
$$+14n\mu_2^2\mu_6 + 5n^2\mu_5^2 - 5n\mu_5^2 + 19n^3\mu_2\mu_3\mu_5$$
$$-57n^2\mu_2\mu_3\mu_5 + 38n\mu_2\mu_3\mu_5 + 9n^3\mu_2\mu_4^2 - 27n^2\mu_2\mu_4^2$$
$$+18n\mu_2\mu_4^2 + 13n^3\mu_3^2\mu_4 - 39n^2\mu_3^2\mu_4 + 26n\mu_3^2\mu_4$$
$$+3n^4\mu_2^3\mu_4 - 18n^3\mu_2^3\mu_4 + 33n^2\mu_2^3\mu_4 - 18n\mu_2^3\mu_4$$
$$+7n^4\mu_2^2\mu_3^2 - 42n^3\mu_2^2\mu_3^2 + 77n^2\mu_2^2\mu_3^2 - 42n\mu_2^2\mu_3^2$$

---

51.

$$\mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^3 \sum_{i=1}^{n} Y_i^2 \sum_{i=1}^{n} Y_i^5\right] = n\mu_{10} + n^2\mu_5^2 - n\mu_5^2 + 4n^2\mu_2\mu_8 - 4n\mu_2\mu_8 + 4n^2\mu_3\mu_7 - 4n\mu_3\mu_7 + 4n^3\mu_2\mu_3\mu_5 - 12n^2\mu_2\mu_3\mu_5 + 8n\mu_2\mu_3\mu_5 + 3n^2\mu_4\mu_6 - 3n\mu_4\mu_6 + 3n^3\mu_2^2\mu_6 - 9n^2\mu_2^2\mu_6 + 6n\mu_2^2\mu_6$$

---

52.

$$\mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^3 \sum_{i=1}^{n} Y_i^3 \sum_{i=1}^{n} Y_i^4\right] = n\mu_{10} + 4n^2\mu_4\mu_6 - 4n\mu_4\mu_6 + 2n^2\mu_3\mu_7 - 2n\mu_3\mu_7 + n^3\mu_3^2\mu_4 - 3n^2\mu_3^2\mu_4 + 2n\mu_3^2\mu_4 + 3n^2\mu_5^2 - 3n\mu_5^2 + 3n^2\mu_2\mu_8 - 3n\mu_2\mu_8 + 3n^3\mu_2\mu_4^2 - 9n^2\mu_2\mu_4^2 + 6n\mu_2\mu_4^2 + 3n^3\mu_2\mu_3\mu_5 - 9n^2\mu_2\mu_3\mu_5 + 6n\mu_2\mu_3\mu_5$$

---

53.

$$\mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^3 \sum_{i=1}^{n} Y_i^3 \sum_{i=1}^{n} Y_i^4\right] = n\mu_{10} + 4n^2\mu_4\mu_6 - 4n\mu_4\mu_6 + 2n^2\mu_3\mu_7 - 2n\mu_3\mu_7 + n^3\mu_3^2\mu_4 - 3n^2\mu_3^2\mu_4 + 2n\mu_3^2\mu_4 + 3n^2\mu_5^2 - 3n\mu_5^2 + 3n^2\mu_2\mu_8 - 3n\mu_2\mu_8 + 3n^3\mu_2\mu_4^2 - 9n^2\mu_2\mu_4^2 + 6n\mu_2\mu_4^2 + 3n^3\mu_2\mu_3\mu_5 - 9n^2\mu_2\mu_3\mu_5 + 6n\mu_2\mu_3\mu_5$$

---

54.

$$\mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^2 \left(\sum_{i=1}^{n} Y_i^2\right)^4\right] = n\mu_{10} + 5n^2\mu_2\mu_8 - 5n\mu_2\mu_8 + 8n^2\mu_3\mu_7 - 8n\mu_3\mu_7 + 10n^2\mu_4\mu_6 - 10n\mu_4\mu_6 + 10n^3\mu_2^2\mu_6 - 30n^2\mu_2^2\mu_6 + 20n\mu_2^2\mu_6 + 6n^2\mu_5^2 - 6n\mu_5^2 + 15n^3\mu_2\mu_4^2 - 45n^2\mu_2\mu_4^2 + 30n\mu_2\mu_4^2 + 24n^3\mu_2\mu_3\mu_5 - 72n^2\mu_2\mu_3\mu_5 + 48n\mu_2\mu_3\mu_5 + 12n^3\mu_3^2\mu_4 - 36n^2\mu_3^2\mu_4 + 24n\mu_3^2\mu_4 + 10n^4\mu_2^3\mu_4 - 60n^3\mu_2^3\mu_4 + 110n^2\mu_2^3\mu_4 - 60n\mu_2^3\mu_4 + 12n^4\mu_2^2\mu_3^2 - 72n^3\mu_2^2\mu_3^2 + 132n^2\mu_2^2\mu_3^2 - 72n\mu_2^2\mu_3^2 + n^5\mu_2^5 - 10n^4\mu_2^5 + 35n^3\mu_2^5 - 50n^2\mu_2^5 + 24n\mu_2^5$$

---

55.

$$\mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^2 \sum_{i=1}^{n} Y_i^2 \left(\sum_{i=1}^{n} Y_i^3\right)^2\right] = n\mu_{10} + 2n^2\mu_2\mu_8 - 2n\mu_2\mu_8 + 4n^2\mu_3\mu_7 - 4n\mu_3\mu_7$$

$$+5n^2\mu_4\mu_6 - 5n\mu_4\mu_6 + n^3\mu_2^2\mu_6 - 3n^2\mu_2^2\mu_6 + 2n\mu_2^2\mu_6 + 2n^2\mu_5^2 - 2n\mu_5^2 + 4n^3\mu_2\mu_3\mu_5 - 12n^2\mu_2\mu_3\mu_5 + 8n\mu_2\mu_3\mu_5 + 2n^3\mu_2\mu_4^2 - 6n^2\mu_2\mu_4^2 + 4n\mu_2\mu_4^2 + 5n^3\mu_3^2\mu_4 - 15n^2\mu_3^2\mu_4 + 10n\mu_3^2\mu_4 + n^4\mu_2^2\mu_3^2 - 6n^3\mu_2^2\mu_3^2 + 11n^2\mu_2^2\mu_3^2 - 6n\mu_2^2\mu_3^2$$

---

56.

$$\mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^2 \sum_{i=1}^{n} Y_i^2 \sum_{i=1}^{n} Y_i^6\right] = n\mu_{10} + n^2\mu_4\mu_6 - n\mu_4\mu_6 + 2n^2\mu_2\mu_8 - 2n\mu_2\mu_8 + n^3\mu_2^2\mu_6 - 3n^2\mu_2^2\mu_6 + 2n\mu_2^2\mu_6 + 2n^2\mu_3\mu_7 - 2n\mu_3\mu_7$$

---

57.

$$\mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^2 \left(\sum_{i=1}^{n} Y_i^2\right)^2 \sum_{i=1}^{n} Y_i^4\right] = n\mu_{10} + 4n^2\mu_4\mu_6 - 4n\mu_4\mu_6 + 3n^2\mu_2\mu_8 - 3n\mu_2\mu_8 + 3n^3\mu_2\mu_4^2 - 9n^2\mu_2\mu_4^2 + 6n\mu_2\mu_4^2 + 3n^3\mu_2^2\mu_6 - 9n^2\mu_2^2\mu_6 + 6n\mu_2^2\mu_6 + n^4\mu_2^3\mu_4 - 6n^3\mu_2^3\mu_4 + 11n^2\mu_2^3\mu_4 - 6n\mu_2^3\mu_4 + 2n^2\mu_5^2 - 2n\mu_5^2 + 4n^2\mu_3\mu_7 - 4n\mu_3\mu_7 + 2n^3\mu_3^2\mu_4 - 6n^2\mu_3^2\mu_4 + 4n\mu_3^2\mu_4 + 4n^3\mu_2\mu_3\mu_4 - 12n^2\mu_2\mu_3\mu_4 + 8n\mu_2\mu_3\mu_4$$

---

58.

$$\mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^2 \sum_{i=1}^{n} Y_i^3 \sum_{i=1}^{n} Y_i^5\right] = n\mu_{10} + n^2\mu_5^2 - n\mu_5^2 + n^2\mu_3\mu_7 - n\mu_3\mu_7 + n^2\mu_2\mu_8 - n\mu_2\mu_8 + n^3\mu_2\mu_3\mu_5 - 3n^2\mu_2\mu_3\mu_5 + 2n\mu_2\mu_3\mu_5 + 2n^2\mu_4\mu_6 - 2n\mu_4\mu_6$$

---

59.

$$\mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^2 \left(\sum_{i=1}^{n} Y_i^4\right)^2\right] = n\mu_{10} + 2n^2\mu_4\mu_6 - 2n\mu_4\mu_6 + n^2\mu_2\mu_8 - n\mu_2\mu_8 + n^3\mu_2\mu_4^2 - 3n^2\mu_2\mu_4^2 + 2n\mu_2\mu_4^2 + 2n^2\mu_5^2 - 2n\mu_5^2$$

---

60.

$$\begin{aligned} \mathrm{E}\left[\sum_{i=1}^{n} Y_i \left(\sum_{i=1}^{n} Y_i^2\right)^3 \sum_{i=1}^{n} Y_i^3\right] &= n\mu_{10} + 3n^2\mu_2\mu_8 - 3n\mu_2\mu_8 + 4n^2\mu_3\mu_7 - 4n\mu_3\mu_7 \\ &+4n^2\mu_4\mu_6 - 4n\mu_4\mu_6 + 3n^2\mu_5^2 - 3n\mu_5^2 + 3n^3\mu_2^2\mu_6 \\ &-9n^2\mu_2^2\mu_6 + 6n\mu_2^2\mu_6 + 9n^3\mu_2\mu_3\mu_5 - 27n^2\mu_2\mu_3\mu_5 \\ &+18n\mu_2\mu_3\mu_5 + 3n^3\mu_2\mu_4^2 - 9n^2\mu_2\mu_4^2 + 6n\mu_2\mu_4^2 \\ &+3n^3\mu_3^2\mu_4 - 9n^2\mu_3^2\mu_4 + 6n\mu_3^2\mu_4 + n^4\mu_2^3\mu_4 - 6n^3\mu_2^3\mu_4 \\ &+11n^2\mu_2^3\mu_4 - 6n\mu_2^3\mu_4 + 3n^4\mu_2^2\mu_3^2 - 18n^3\mu_2^2\mu_3^2 \\ &+33n^2\mu_2^2\mu_3^2 - 18n\mu_2^2\mu_3^2 \end{aligned}$$

---

61.

$$\begin{aligned} \mathrm{E}\left[\sum_{i=1}^{n} Y_i \left(\sum_{i=1}^{n} Y_i^2\right)^2 \sum_{i=1}^{n} Y_i^5\right] &= n\mu_{10} + n^2\mu_5^2 - n\mu_5^2 + 2n^2\mu_2\mu_8 - 2n\mu_2\mu_8 + 2n^2\mu_3\mu_7 \\ &-2n\mu_3\mu_7 + 2n^3\mu_2\mu_3\mu_5 - 6n^2\mu_2\mu_3\mu_5 + 4n\mu_2\mu_3\mu_5 \\ &+n^2\mu_4\mu_6 - n\mu_4\mu_6 + n^3\mu_2^2\mu_6 - 3n^2\mu_2^2\mu_6 + 2n\mu_2^2\mu_6 \end{aligned}$$

---

62.

$$\begin{aligned} \mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i^3\right)^3 \sum_{i=1}^{n} Y_i\right] &= n\mu_{10} + 3n^2\mu_3\mu_7 - 3n\mu_3\mu_7 + 3n^2\mu_4\mu_6 - 3n\mu_4\mu_6 \\ &+3n^3\mu_3^2\mu_4 - 9n^2\mu_3^2\mu_4 + 6n\mu_3^2\mu_4 \end{aligned}$$

---

63.

$$\begin{aligned} \mathrm{E}\left[\sum_{i=1}^{n} Y_i \sum_{i=1}^{n} Y_i^2 \sum_{i=1}^{n} Y_i^3 \sum_{i=1}^{n} Y_i^4\right] &= n\mu_{10} + n^2\mu_2\mu_8 - n\mu_2\mu_8 + 2n^2\mu_3\mu_7 - 2n\mu_3\mu_7 \\ &+2n^2\mu_4\mu_6 - 2n\mu_4\mu_6 + n^2\mu_5^2 - n\mu_5^2 + n^3\mu_2\mu_3\mu_5 \\ &-3n^2\mu_2\mu_3\mu_5 + 2n\mu_2\mu_3\mu_5 + n^3\mu_2\mu_4^2 - 3n^3\mu_2\mu_4^2 \\ &+2n\mu_2\mu_4^2 + n^3\mu_3^2\mu_4 - 3n^2\mu_3^2\mu_4 + 2n\mu_3^2\mu_4 \end{aligned}$$

---

64.

$$\begin{aligned} \mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i^2\right)^5\right] &= n\mu_{10} + 5n^2\mu_2\mu_8 - 5n\mu_2\mu_8 + 10n^2\mu_4\mu_6 - 10n\mu_4\mu_6 + 10n^3\mu_2^2\mu_6 \\ &-30n^2\mu_2^2\mu_6 + 20n\mu_2^2\mu_6 + 15n^3\mu_2\mu_4^2 - 45n^2\mu_2\mu_4^2 + 30n\mu_2\mu_4^2 \end{aligned}$$

$$+10n^4\mu_2^3\mu_4 - 60n^3\mu_2^3\mu_4 + 110n^2\mu_2^3\mu_4 - 60n\mu_2^3\mu_4$$
$$+n^5\mu_2^5 - 10n^4\mu_2^5 + 35n^3\mu_2^5 - 50n^2\mu_2^5 + 24n\mu_2^5$$

---

65.

$$\mathrm{E}\left[\left(\sum_{i=1}^n Y_i^2\right)^3 \sum_{i=1}^n Y_i^4\right] = n\mu_{10} + 4n^2\mu_4\mu_6 - 4n\mu_4\mu_6 + 3n^2\mu_2\mu_8 - 3n\mu_2\mu_8 + 3n^3\mu_2\mu_4^2$$
$$-9n^2\mu_2\mu_4^2 + 6n\mu_2\mu_4^2 + 3n^3\mu_2^2\mu_6 - 9n^2\mu_2^2\mu_6 + 6n\mu_2^2\mu_6$$
$$+n^4\mu_2^3\mu_4 - 6n^3\mu_2^3\mu_4 + 11n^2\mu_2^3\mu_4 - 6n\mu_2^3\mu_4$$

---

66.

$$\mathrm{E}\left[\left(\sum_{i=1}^n Y_i^2\right)^2 \sum_{i=1}^n Y_i^6\right] = n\mu_{10} + n^2\mu_4\mu_6 - n\mu_4\mu_6 + 2n^2\mu_2\mu_8 - 2n\mu_2\mu_8$$
$$+n^3\mu_2^2\mu_6 - 3n^2\mu_2^2\mu_6 + 2n\mu_2^2\mu_6$$

---

67.

$$\mathrm{E}\left[\left(\sum_{i=1}^n Y_i^2\right)^2 \left(\sum_{i=1}^n Y_i^3\right)^2\right] = n\mu_{10} + 2n^2\mu_2\mu_8 - 2n\mu_2\mu_8 + n^2\mu_4\mu_6 - n\mu_4\mu_6 + n^3\mu_2^2\mu_6$$
$$-3n^2\mu_2^2\mu_6 + 2n\mu_2^2\mu_6 + 2n^2\mu_3\mu_7 - 2n\mu_3\mu_7 + 2n^2\mu_5^2 - 2n\mu_5^2$$
$$+4n^3\mu_2\mu_3\mu_5 - 12n^2\mu_2\mu_3\mu_5 + 8n\mu_2\mu_3\mu_5 + n^3\mu_3^2\mu_4 - 3n^2\mu_3^2\mu_4$$
$$+2n\mu_3^2\mu_4 + n^4\mu_2^2\mu_3^2 - 6n^3\mu_2^2\mu_3^2 + 11n^2\mu_2^2\mu_3^2 - 6n\mu_2^2\mu_3^2$$

---

68.

$$\mathrm{E}\left[\sum_{i=1}^n Y_i^2 \sum_{i=1}^n Y_i^3 \sum_{i=1}^n Y_i^5\right] = n\mu_{10} + n^2\mu_5^2 - n\mu_5^2 + n^2\mu_3\mu_7 - n\mu_3\mu_7 + n^2\mu_2\mu_8 - n\mu_2\mu_8$$
$$+n^3\mu_2\mu_3\mu_5 - 3n^2\mu_2\mu_3\mu_5 + 2n\mu_2\mu_3\mu_5$$

---

69.

$$\mathrm{E}\left[\sum_{i=1}^n Y_i^2 \left(\sum_{i=1}^n Y_i^4\right)^2\right] = n\mu_{10} + n^2\mu_2\mu_8 - n\mu_2\mu_8 + 2n^2\mu_4\mu_6 - 2n\mu_4\mu_6$$
$$+n^3\mu_2\mu_4^2 - 3n^2\mu_2\mu_4^2 + 2n\mu_2\mu_4^2$$

---

70.

$$\mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i^3\right)^2 \sum_{i=1}^{n} Y_i^4\right] = n\mu_{10} + n^2\mu_4\mu_6 - n\mu_4\mu_6 + 2n^2\mu_3\mu_7 - 2n\mu_3\mu_7 + n^3\mu_3^2\mu_4 - 3n^2\mu_3^2\mu_4 + 2n\mu_3^2\mu_4$$

---

# Bibliography


Allison, J. S. (2008). *Bootstrap based hypothesis testing*, PhD thesis, North-west University, Potchefstroom Campus, South Africa.

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

Arcones, M. A. (1991). *On the asymptotic theory of the bootstrap*, PhD thesis, City University of New York.

Arcones, M. A. and Giné, E. (1989). The bootstrap of the mean with arbitrary bootstrap sample size, *Annales de l'Institute Henri Poincaré* **25**(4): 457 – 481.

Arcones, M. A. and Giné, E. (1992). On the bootstrap of $U$ and $V$ statistics, *The Annals of Statistics* **20**(2): 655–674.

Athreya, K. B. (1987). Bootstrap of the mean in the infinite variance case, *The Annals of Statistics* **15**(2): 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 Inferences* **58**: 299–320.

Babu, G. J. (1984). Bootstrapping statistics with linear combinations of chi-squares as weak limit, *Sankhyā: The Indian Journal of Statistics A* **46**(1): 85–93.

Beran, R. (1988). Prepivoting test statistics: A bootstrap view of asymptotic refinements, *Journal of the American Statistical Association* **83**(403): 687–697.

Beran, R. (1997). Diagnosing bootstrap success, *Annals of the Institute of Statistical Mathematics* **49**(1): 1–24.

Beran, R. and Srivastava, M. S. (1985). Bootstrap tests and confidence regions for functions of a covariance matrix, *The Annals of Statistics* **13**(1): 95–115.

Berrendero, J. R. (2007). The bagged median and the bragged mean, *The American Statistician* **61**(4): 325–330.

Bhattacharya, R. N. and Ghosh, J. K. (1978). On the validity of the formal Edgeworth expansion, *The Annals of Statistics* **6**(2): 434 – 451.

Bickel, P. J. and Freedman, D. A. (1981). Some asymptotic theory for the bootstrap, *The Annals of Statistics* **9**(6): 1196 – 1217.

Bickel, P. J. and Ren, J.-J. (1996). The $m$ out of $n$ bootstrap and goodness of fit tests with doubly censored data, *Robust Statistics: data, analysis and computer intensive methods. Lecture Notes in Statistics* .(109): 34–47.

Bickel, P. J. and Sakov, A. (1999). On the choice of $m$ in the $m$ out of $n$ bootstrap in estimation problems, Unpublished (See Proceedings of the American Statistical Association, Sakov and Bickel, 1999).

Bickel, P. J. and Sakov, A. (2002). Extrapolation and the bootstrap, *Sankhyā: The Indian Journal of Statistics* **64**: 640 – 652. Series A, Pt. 3.

Bickel, P. J. and Sakov, A. (2008). On the choice of $m$ in the $m$ out of $n$ bootstrap and its applications to confidence bounds for extrema, *Statistica Sinica* **18**(3): 967–985.

Bickel, P. J. and Yahav, J. A. (1988). Richardson extrapolation and the bootstrap, *Journal of the American Statistical Association* **86**(402): 387–393.

Bickel, P. J., Götze, F. and van Zwet, W. R. (1997). Resampling fewer than $n$ observations: Gains, losses, and remedies for losses, *Statistica Sinica* **7**: 1 – 31.

Boos, D. D. and Brownie, C. (1989). Bootstrap methods for testing homogeneity of variances, *Technometrics* **13**: 69–82.

Boos, D. 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.

Breiman, L. (1994). Bagging predictors, *Technical Report 421*, University of California, Berkley, California.

Bretagnolle, J. (1983). Lois limites du Bootstrap de certaines fonctionnelles, *Annales de l'Institute Henri Poincaré, Section B* **19**(3): 281–296.

Bühlmann, P. (2002). Bootstraps for time series, *Statistical Science* **17**(1): 52–72.

Bühlmann, P. (2003). Bagging, subagging and bragging for improving some prediction algorithms, *in* M. G. Akritas and D. N. Politis (eds), *Recent Advances and Trends in Nonparametric Statistics*, Elsevier.

Buja, A. and Stuetzle, W. (2006). Observations on Bagging, *Statistica Sinica* **16**: 323–351.

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.

Chernick, M. R. (2008). *Bootstrap methods – A guide for practitioners and researchers*, Wiley Inter-science.

Cheung, K. Y. and Lee, S. M. S. (2005). Variance estimation for sample quantiles using the $m$ out of $n$ bootstrap, *Annals of the Institute of Statistical Mathematics* **57**(2): 279 – 290.

Chung, K. H. and Lee, S. M. S. (2001). Optimal bootstrap sample size in construction of percentile confidence bounds, *Scandanavian Journal of Statistics* **28**(1): 225 – 239.

Croux, C., Joossens, K. and Lemmens, A. (2007). Trimmed bagging, *Computational Statistics & Data Analysis* **52**: 362 – 368.

D'agostino, R. B. and Stephens, M. A. (1986). *Goodness-of-fit techniques*, Vol. 68 of *Statistics: textbooks and monographs*, Marcel Dekker, inc.

Datta, S. (1995). On a modified bootstrap for certain asymptotically nonnormal statistics, *Statistics and Probability Letters* **24**: 91–98.

Davidson, R. and MacKinnon, J. G. (2004). *Econometric Theory and Methods*, Oxford University Press, New York.

Davison, A. and Hinkley, D. (1997). *Bootstrap Methods and their Applications*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press.

Deheuvels, P., Mason, D. M. and Shorack, G. R. (1993). Some results on the influences of extremes on the bootstrap, *Annales de l'Institute Henri Poincaré* **29**(1): 83 – 103.

Dehling, H. and Mikosch, T. (1994). Random quadratic forms and the bootstrap for $U$-statistics, *Journal of Multivariate Analysis* **51**(2): 392–413.

del Barrio, E., Cuesta-Albertos, J. A. and Matrán, C. (2002). Asymptotic stability of the bootstrap sample mean, *Stochastic Processes and their Applications* **97**: 289–306.

DiCiccio, T. J. and Efron, B. (1996). Bootstrap confidence intervals, *Statistical Science* **11**(3): 189–228.

Dodd, L. E. and Korn, E. L. (2007). The bootstrap variance of the square of a sample mean, *The American Statistician* **61**(2): 127 – 131.

Dümbgen, L. (1993). On nondifferentiable functions and the bootstrap, *Probability Theory and Related Fields* **95**: 125–140.

Efron, B. (1979). Bootstrap methods: Another look at the Jackknife, *The Annals of Statistics* **7**(1): 1 – 26.

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

Efron, B. (1992). Jackknife-after-bootstrap standard errors and influence functions, *Journal of the Royal Statistical Society* **54**(1): 83–127.

Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap*, number 57 in *Monographs on Statistics and Applied Probability*, Chapman & Hall.

Fisher, N. I. and Hall, P. (1990). On bootstrap hypothesis testing, *Australian & New Zealand Journal of Statistics* **32**(2): 177–190.

Fukuchi, J. I. (1994). *Bootstrapping extremes of random variables*, PhD thesis, Iowa State University.

Giné, E. and Zinn, J. (1989). Necessary conditions for the bootstrap of the mean, *The Annals of Statistics* **17**(2): 684–691.

Götze, F. and Račkauskas, A. (2001). Adaptive choice of bootstrap sample sizes, *State of the art in probability and statistics*, 36, I.M.S. publications, I.M.S. lectures notes, pp. 286 – 309.

Hall, P. (1990). Asymptotic properties of the bootstrap for heavy-tailed distributions, *The Annals of Probability* **18**(3): 1342–1360.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*, Springer Series in Statistics, Springer-Verlag.

Hall, P. and Martin, M. A. (1988). Exact convergence rate of bootstrap quantile variance estimator, *Probability Theory and Related Fields* **80**: 261–268.

Hall, P. and Wilson, S. R. (1991). Two guidelines for bootstrap hypothesis testing, *Biometrics* **47**(2): 757–762.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution, *The Annals of Mathematical Statistics* **19**(3): 293–325.

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.

Janssen, P. (1997). Bootstrapping *U*-statistics, *South African Statistical Journal* **31**: 185–216.

Jiménez-Gamero, M. D., Muñoz-García, J. and Pino-Mejías, R. (2003). Bootstrapping parameter estimated degenerate *U* and *V* statistics, *Statistics & Probability Letters* **61**: 61–70.

Knight, K. (1989). On the bootstrap of the sample mean in the infinite variance case, *The Annals of Statistics* **17**(3): 1168–1175.

Kotze, S., Johnson, N. L. and Read, C. B. (eds) (1983). *Encyclopedia of Statistical Sciences*, Vol. 3: Faà di Bruno Formula – Hypothesis testing, Wiley-Interscience.

Lee, A. J. (1990). *U-Statistics*, Vol. 110 of *STATISTICS: Textbooks and monographs*, Marcel Dekker.

Mammen, E. (1992). *When does bootstrap work? Asymptotic results and simulations*, Lecture Notes in Statistics, Springer-Verlag.

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.

Ngatchou-Wandji, J. (2006). On testing for the nullity of some skewness coefficients, *International Statistics Review* **74**(1): 47 – 65.

Polansky, A. M. (1999). Upper bounds on the true convergence of bootstrap percentile type confidence intervals, *The American Statistician* **53**(4): 362–369.

Politis, D. N. and Romano, J. P. (1994). Large sample confidence regions based on subsamples under minimal assumptions, *The Annals of Statistics* **22**(4): 2031–2050.

Politis, D. N., Romano, J. P. and Wolf, M. (1999). *Subsampling*, Springer-Verlag.

Politis, D. N., Romano, J. P. and Wolf, M. (2001). On the asymptotic theory of subsampling, *Statistica Sinica* **11**: 1105 – 1124.

Putter, H. and van Zwet, W. R. (1996). Resampling: Consistency of substitution estimators, *The Annals of Statistics* **24**(6): 2297–2318.

Romano, J. P. (1988). Bootstrapping the mode, *Annals of the Institute of Statistical Mathematics* **40**(3): 565–586.

Sakov, A. (1998). *Using the m out of n Bootstrap in hypothesis testing*, PhD thesis, University of California, Berkley.

Samworth, R. (2003). A note on methods of restoring consistency to the bootstrap, *Biometrika* **90**(4): 985 – 990.

Serfling, R. J. (1980). *Approximation theorems of mathematical statistics*, Wiley, New York.

Shao, J. (1994). Bootstrap sample size in nonregular cases, *Proceedings of the American Mathematical Society* **122**(4): 1251 – 1262.

Shao, J. and Tu, D. (1995). *The Jackknife and Bootstrap*, Springer Series in Statistics, Springer - Verlag.

Singh, K. (1981). On the asymptotic accuracy of Efron's bootstrap, *The Annals of Statistics* **9**(6): 1187 – 1195.

Swanepoel, J. W. H. (1986). A note on proving that the (modified) bootstrap works, *Communications in Statistics – Theory and Methods* **15**(11): 3193 – 3203.

Swanepoel, J. W. H. (1988). A new nonparametric estimation technique under asymmetry using the bootstrap, *Technical report*, Potchefstroom University for Christian Higher Education.

Swanepoel, J. W. H. (1990). A review of bootstrap methods, *South African Statistical Journal*.

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. (1990). On the asymptotic properties of the Jackknife histogram, *The Annals of Statistics* **18**(3): 1438–1452.

Wu, W., Carlstein, E. and Cambanis, S. (1989). Bootstrapping the sample mean for data with infinite variance., *Technical Report 296*, North Carolina University at Chapel Hill Center for Stochastic Processes.

Young, A. (1986). Conditioned data-based simulations: Some examples from geometrical statistics, *International Statistical Review / Revue Internationale de Statistique* **54**(1): 1–13.