POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Importance Resampling: two new algorithms for Bootstrap Estimation in high dimensions

Tesi di Laurea Magistrale in
Mathematical Engineering - Ingegneria Matematica

Author: **Alfredo Gimenez Zapiola**

Student ID: 974770
Advisor: Prof. Simone Vantini
Academic Year: 2022-23

Abstract

The Bootstrap is well-known nonparametric method with a wide range of applications. It is exploited to estimate statistics and related quantities of interest based on a random sample. In practice, a Monte Carlo estimate of the Bootstrap estimate is utilised, which implies an additional error layer. Importance Resampling, a variance reduction technique, is applied in this thesis to the Monte Carlo estimates of a given quantile of statistics that are used in high-dimensional data to construct Simultaneous Confidence Bands (SCB)s. In the original works of Johns (1988), Do and Hall (1991) and Davison (1988), a technique known as Exponential tilting is used for such task, which fails in this context. We propose two new algorithms, namely Contribution Tilted Mixture (CTM) and Loss Tilting (LT), and show through a simulation study effectively reduce the variance of the Monte Carlo estimate of the Bootstrap estimate for statistics used in SCB construction, demonstrating it through a simulation study. We also run a brief experiment to show the need of Importance Resampling when using the Bootstrap for SCBs.

Keywords: Bootstrap, importance resampling, Simultaneous Confidence Bands, exponential tilting, nonparametric delta method, functional data

Abstract in lingua italiana

Il Bootstrap è un noto metodo non parametrico con una vasta gamma di applicazioni. Viene sfruttato per stimare statistiche e relative quantità d'interesse sulla base di un campione casuale. In pratica, viene utilizzata una stima Monte Carlo della stima Bootstrap, il che implica un ulteriore livello di errore. Il ricampionamento dell'importanza, una tecnica di riduzione della varianza, viene applicato in questa tesi alla stima Monte Carlo di un dato quantile di statistiche che vengono utilizzate in dati ad alta dimensione per costruire bande di confidenza simultanee (SCB). Nei lavori originali di Johns (1988), Do and Hall (1991) e Davison (1988), viene utilizzata una tecnica nota come tilting esponenziale per tale compito, che fallisce in questo contesto. Proponiamo due nuovi algoritmi, ovvero Contribution Tilted Mixture (CTM) e Loss Tilting (LT), e mostriamo tramite uno studio di simulazione che riduce efficacemente la varianza della stima Monte Carlo della stima Bootstrap per statistiche che si usano nella costruzione di SCB, dimostrandolo tramite uno studio di simulazione. Eseguiamo anche un breve esperimento per mostrare la necessità del ricampionamento dell'importanza quando si utilizza Bootstrap per SCBs.

Parole chiave: bootstrap, ricampionamento d'importanza ,bande di confidenza simultanee, tilting esponenziale, metodo delta nonparametrico, dati funzionali

Contents

Introduction

This thesis concerns a noticeably useful device in Nonparametric Statistics: the Bootstrap.
Originally proposed by Efron (1979), it is a nonparametric tool with a wide range of applications. It allows the estimation of a statistic and quantities related to it, starting from a single random sample and sampling with replacement (resampling) from it. It has consequently been exploited in the construction of confidence intervals (DiCiccio and Efron (1996)), hypothesis testing (Davison and Hinkley (1997)) and most recently in the construction of Simultaneous Confidence Bands (SCBs) for functional data (Degras (2009), Telschow and Schwartzman (2022), Bunea et al. (2011), and many others), amongst several other applications.
Its relevance relies on the fact that it requires minimal assumptions on the underlying distribution of the sample, is simple to apply, has an asymptotic validity and at the same time provides estimates of quantities which otherwise would be cumbersome to obtain with the same accuracy.
The estimation of a random quantity through the Bootstrap method almost always requires a Monte Carlo simulation, which adds another error layer: in practice, the Monte Carlo estimate of the Bootstrap estimate is used. The mitigation of the first error, that is, the Monte Carlo error, which is the scope of this thesis, has been of great interest for the contributors of the Bootstrap, and several variance reduction techniques (efficiency improvement) have been researched.
On the one hand, there are methods that are exclusive to the Bootstrap world. Efron (1990) provides several techniques to reduce the variance of the Monte Carlo (MC) estimate of the Bootstrap approximation through *a posteriori* calculations, as well as diagnostics for the quality of the MC estimate. Davison and Hinkley (1988) directly avoid such simulation through the so-called Saddlepoint Methods to yield the Bootstrap estimate.
On the other hand, there are methods belonging to the more general Monte Carlo world adapted to the particular case of the Bootstrap estimate calculation. These are Antithetic Sampling, Control Variates and Importance Sampling, which are explained in textbooks

Hall (1992) and Davison and Hinkley (1997).
In this thesis, we restrict our attention to Importance Sampling, which in the case of the Bootstrap is called Importance Resampling. It was originally proposed by Johns (1988), with quite strong assumptions on the statistic of interest and further developed by Do and Hall (1991), who provided an empirical version, and Davison (1988) with a slight modification to the other two proposals. What these articles share is the use of Exponential Tilting, a method used to shift the Bootstrap distribution of the statistic of interest to get quantile estimates with lower variance.
Given that very few papers have been written recently on the subject matter, this topic could be asserted to be in the consolidation stage. However, with the advent of high-dimensional data (*e.g.* functional data: Ramsay and Silverman (2005)), calculations have become more computationally intensive, and the problem has regained importance.
Consequently, we restrict our attention to one particular use case of the Bootstrap method: the construction of Simultaneous Confidence Bands (SCBs) (*i.e.*, confidence regions for highly dimensional data).
The current reference paper on the subject matter is Degras (2011), which utilises a point-wise Student's t statistic to generalise the confidence intervals as constructed in DiCiccio and Efron (1996) to the multivariate case. This scope of this thesis is to apply Importance Resampling to estimate SCBs that utilise statistics of this type.
The main contribution of this thesis are two new algorithms, namely **Contribution Tilted Mixture** (CTM) and **Loss Tilting**, which as will be shown on Chapter 2 contrary to Exponential Tilting do increase the efficiency when working with high-dimensional data. As a matter of fact, as it will be illustrated in Chapter 2, Exponential Tilting fails for Degras (2009)-like statistics when the dimensionality increases.
The remainder of this thesis is organised as follows:

1. In Chapter 1 the concepts of the Bootstrap, Importance Sampling and the so-called Nonparametric Delta method, a device used in Exponential tilting, as well as high dimensional data are explained.

2. Chapter 2 focuses on Importance Resampling. We show how it can be used for quantile estimation, a task necessary for building SCBs. Exponential tilting is explained, and the new algorithms, namely **Loss tilting** and **Contribution Tilting Mixture** are described. A simulation study is carried out focusing on the application of Importance Resampling for the construction of SCBs and the three algorithms (Exponential Tilting, Loss Tilting and Contribution Tilted Mixture) are compared.

3. In Chapter 3 we provide an overview of the different methods present in the scientific community that exploit the Bootstrap for SCBs, and an experiment that

demonstrates the effect of the Monte Carlo error in such task is shown.

4. Appendix A provides an overview of the different variance reduction methods for the Monte Carlo approximation of the Bootstrap.

1 | Theoretical Background

1.1. The Bootstrap

As aforementioned, it is a nonparametric method (i.e. it makes minimal assumptions on the underlying distribution F of the data) that estimates the sampling distribution of a statistic using the observed data.
The underlying intuition is the following: given a sample $\mathbf{X} = \{\mathbf{x}_1, ..., \mathbf{x}_i, ..., \mathbf{x}_N\}$ where N denotes the sample size and each $\mathbf{x}_i$ is an i.i.d p-dimensional vector (p integer, possibly a high value), we want to estimate certain properties of the unknown underlying distribution of the sample F. Such property is usually a functional of F, and we can denote it as:

$$t(F) = \int_{\Omega} t(\mathbf{x}) f(\mathbf{x}) \, d\mathbf{x} \tag{1.1}$$

where Ω denotes the support of distribution F and f its probability density (mass in the discrete case) function, and the bold notation signifies a p-dimensional vector.
The reasoning behind the Bootstrap method is the following: given the sample $\mathbf{X}$, we assume the $ECDF$ (empirical cumulative distribution function) $\hat{F}$ approximates the actual CDF (cumulative distribution function) F. Hence, sampling i.i.d from $\hat{F}$, which in practice means sampling **with replacement** from the observed data, is approximately tantamount to sampling from F. Of course, such assumption becomes valid as N goes to $+\infty$, which is true thanks to the Law of Large Numbers. For details on this result and the associated rate of convergence, the reader is referred to Singh (1981).

Assumption 1 (The Bootstrap assumption)**.** *The ECDF $\hat{F}$ given by sample $\mathbf{X}$ approximates the CDF F of the underlying distribution of the data, so that the sampling distribution of a statistic can be estimated by sampling with replacement from $\hat{F}$, using the plug-in principle.*

A basic yet powerful application of the **Bootstrap** method could be the estimation of the mean squared error (MSE) of a sample estimator $\hat{t}(\hat{F})$, whose value is $\hat{\theta}$, of functional

$t(F)$, with value θ. We know that:

$$MSE(\hat{t}(\hat{F})) = MSE(\hat{\theta}) = \mathbb{E}_F[(\hat{\theta} - \theta)^2] = \mathbb{E}_F[(\hat{t}(\mathbf{X}) - t(F))^2] \tag{1.2}$$

where $\mathbf{X}$ denotes a sample drawn from F, and $\hat{t}$ is a sample estimator of t.
Since instead of the true F we only know the $\hat{F}$ given by the sample, we apply the **Bootstrap assumption** and estimate the MSE in the following way:

$$M\hat{S}E(\hat{\theta}) = \mathbb{E}_{\hat{F}}[(\hat{t}(\mathbf{X}^*) - \hat{t}(\hat{F}))^2] = \mathbb{E}_{\hat{F}}[(\hat{\theta}^* - \hat{\theta})^2] \tag{1.3}$$

where $\hat{t}(\mathbf{X}^*)$ denotes the value of statistic t evaluated on a Bootstrap sample (a sample with replacement) from originally available data $\mathbf{X}$, and $\hat{\theta}^*$ its value. Alternatively, notation $\hat{T}^*$ can be used for the same quantity.
If we expand Equation (1.3):

$$\mathbb{E}_{\hat{F}}[(\hat{\theta}^* - \hat{\theta})^2] = \int_{\Omega(\hat{F})} (\hat{t}(\mathbf{X}^*) - \hat{t}(\hat{F}))^2 \; d\hat{F}(\mathbf{X}^*) \tag{1.4}$$

with $\Omega(\hat{F})$ being the support of $d\hat{F}(\mathbf{X}^*)$, the *p.m.f.* (probability mass function) of each re-sample of $\hat{F}$. Note that $\Omega(\hat{F})$ is of finite counting measure, since it is the space of all possible distinct samples with replacement from original sample $\mathbf{X}$.
What is more, $d\hat{F}(\mathbf{X}^*)$ corresponds to a multinomial distribution:

$$d\hat{F}(\mathbf{X}^*) = \prod_{j=1}^{N} p_j^{f_j^*} \tag{1.5}$$

where f_j^* denotes the frequency of the jth statistical unit in re-sample $\mathbf{X}^*$, and p_j is naturally $\frac{1}{N}$, the probability of resampling under the so-called Ordinary Bootstrap.
However, as N increases, the cardinality of $\Omega(\hat{F})$ becomes too large to evaluate the quantity of interest on all the elements of its support. Therefore, the Bootstrap estimate (1.3) is approximated by its Monte Carlo estimate:

$$M\hat{S}E(\hat{\theta}) = \int_{\Omega(\hat{F})} (\hat{t}(\mathbf{X}^*) - \hat{t}(\hat{F}))^2 \, d\hat{F}(\mathbf{X}^*) \approx \frac{1}{B}\sum_{b=1}^{B}(\hat{t}(\mathbf{X}^{*\mathbf{b}}) - \hat{t}(\hat{F}))^2 =: MS\hat{\hat{E}}_{MC}(\hat{\theta}) \tag{1.6}$$

where $\mathbf{X}^{*\mathbf{b}}$ is a Monte Carlo (re)sample (*i. e.* a sample with replacement) from $\hat{F}$, B the number of Monte Carlo iterations, and we denote the Monte Carlo estimate of Bootstrap estimate $M\hat{S}E(\hat{\theta})$ with $MS\hat{\hat{E}}_{MC}(\hat{\theta})$ (the double hat notation helps give the intuition we are providing the MC estimate of a Bootstrap estimate).

Therefore, apart from the error implied in the Bootstrap Assumption (1), the Bootstrap estimate is affected by the **Monte Carlo** (MC) error as well.

Remark 1 (Monte Carlo error *versus* Bootstrap error)**.** *Let us denote μ the true value of a functional t of F we are estimating, namely:*

$$\mathbb{E}_F[t] = \mu \tag{1.7}$$

with sample estimator $\hat{t}$. Denote $\hat{\mu}$ as its Bootstrap estimate, and $\hat{\hat{\mu}}_{MC}$ the Monte Carlo estimate of $\hat{\mu}$. We distinguish the ***Bootstrap error****, given by*

$$\epsilon_B = \mu - \hat{\mu} \tag{1.8}$$

from the ***Monte Carlo error*** *(MC error), which arises from the approximation of the Bootstrap integral:*

$$\epsilon_{MC} = \hat{\mu} - \hat{\hat{\mu}}_{MC} = \int_{\Omega(\hat{F})} \hat{t}(\mathbf{X}^*)\, d\hat{F}(\mathbf{X}^*) - \frac{1}{B}\sum_{b=1}^{B} \hat{t}(\mathbf{X}^{*(\mathbf{b})}) \tag{1.9}$$

Thus, the Bootstrap estimate of any statistic of interest (for e.g. the MSE) will be subject to both the Bootstrap error and the MC error.

As mentioned before, reducing the MC error is the scope of this thesis. In particular, we focus on the technique named Importance Resampling, which is explained in the next subsection.

1.2. Importance Sampling and Importance Resampling

Importance Sampling is a technique developed to reduce the variance of the Monte Carlo estimate of statistics of the kind as Equation (1.1). In particular, whenever $t(\mathbf{x})$ and $f(\mathbf{x})$ are such that when one is of large value, the other one is small, which translates into MC iterations that contribute little to the integral (see Zio (2013)).
The idea is thus to sample instead of from $f(\mathbf{x})$, from the so-called Importance distribution $h(\mathbf{x})$ and adjusting the statistic $t(\mathbf{x})$ to compensate from the bias we get by sampling from

another distribution. Following Equation (1.1), we would have:

$$\mu := t(F) = \int_{\Omega} t(\mathbf{x})f(\mathbf{x})d(\mathbf{x}) = \int_{\Omega} t(\mathbf{x})\frac{f(\mathbf{x})}{h(\mathbf{x})}h(\mathbf{x})\ d\mathbf{x} \approx \frac{1}{B}\sum_{b=1}^{B} t(\mathbf{x}^{(b)})\frac{f(\mathbf{x}^{(b)})}{h(\mathbf{x}^{(b)})} =: \hat{\mu}_{IS} \tag{1.10}$$

where $\hat{\mu}_{IS}$ is the Importance Sampling estimate, and $\mathbf{x}^{(b)}$ the bth MC sample from $h(\mathbf{x})$. Moreover, the ratio $\frac{f(\mathbf{x})}{h(\mathbf{x})}$ is usually called the likelihood ratio, and denoted by:

$$w(\mathbf{x}) := \frac{f(\mathbf{x})}{h(\mathbf{x})} \tag{1.11}$$

Whereas the variance of the MC estimate under Ordinary MC, meaning when sampling from $f(\mathbf{x})$ is given by:

$$Var_F[\hat{\mu}_{MC}] = B^{-1}\Big(\int_{\Omega} t(\mathbf{x})^2 f(\mathbf{x})\ d\mathbf{x} - \mu^2\Big) \tag{1.12}$$

, the variance of the Importance Sampling estimate $\hat{\mu}_{IS}$ is

$$Var_H[\hat{\mu}_{IS}] = B^{-1}\Big(\int_{\Omega} t(\mathbf{x})^2 w(\mathbf{x})^2 h(\mathbf{x})\ d\mathbf{x} - \mu^2\Big) \tag{1.13}$$

We remark that the expected value of both the Ordinary and the Importance Sampling MC estimates is the same. Naturally, the idea is to choose $p.d.f.$ (probability density function) h with CDF H such that the variance (1.13) is smaller than (1.12). It can actually be proven (see Zio (2013)) that there exists an optimal importance distribution, which is proportional to $|h(\mathbf{x})|f(\mathbf{x})$. Nonetheless, finding such optimal importance distribution requires a similar computational effort to that of computing $\hat{\mu}_{MC}$, which is why in practice the choice of H becomes a rather creative and problem-specific task (Zio (2013)).

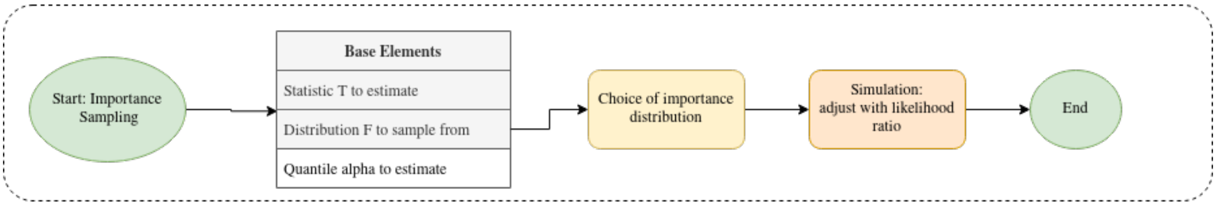

Figure 1.1: General Importance sampling algorithm to compute the quantile of a statistic
1

The special case of importance sampling applied to the Bootstrap method is called **Importance Resampling** (IR). Firstly escogitated by Johns (1988) and furtherly developed by Do and Hall (1991), Davison (1988), and Hinkley and Shi (1989), it is the main focus of chapter 2.

Let us first generalise the MSE example (1.4) for a general statistical function m with sample estimator $\hat{m}$ and have the general formula for the MC estimate $\hat{\hat{\mu}}_{MC}$ of Bootstrap estimate $\hat{\mu}$:

$$\hat{\mu} = \int_{\Omega(\hat{F})} \hat{m}(\mathbf{X}^*)\, d\hat{F}(\mathbf{X}^*) \approx B^{-1} \sum_{b=1}^{B} \hat{m}(\mathbf{X}^{*\mathbf{b}}) =: \hat{\hat{\mu}}_{MC} \tag{1.14}$$

where $d\hat{F}(\mathbf{X}^*)$ corresponds to a multinomial distribution:

$$d\hat{F}(\mathbf{X}^*) = \prod_{j=1}^{N} p_j^{f_j^*} \tag{1.15}$$

; f_j^* denotes the frequency of the jth statistical unit in re-sample $\mathbf{X}^*$, and p_j is $\frac{1}{N}$ under Ordinary MC (the analogue of $f(\mathbf{x})$ in (1.10)). In the case of **Importance Resampling**, the rationale is to alter the values p_j such that the variance of estimate $\hat{\hat{\mu}}_{IS}$ is lower than that of $\hat{\hat{\mu}}_{MC}$. Therefore, the IR estimate of $\hat{\mu}$ becomes:

$$\hat{\mu} = \int_{\Omega(\hat{F})} \hat{m}(\mathbf{X}^*)\, d\hat{F}(\mathbf{X}^*) = \int_{\Omega(\hat{F})} \hat{m}(\mathbf{X}^*) \frac{d\hat{F}(\mathbf{X}^*)}{dH(\mathbf{X}^*)}\, dH(\mathbf{X}^*) \tag{1.16}$$

where the likelihood ratio is:

$$w(\mathbf{X}^*) := \frac{d\hat{F}(\mathbf{X}^*)}{dH(\mathbf{X}^*)} \tag{1.17}$$

, and H is a multinomial distribution:

$$d\hat{H}(\mathbf{X}^*) = \prod_{j=1}^{N} g_j^{f_j^*} \tag{1.18}$$

where g_j are to be defined since they constitute the Importance distribution. The Monte Carlo approximation $\hat{\hat{\mu}}_{IR}$ (where IR stands for Importance Resampling):

$$\hat{\hat{\mu}}_{IR} := \frac{1}{B} \sum_{b=1}^{B} \hat{m}(\mathbf{X}^{*\mathbf{b}}) w(\mathbf{X}^{*\mathbf{b}}) \tag{1.19}$$

where $\mathbf{X}^{*\mathbf{b}}$ is a sample of H for each $b = 1, \dots, B$; and whose variance,

$$Var_H[\hat{\hat{\mu}}_{IR}] = B^{-1} \Big\{ \int_{\Omega(\hat{F})} (\hat{m})^2(\mathbf{X}^*) w^2(\mathbf{X}^*)\, dH(\mathbf{X}^*) - \mu^2 \Big\} \tag{1.20}$$

is hopefully lower than under Ordinary MC:

$$Var_H[\hat{\hat{\mu}}_{MC}] = B^{-1}\Big\{\int_{\Omega(\hat{F})} \hat{m}^2(\mathbf{X}^*)\, d\hat{F}(\mathbf{X}^*) - \mu^2\Big\} \tag{1.21}$$

Remark 2 (Bootstrap procedure under Importance Resampling). *At each MC iteration, we draw a sample with replacement from the original sample* $\mathbf{X}$ *of size* N*, but the probability of re-sampling the* j*th statistical unit is* g_j *(given by the importance distribution) instead of* $p_j = \frac{1}{N}$

What the classic papers (*i.e.* Johns (1988), Davison (1988) and Do and Hall (1991)) have in common is the fact that they recur to a technique named Exponential Tilting (shown in Chapter 2) to yield resampling probabilities g_j present in (1.18). As it will also be shown in Chapter (2), this is not an exclusive choice, and we provide indeed two new algorithms that work better than Exponential Tilting in the setting of SCB construction, as mentioned in the Introduction. Before moving to such explanation, we outline below the Nonparametric Delta method, the mechanism behind all three algorithms, namely Exponential Tilting, Loss Tilting and Contribution Tilted Mixture.

1.3. The Nonparametric Delta Method

The delta method is an approach to calculate Edgeworth expansions (see Hall (1992)), yet it is most commonly known for the case of the linearisation of a statistic, which is the case in Davison and Hinkley (1997), Do and Hall (1991), as well as in this thesis. It consists on the application of the Taylor series expansions to statistical functions, which indeed are nothing but operators mapping the space of probability functions to $\mathbb{R}$ (*i.e.* functionals).

In this work, we shall only consider the linear expansion.[2]. Suppose we have a statistic t valued at distribution CDF G_1, and we want to evaluate it at G_2. Then, we can exploit first order Taylor expansion, *id est*, we can **approximate it** with its linearised version t_L in the following way (see Davison and Hinkley (1997))

$$t_L(G_2) := t(G_1) + \int_{\Omega(G_2)} L_t(y; G_1)\, dG_2(y) \tag{1.22}$$

[2]The quadratic approximation of statistics is mentioned as a possibility in Davison and Hinkley (1997), but it is not utilised.

where $L_t(y; G_1)$ is the derivative of t evaluated at G_1, and is called the **influence function**:

$$L_t(x; G_1) = \lim_{\epsilon \to 0} \frac{t\Big((1-\epsilon)G_1 + \epsilon H_{\mathbf{x}}\Big) - t(G_1)}{\epsilon} = \frac{\partial t\Big((1-\epsilon)G_1 + \epsilon H_{\mathbf{x}}\Big)}{\partial \epsilon}\Bigg|_{\epsilon=0} \tag{1.23}$$

where $H_{\mathbf{x}}$ is the Heavyside function. Usually, it will be the case that instead of G_1, we only know its $ECDF$ through an available sample. That is, we deal with $L_t(y; \hat{G}_1)$ and we call it the **empirical influence function** $l(y)$, which evaluated at a particular value $y_j,\ j \in \{1, \ldots, N\}$ (N being the sample size), is called the empirical influence value $l_j = l(y_j)$.

Of course, when $G_1 = G_2$, the right term of (1.22) equals 0, that is:

$$\int_{\Omega(G_2)} L_t(y; G_1)\ dG_2(y) = 0 \tag{1.24}$$

(indeed the first order Taylor expansion of a function about a point (G_2) evaluated at that point (G_2) matches the function at such point (G_2)).

When we substitute G_2 with $\hat{G}_1$ (as said before, the $ECDF$ of the sample of size N), the delta method is called the nonparametric delta method, and we have:

$$t_L(\hat{G}_1) := t(G_1) + \int_{\Omega(\hat{G}_2)} L_t(y; G_1)\ d\hat{G}_1(y) = t(G_1) + n^{-1} \sum_{j=1}^{N} l_j \tag{1.25}$$

What is more, the Nonparametric Delta Method Result (see Davison and Hinkley (1997) and Wasserman (2006)) states that if statistic t is a smooth functional[3], then through the Central Limit Theorem, asymptotically:

$$t_L(\hat{G}_1) - t(G_1) \sim N(0, v_L(G_1)) \tag{1.26}$$

where $v_L(F)$ is the delta method variance:

$$v_L(G_1) = n^{-1} Var[L_t(Y)] = n^{-1} \int_{\Omega(G_1)} L_t^2(y) dG_1(y) \tag{1.27}$$

[3]More specifically, it should be Hadamard differentiable. Textbooks Davison and Hinkley (1997) and Hall (1992) are not that specific in the sense they only state it should be a smooth statistic. See Wasserman (2006) for details.

which is approximated by the sample version:

$$v_L(\hat{G}_1) := n^{-2} \sum_{i=1}^{N} l_i^2 \tag{1.28}$$

It remains a problem to know l_i. Whereas they could be calculated analytically (computing the derivative of the statistic w.r.t. each statistical unit), due to the nature of the statistics used for the construction of Bootstrap SCBs, in this thesis we focus on the empirical methods to estimate them, as it is shown in Chapter 2.

1.4. High-dimensional data

In Statistics, a usual notation is N for sample size, and p for the number of random variables (or dimensionality) per statistical unit (Secchi et al. (2013)). We talk about high-dimensional data whenever we are dealing with samples whose underlying distribution F is of large (and possibly infinite) p. In such situation, different models are available. A statistical unit from random sample $\mathbf{X}$ of distribution F of large p could be represented as a p-dimensional vector, namely $\mathbf{x} = \begin{bmatrix} x^{(1)} \\ \vdots \\ x^{(p)} \end{bmatrix} \in \mathbb{R}^p$, and the whole sample with matrix $\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^\intercal \\ \vdots \\ \mathbf{x}_N^\intercal \end{bmatrix}$.

Another possibility is provided in Functional Data Analysis (FDA). Under such model, each statistical unit is a function, for *e.g.* with domain $\mathcal{I} = (a, b) \subset \mathbb{R}$; $a, b \in \mathbb{R} : b > a$, we assume statistical unit $\mathbf{x}$ is a function of the L^q (for a given $q \in \mathbb{N}^+$) space on $\mathcal{I}$, and the random sample $\mathbf{X}$ contains a discrete set of observations for each statistical unit. For convenience, we will use the following notation: let $\mathcal{S}$ be the set of random variables (dimensions) of each statistical unit. Thus, in the univariate case it will be a singleton; in the multivariate case $\mathcal{S} = \{1, ..., p\}$ and in the functional case $\mathcal{S} = (a, b)$.

In this thesis, we are interested in the high-dimensional Bonferroni simultaneous confidence intervals for the the mean $\mu(s)$, $s \in \mathcal{S}$ of F, which in FDA are called called Simultaneous Confidence Band (SCBs).
We mainly focus on the method proposed by Degras (2009): the objective is to obtain

bands of the form:

$$[\hat{\mu}(s) - c_\alpha \frac{\hat{\sigma}(s)}{\sqrt{N}},\ \hat{\mu}(s) + c_\alpha \frac{\hat{\sigma}(s)}{\sqrt{N}}],\ s \in \mathcal{S} \tag{1.29}$$

with $\hat{\mu}(s)$ being an element-wise (point-wise), i.e. $\forall s \in \mathcal{S}$ estimate of the mean; $\hat{\sigma}(s)$ the element-wise scale estimator; and c_α is the so-called critical value, to be estimated, such that:

$$\mathbb{P}\Big\{\mu(s) \in \Big[\hat{\mu}(s) - c_\alpha \frac{\hat{\sigma}(s)}{\sqrt{N}},\ \hat{\mu}(s) + c_\alpha \frac{\hat{\sigma}(s)}{\sqrt{N}}\Big]\Big\} \approx 1 - \alpha \tag{1.30}$$

where $1-\alpha$ is the confidence level. Moreover, c_α is the $(1-\alpha)$th quantile of the following statistic:

$$\sup_{s \in \mathcal{S}} |\sqrt{N} \frac{\hat{\mu}(s) - \mu(s)}{\hat{\sigma}(s)}| \tag{1.31}$$

which of course corresponds to the L^∞ norm of $|\sqrt{N}\frac{\hat{\mu}(s)-\mu(s)}{\hat{\sigma}(s)}|$. Degras (2011) also provides its Bootstrap estimate:

$$\sup_{s \in \mathcal{S}} |\sqrt{N} \frac{\hat{\mu}^*(s) - \hat{\mu}(s)}{\hat{\sigma}^*(s)}| \tag{1.32}$$

(where, as said before $\hat{\mu}^*(s)$ is the value of the sample estimator on a Bootstrap sample), which in turn is to be estimated via Monte Carlo. We remark that apart from several assumptions (Degras (2011)), SCBs as in Equation (1.29) assume symmetry of the stochastic process $\sqrt{N}\frac{\hat{\mu}(s)-\mu(s)}{\hat{\sigma}(s)}$. Being coherent with Bootstrap T confidence intervals (Efron and Tibshirani (1993)), we build SCBs as:

$$[\hat{\mu}(s) - c_\alpha^+ \frac{\hat{\sigma}(s)}{\sqrt{N}},\ \hat{\mu}(s) - c_\alpha^- \frac{\hat{\sigma}(s)}{\sqrt{N}}],\ s \in \mathcal{S} \tag{1.33}$$

where c_α^+ is the $\frac{1-\alpha}{2}$th quantile of:

$$\sup_{s \in \mathcal{S}} \sqrt{N} \frac{\hat{\mu}(s) - \mu(s)}{\hat{\sigma}(s)} \tag{1.34}$$

and c_α^- is the $\frac{\alpha}{2}$th quantile of:

$$\inf_{s \in \mathcal{S}} \sqrt{N} \frac{\hat{\mu}(s) - \mu(s)}{\hat{\sigma}(s)} \tag{1.35}$$

whose Bootstraps estimates are, respectively:

$$\sup_{s \in \mathcal{S}} \sqrt{N} \frac{\hat{\mu}^*(s) - \hat{\mu}(s)}{\hat{\sigma}^*(s)};\ \inf_{s \in \mathcal{S}} \sqrt{N} \frac{\hat{\mu}^*(s) - \hat{\mu}(s)}{\hat{\sigma}^*(s)} \tag{1.36}$$

In addition (this will be useful to evaluate Importance Resampling in Chapter 2), we also generalise reverse percentile confidence intervals, (known as basic Bootstrap confidence limits in Davison and Hinkley (1997), see also Hesterberg (2014)) to obtain SCBs , that is:

$$[\hat{\mu}(s) - q_\alpha^+,\ \hat{\mu}(s) - q_\alpha^-],\ s \in \mathcal{S} \tag{1.37}$$

such that:

$$\mathbb{P}\Big\{\mu(s) \in \Big[\hat{\mu}(s) - q_\alpha^+,\ \hat{\mu}(s) - q_\alpha^-\Big]\Big\} \approx \alpha \tag{1.38}$$

where q_α^+ is the $\frac{1-\alpha}{2}$th quantile of the *sup* of the so-called element-wise bias[4]:

$$\sup_{s \in \mathcal{S}} \ \hat{\mu}(s) - \mu(s) \tag{1.39}$$

and q_α^- is the $\frac{\alpha}{2}$th quantile of:

$$\inf_{s \in \mathcal{S}} \ \hat{\mu}(s) - \mu(s) \tag{1.40}$$

with Bootstrap estimates, respectively

$$\sup_{s \in \mathcal{S}} \ \hat{\mu}^*(s) - \hat{\mu}(s); \ \ \inf_{s \in \mathcal{S}} \ \hat{\mu}^*(s) - \hat{\mu}(s) \tag{1.41}$$

We justify such choices and overview the different methods to build SCBs through Bootstrap estimates in Section 3.1.

[4]This name is used in Davison and Hinkley (1997). Another possible name could be "difference" instead of "bias".

2 | Importance resampling

In this chapter, the general algorithm to perform Importance Resampling is explained. To yield an importance resampling distribution, we explain Exponential Tilting, which is used in the classic papers: Johns (1988), Do and Hall (1991) and Davison (1988); as well as our two proposals: Contribution Tilted Mixture (CTM) and Loss Tilting (LT). As uttered above, in this thesis we concentrate in the estimation of statistics that are used in the construction of SCBs through the Bootstrap. In particular, this will imply estimating the $(1-\alpha)$th quantile, where α is the desired confidence level, such as (1.41) and (1.36). We remark that Importance Resampling is not the only method to reduce the variance of an MC estimate. Indeed, a plethora of approaches are available to obtain a lower variance than Equation (1.21) for a Bootstrap estimate. In **Appendix A** an overview of the techniques to make the MC estimate of the Bootstrap estimate more efficient (*i.e.* with less variance w.r.t crude MC) is provided. The reason for which we choose Importance Resampling is that it yields the highest efficiencies for tail probability (and thus extreme quantile estimation) amongst the available techniques (see Davison and Hinkley (1997)).

2.1. Importance Resampling for Bootstrap Quantile Estimation

We have been over the fact Importance Resampling provides MC estimates with less variance than Ordinary Monte Carlo, as long as an adequate importance disstribution is given. In this section, we present the core focus of this research, which is the use of this variance reduction technique to yield Bootstrap estimates for quantiles of a statistic, later employed in the construction of simultaneous confidence regions (SCBs).

Once the desired quantile level α is chosen, one must reason about the choice of the importance distribution H.

We first view the tail probability estimation. Given a statistic $T = t(\mathbf{x})$, where $CDF\ F$, $\mathbf{x} \in \Omega(F)$, of the form (1.1), for a given θ, we are interested in estimating:

$$\pi = \mathbb{P}_F[T \leq \theta] \tag{2.1}$$

for which we use the Bootstrap estimate from available sample $\mathbf{X}$ of F of size N and $ECDF$ $\hat{F}$, and sample estimator for T being $\hat{T} = \hat{t}(\mathbf{X})$ (with $\mathbf{X}$ a random sample of size N)

$$\hat{\pi} = \mathbb{P}_{\hat{F}}[\hat{t}^*(\mathbf{X}^*) \leq \theta] = \mathbb{P}_{\hat{F}}[\hat{T}^* \leq \theta] \tag{2.2}$$

denoting with $\hat{T}^*$ the Bootstrap distribution of the statistic conditioned to the available sample.
To obtain the MC estimate of (2.2), the function to use in the Bootstrap integral (1.14) is:

$$\hat{m}(\mathbf{X}^*) = \mathbb{1}\{\hat{t}(\mathbf{X}^*) \leq \theta\} \tag{2.3}$$

Whence one can realise that when the tail probability is being estimated, most Bootstrap iterations will be 0 when $\theta > t(\mathbf{X}^*)$ when the order of the quantile α is $\ll 0.5$, meaning they will not contribute to the integral (an analogue argument holds when the order of the quantile is $\alpha \gg 0.5$). Consequently, an idea is to recenter it at the tail value θ (or quantile say $\alpha = 0.05$), *"so that the estimation of the tail quantile becomes more like estimating the median"* (Johns (1988)), provided we adjusted with the likelihood ratio from sampling from this other distribution, an unbiased estimator with less variance of the tail probability (quantile) would be obtained.
In other words, when dealing with tail probabilities or extreme quantiles, (2.3) will have the same value in too many MC iterations, which translates into low efficiency. Indeed, the MC variance of such quantity is given by (plug (2.2) into (1.21)):

$$Var[\hat{\hat{\pi}}_{MC}] = \frac{\hat{\pi}(1-\hat{\pi})}{B} \tag{2.4}$$

(where B is the number of MC iterations) which implies the farther the tail probability (quantile order) is from 0.5, the more inefficient Ordinary Monte Carlo is. This of course arouses the need of Importance Resampling.
Quantile estimation, however, requires an extra step to be performed with Importance Resampling (see Hall (1992)). Letting α be order of the quantile of the statistic to estimate, the quantity of interest $\hat{\xi}_\alpha$ is the solution to :

$$\mathbb{P}_{\hat{F}}(\hat{T}^* \leq \hat{\xi}_\alpha) = \alpha \tag{2.5}$$

and we can denote the Bootstrap estimate of the CDF of functional T by:

$$\hat{Q}(y) = \mathbb{P}_{\hat{F}}(\hat{T}^* \leq y) \tag{2.6}$$

[1]

Remark 3 (Discreteness of $\hat{\hat{Q}}_{MC}(y)$)**.** *The support of the MC estimate of $\hat{Q}(y)$, that is $\hat{\hat{Q}}_{MC}(y)$ is of exactly B counting measure, with B being the number of MC iterations. Indeed:*

$$Supp[\hat{\hat{Q}}_{MC}(y)] = \{\hat{t}(\mathbf{X}^{*\mathbf{b}}),\ b = 1, \ldots, B\} \tag{2.7}$$

From Equation (2.6) it easy to check that the Bootstrap quantile estimate will be given by:

$$\hat{\xi}_\alpha = \inf\{y : \hat{Q}(y) \geq \alpha\} \tag{2.8}$$

and the MC estimate of such estimate:

$$\hat{\hat{\xi}}_{MC,\alpha} = \inf\{y^{(b)} : \hat{\hat{Q}}_{MC}(y^{(b)}) \geq \alpha\ ,\ b = 1, \ldots, B\} \tag{2.9}$$

Under Importance Resampling, the fact that different weights (see Equation 1.17) are associated to each Bootstrap replicate has to be taken into account. Thus, the values of the statistic evaluated at each Bootstrap sample are ordered: $\hat{T}_1^* < \ \ldots\ < \hat{T}_B^*$ and have corresponding weights $w_1^* < \ \ldots\ < w_B^*$.
When the order of the quantile of interest α is < 0.5, the Importance Resampling estimate is $\hat{T}_M^*$, where M is **such that** (see Davison and Hinkley (1997) or Johns (1988)):

$$\frac{1}{B}\sum_{b=1}^{M} w_b^* \leq \alpha < \frac{1}{B}\sum_{b=1}^{M+1} w_b^* \tag{2.10}$$

However, care must be taken when the order α of the quantile is such that $\alpha > 0.5$, and the following estimate is used:

$$\frac{1}{B}\sum_{b=M}^{B} w_b^* \leq 1 - \alpha < \frac{1}{B}\sum_{b=M+1}^{B} w_b^* \tag{2.11}$$

Remark 4 (Importance Resampling when the quantile is $\alpha > 0.5$)**.** *In such case, estimating (2.10) would actually lead to a loss of efficiency with respect to ordinary Resampling. Indeed, if the Boostrap distribution T^* is tilted to the right, then the likelihood ratio $w(\mathbf{X}^*) = \frac{d\hat{F}(\mathbf{X}^*)}{dH(\mathbf{X}^*)}$ (1.17) might explode for the lowest values of T^*: the further the importance distribution of T^* is tilted to the right to be centered at the desired quantile, the lower of the denominator is for its lowest values, and their associated weights skyrocket.*

[1]Note it is the function whose efficiency through Importance Resampling is shown in figure A.1

Thus, it is better not to include those weights in the estimation of the upper ($\alpha \gg 0.5$ quantile by estimating $1 - \hat{Q}(y) = \mathbb{P}_{\hat{F}}[\hat{T}^* > y]$, and utilise the estimate in (2.11) This will be visualised in section 2.1

2.2. Choice of the Importance Resampling Distribution

2.2.1. Exponential Tilting

We now delineate Exponential Tilting, the technique utilised in the three main papers on Importance Resampling to provide the importance distribution: Johns (1988), Do and Hall (1991) and Davison (1988).
Firstly, we resume our speech on Section 1.3. Applying the Bootstrap Assumption (1) on Equation (1.25), we can approximate a sample estimator $\hat{t}$ of statistic t with its linear approximation $\hat{t}_L$. Letting $\hat{F}$ denote the ECDF of a random sample $\mathbf{X}$, and $\hat{F}^*$ the ECDF of a sample with replacement from it $\mathbf{X}^*$, we can define the first order approximation of a sample statistic $\hat{t}$:

$$\hat{t}_L(\hat{F}^*) := \hat{t}(\hat{F}) + n^{-1} \sum_{i=1}^{N} l_j^* \tag{2.12}$$

[2] in alternative notation:

$$\hat{t}_L(\mathbf{X}^*) := \hat{t}(\mathbf{X}) + n^{-1} \sum_{i=1}^{N} l_j^* \tag{2.13}$$

where the l_j^* represent the empirical influence values (i.e. estimates of the influence values) of the first order derivative of statistic $\hat{t}$ (i.e an operator from the space of distributions to the space of $\mathbb{R}$) evaluated at $\hat{F}$, for the jth statistical unit of the Bootstrap sample $\mathbf{X}^*$ (hence the asterisk in the notation). This means that if the both the first and the second statistical units of the Bootstrap sample are the same statistical unit of the original sample, then $l_1^* = l_2^*$.
If the influence values are not known analytically, they can be obtained in several empirical ways (see Davison and Hinkley (1997) and Canty and Ripley (2022)):

- **Jacknife method**. Similar to the total order Sobol indices (see Manzoni (2022)), the influence value of the jth statistical unit is proportional to the change in the statistic when the jth statistical unit is not present in the bootstrap sample that calculates it. Related methods are the **infinitesimal** jacknife and the **positive**

[2]This notation is the one used in Davison and Hinkley (1997)

jacknife (see Canty and Ripley (2022) and Davison and Hinkley (1997) for more details).

- **Regression**. The idea is to firstly a pilot MC run with B_1 iterations for the Bootstrap estimate of $\hat{t}$. This yields:
  - $\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_{B_1} \end{bmatrix}$ where $y_b = \hat{t}(\mathbf{X}^{*b})$, $b = 1, \ldots, B_1$, where $\mathbf{X}^{*b}$ is the bth re-sample of original sample $\mathbf{X}$;
  - Design matrix $\mathbf{Z}$ of dimension $B_1 \times N$, (N sample size of $\mathbf{X}$)
    $\mathbf{Z} = \begin{bmatrix} f_1^{*1} & f_2^{*1} & \ldots & f_N^{*1} \\ \vdots & \vdots & \ddots & \vdots \\ f_1^{*B_1} & f_2^{*B_1} & \ldots & f_N^{*B_1} \end{bmatrix}$ where f_j^{*b}, $b = 1, \ldots, B_1$, $j = 1, \ldots, N$ is the frequency of the jth statistical unit on the bth Bootstrap sample. Note that $\sum_{j=1}^{N} f_j = N, \forall b \in \{1, \ldots, B_1\}$

  Whence we have the necessary elements to fit a linear regression, and the estimated (through Ordinary Least Squares) coefficients vector $\hat{\mathbf{b}}$ is nothing but the vector of the empirical influence values: $\hat{\mathbf{b}} = \begin{bmatrix} \hat{l}_1 \\ \vdots \\ \hat{l}_N \end{bmatrix}$

Now we are ready to outline the main reasoning made in Davison and Hinkley (1997) for Importance Resampling, where $\hat{T}^*$ is the Bootstrap estimator of sample version $\hat{T}$ of statistic T on sample $\mathbf{X}$

1. We approximate the statistic $\hat{T}^*$ by its linearised version $\hat{T}_L^*$ (as seen on Equation (2.13)) which is an accurate approximation of itself.
2. Such statistic $\hat{T}_L^*$ follows approximately a normal distribution, which is the case asymptotically. *Id est*, as seen in Equation (2.13), and applying the nonparametric delta method result of Equation (1.26), asymptotically $\hat{T}_L^* \sim N(\hat{T}(\hat{F}), v_L(\hat{F}))$
3. Exponential tilting is used to define the g_j in (1.18) **s.t.** the importance distribution yields values $\hat{t}(\mathbf{X}^{*\mathbf{b}})$ centered at value $\hat{\xi}_{I,\alpha}$, which is an initial rough estimate of the αth quantile of $\hat{T}^*$ we will estimate better through Importance Resampling.

That is, in the classical papers, we make the following assumptions:

Assumption 2 (Accuracy of the Linear Approximation)**.** *The linear approximation* $\hat{T}_L^*$ *of* $\hat{T}^*$ *is accurate.*

Assumption 3 (Normality of the Linearised Statistic)**.** *The linearised statistic* $\hat{T}_L^*$ *under ordinary resampling is approximately normal.*

Moreover, from the "classic" papers (Johns (1988), Davison (1988) & Do and Hall (1991)), two main approaches for Importance Resampling arise:

1. **Direct**. With both Assumptions (2) and (3), a single MC simulation is done directly, centering the Bootstrap distribution $\hat{T}^*$of statistic of interest T on the αth quantile of the normal approximation of $\hat{T}_L^*$ (see Davison and Hinkley (1997)), or on the value such that the estimated variance (computed analytically with the normal approximation) of the quantile estimate is minimised (see Do and Hall (1991) and Johns (1988)). However, this method requires the availability of the influence values and its validity is asymptotic (indeed, assumption 3 is valid under the C.L.T for several location estimators, as seen in Johns (1988)).

2. **Empirical**. The idea is to run a pilot MC run with Ordinary resampling, whence both the influence values and quantile of the Bootstrap estimate of the statistic of interest $\hat{\xi}_\alpha$ can be estimated. Then the importance distribution is derived such that: (1) the center of $\hat{T}_L^*$ is the pilot estimate denoted as $\hat{\hat{\xi}}_{B_1,\alpha}$ (B_1 number of iteration sin the pilot run) as in Davison and Hinkley (1997); (2) the variance of the estimate is minimised as in Hall (1992).

In this thesis we focus on the **Empirical** Importance Resampling, and use as baseline the implementation as in Davison and Hinkley (1997), *id est*, where the re-centering of the Bootstrap distribution of $\hat{T}_L^*$ is such that the quantile estimate of the pilot run $\hat{\hat{\xi}}_{B_1,\alpha}$ is at its center (of course, the center of the Bootstrap estimate of a statistic $\hat{T}$ under ordinary resampling is $\hat{t}(\mathbf{X})$ with $(\mathbf{X})$ being the available random sample).
Therefore, whereas failure to comply with Assumption (2) might lead to decreased efficiency with Importance Resampling, not fulfilling Assumption (3) might have a lesser effect on the variance, since the location of the re-centering does not depend on this assumption.

The above can be visualised in the following image:

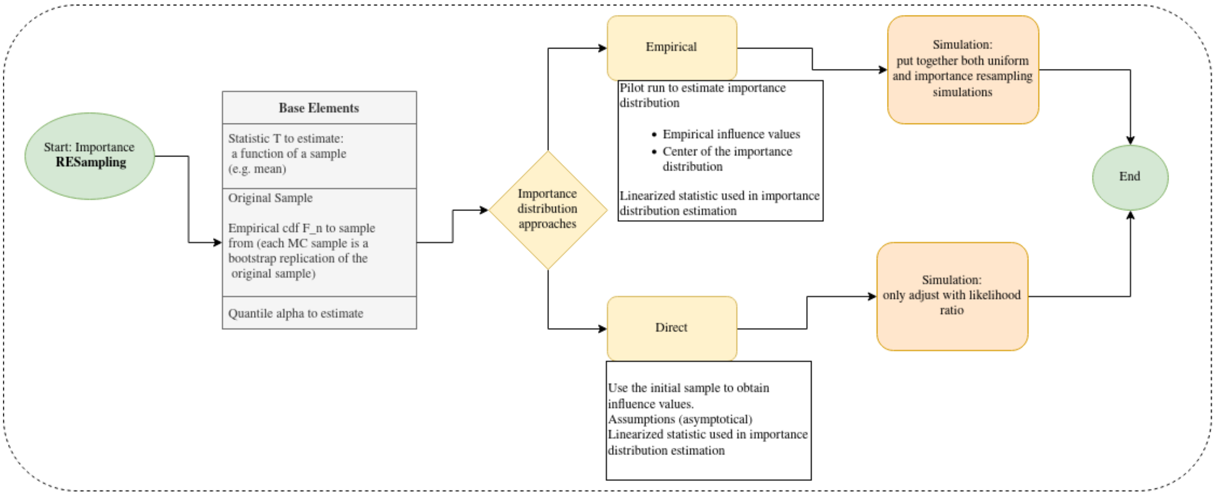

Figure 2.1: Importance resampling scheme for quantile estimation

We now move into the details behind the generation of the importance distribution in Importance Resampling.

We are now ready to go into the details of Exponential Tilting. To choose importance resampling distribution H, the idea is to assign, for Equation (1.18)

$$g_j \;\propto\; (\lambda l_j),\; j = 1,\, ...,\, N \tag{2.14}$$

where λ is a variable to tune such that the distribution is centered in the desired value. This is done through a Newton solver, see Davison and Hinkley (1997) and Canty and Ripley (2022), and in general converges quickly, as mentioned in Do and Hall (1991) Indeed, we want to choose $\lambda \in \mathbb{R}$ such that:

$$\frac{\sum_{i=1}^{N} l_i \exp(\frac{\lambda l_i}{N})}{\sum_{i=1}^{N} \exp(\frac{\lambda l_i}{N})} = \theta_0 \tag{2.15}$$

where θ_0 is the desired center for linearised Boostrap statistic $\hat{T}_L^*$ under Importance Resampling. To accomplish such goal, we to solve the following optimisation problem and trying to minimise the squared error:

$$\arg\min_{\lambda \in \mathbb{R}} \Big(\frac{\sum_{i=1}^{N} l_i \exp(\frac{\lambda l_i}{N})}{\sum_{i=1}^{N} \exp(\frac{\lambda l_i}{N})} - \theta_0 \Big)^2 \tag{2.16}$$

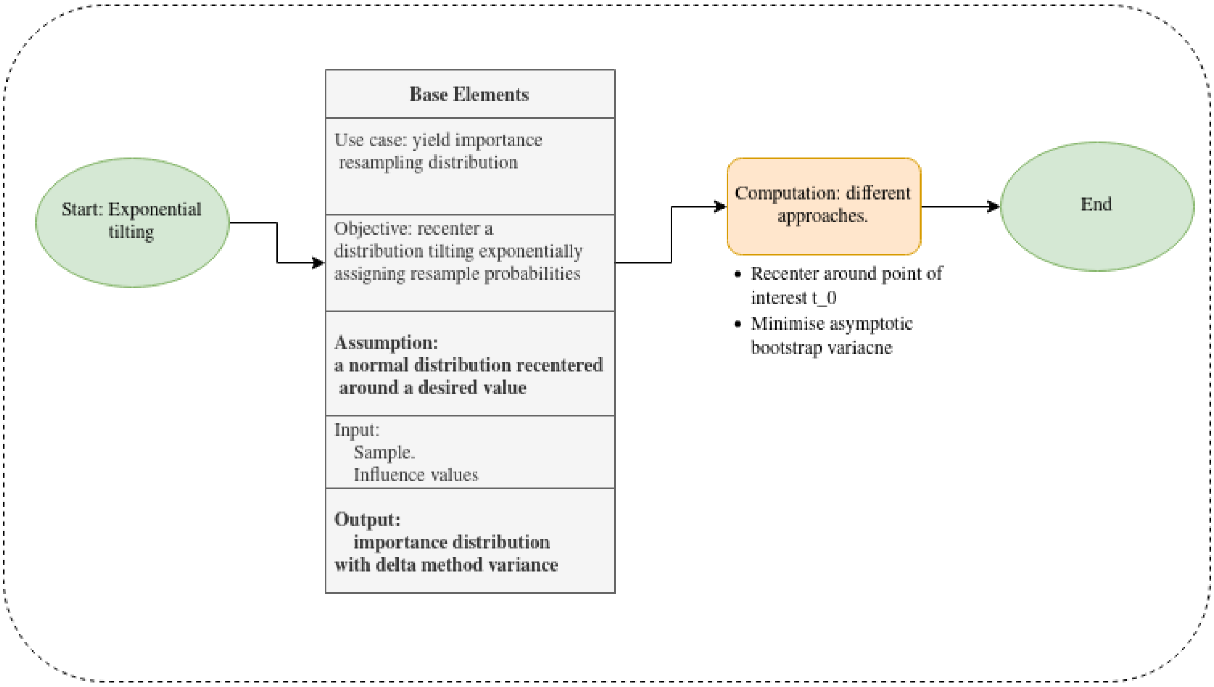

Figure 2.2: Importance resampling scheme for quantile estimation through Exponential Tilting

Exponential tilting is summarised in figure (2.2).

Remark 5 (Why exponential tilting?). *There are two main reasons:*

- *It allows to set the Importance distribution (g_j in Equation (1.18)) such that the linearised statistic $\hat{T}_L^*$ is re-centered to a desired value (although it is not necessarily the only method to do so)*
- *It keeps the variance of the linearised statistic $\hat{T}_L^*$ the same as under Ordinary Resampling ($p_j = \frac{1}{N}$), which is v_L (id est the nonparametric delta method variance).*

However, the variance of the importance distribution need not be the same as the linearised statistic under equally-weighted ($p_j = N^{-1}$) resampling. Indeed, in the general case of Importance sampling, an ideal (but infeasible) importance distribution would yield a dirac mass on the value we want to estimate, requiring one single MC iteration for an estimate without variance. Nonetheless, in this setting we are working with a discrete and finite distribution $\hat{F}$ so such choice may not even be possible. Moreover, as Johns (1988) mentions: "*The method of the exponential tilting leaves the variance of the statistic unchanged, which leaves open the possibility that some other transformation might reduce the variance and perhaps lead to an improved quantile estimator*". Taking into consideration the non-linearity of statistics (1.41) and (1.36), it is likely that in our case Assumption 2 is not fulfilled, and the motivations in Remark 8 may lose validity, which opens the possibility for other procedures.

We are now ready to write the algorithm for importance resampling as presented in Davison and Hinkley (1997). Given a sample $\mathbf{X}$ of an underlying possibly multi or infinite-dimensional distribution F of cardinality N, we want to estimate the α quantile of operator t. We through its sample estimate $\hat{T} = \hat{t}(\mathbf{X})$ whose distribution we approximate through the Bootstrap, so in practice we estimate via the Monte Carlo method the distribution of $\hat{T}^*$

Algorithm 2.1 Importance resampling (Davison)

```
Set B_1 and B_2 for the pilot run and the Importance run, respectively.
for b ∈ {1, ..., B_1} do
    Obtain Bootstrap sample X^{*b} by sampling with replacement from original sample X
    Set T_b^* ← t̂(X^{*b})
end for
Obtain the empirical influence values l̂_j^*, j ∈ {1, ..., N} through regression.
Obtain an estimate of the α quantile ξ̂̂_{B_1,α} using T_b^*, b = 1, ..., B_1
Calculate the probabilities of resampling each statistical unit g_j by solving Problem (2.16), yielding importance distribution H of shape of Equation 1.15 with probabilities g_i ∀i ∈ {1, ..., N}.
for b ∈ {1, ..., B_2} do
    Obtain Bootstrap sample X^{*b} by sampling with replacement from original sample X with probabilities g_i ∀i ∈ {1, ..., N}
    Set T_b^* as t̂(X^{*b})
    Compute the likelihood ratio as w(X^{*b}) ← dF(X^{*b}) / dH(X^{*b})
end for
if α < 0.5 then
    Estimate the αth quantile using Equation (2.10)
else
    Estimate the αth quantile using Equation (2.11)
end if
```

Remark 6 (Simplification of the quantile estimation)**.** *Contrary to the algorithm proposed by Hall (Do and Hall (1991)), which uses both the pilot and the Importance simulation to estimate the Bootstrap quantile of the statistic in a convex combination* [3]*, in this thesis*

[3] The weights are attributed according to an estimate of the variance, see Do and Hall (1991)

we will compare Importance Resampling and Ordinary Bootstrap considering only the iterations B_2 in which the samples were made from the importance distribution for the first (so as if the center and influence values were available), and the same number of iterations B_2 for the Bootstrap with equal probabilities for each statistical unit.

2.2.2. Our first proposal: Loss Tilting

We recall from the Introduction that the scope of this thesis is to apply Importance Resampling for the quantile estimation of statistics that are used to construct Bootstrap SCBs, such as (1.33) and (1.39). In such particular case, Assumption (2) is not valid anymore, since the statistic is the composition of the *supremum* ($infimum$), a non-linear function, and another statistical function (for *e.g.* Student's t statistic, which is non-linear as well). What is more, whereas Student's t is a smooth function (see for *e.g.* Johns (1988)) the *supremum* ($infimum$) is not (of course when $p > 1$), which puts Assumption (3) in jeopardy as well.
We thus make the following remarks:

Remark 7 (Violation of Assumption (2) with statistics used for SCBs)**.** *When using statistics (1.39) and (1.34), the supremum ($infimum$) induces a non-linearity that increases as p increases, so that the linear approximation $\hat{T}_L$ of a sample estimator $\hat{T}$ is not accurate, violating Assumption (2).*
Such violation occurs in Johns (1988), for example, who uses Student's t statistic, but Exponential Tilting still reduces the MC variance. In our scenario of interest, there is a further layer of non-linearity induced by the sup (inf) operator, which worsens such infringement.

Remark 8. *(Non-smoothness of statistics used for SCBs) The delta method result(1.26) is not fulfilled in the statistics used in the construction for SCBs, since statistics (1.39) and (1.34) use the supremum ($infimum$) a non-smooth function. In this thesis we employ the use of influence values to obtain the linearised statistic anyways.*

Consequently, one may anticipate that since the core hypotheses under Exponential Tilting in the classic papers, the reasons for which it is used will no longer hold, see Remark (5). What is more, we will see in Section 2.3 it will lead to a larger variance w.r.t Ordinary Resampling when $p > 1$.
This opens the possibility for other procedures to obtain an Importance distribution to reduce the variance (1.13). We leave the discussion of the optimality of Exponential Tilting in Importance Resampling for future research, and focus on building alternative algorithms that can reduce the MC variance when Assumptions (2) and (3) are not nec-

essarily true.
We thus firstly propose **Loss Tilting**. We recall the objective is to re-center the Bootstrap distribution $\hat{T}^* = \hat{t}(\mathbf{X}^*)$ of a sample estimator $\hat{T} = \hat{t}(\mathbf{X})$ of statistic of interest $T = t(F)$ from center $\hat{t}(\mathbf{X})$ to a desired value θ_0 as a new center, which may be for example the estimate obtained after the pilot run $\hat{\hat{\xi}}_{B_1,\alpha}$ (see Algorithm 2.1).
The intuition is the following: we want to re-sample more frequently the statistical unit j, the more its influence value l_j (or its empirical estimate $\hat{l}_j$) pushes towards θ_0 (which of course happens in the case of Exponential Tilting).
Then, we denote the difference between the desired center for the tilted Bootstrap distribution and the center under ordinary resampling:

$$d := \theta_0 - \hat{t}(\mathbf{X}) \tag{2.17}$$

and assign a re-sampling probability g_j for the jth statistical unit such that the closer (in terms of a possibly symmetric loss function ℓ) its influence value l_j is to d, the higher g_j is.
Consequently, we propose the following procedure:

1. Compute the difference d as in (2.17)
2. For **each** statistical unit j, compute the Loss function of the difference between d and the (empirical if not derived analytically) influence value l_j ($\hat{l}_j$ when estimated empirically)
$$h_j = \ell(d - l_j) \tag{2.18}$$
3. Since we want a probability distribution, we normalise h_i:
$$\tilde{h}_j = \frac{h_j}{\sum_{i=1}^{N} h_i} \tag{2.19}$$
4. Since we want to give less probability the higher the loss is, we compute the complement of each $\tilde{h}_i$:
$$\tilde{h}_j^c = 1 - \tilde{h}_j \tag{2.20}$$
5. And normalise them to get the importance distribution:
$$g_j = \frac{\tilde{h}_j^c}{\sum_{i=1}^{N} \tilde{h}_i^c},\ j \in \{1, \dots, N\} \tag{2.21}$$

which is summarised in the following algorithm:

Algorithm 2.2 Loss tilting

1: Given sample $\mathbf{X}$ of size N, sample estimator $\hat{t}$, desired center θ_0, loss function ℓ
2: Compute $d = \leftarrow \theta_0 - \hat{t}(\mathbf{X})$,
3: Set $h_j = \ell(d - l_j)$, $j \in \{1, \ldots, N\}$
4: Set $\tilde{h}_j = \frac{h_j}{\sum_{i=1}^{N} h_i}$, $j \in \{1, \ldots, N\}$
5: Set $\tilde{h}_j^c = 1 - \tilde{h}_j$, $j \in \{1, \ldots, N\}$
6: Set $g_j = \frac{\tilde{h}_j^c}{\sum_{i=1}^{N} \tilde{h}_i^c}$, $j \in \{1, \ldots, N\}$

Remark 9 (Choice of the loss function)**.** *There are possibly infinite valid choices for a function that gives higher resampling probabilities for statistical units whose influence values push towards the desired center for the tilted distribution.*
What is more, the steps in subsection 2.2.2 are not the only way to obtain the probabilities. For instance, normalisation could be done with a softmax transform, or simply by setting $g_j \propto \frac{1}{\ell(d-l_j)}$
The point is that when we choose an alternative to Exponential tilting, we renounce the guarantee that the tilted Bootstrap distribution of the linearised statistic $\hat{T}_L^*$ *will be effectively centered at desired* θ_0 *, yet a whole new world is unlocked.*
Indeed, one could choose a function with a parameter to control the variance in the tilted distribution, which could for example be tuned according to the uncertainty of estimate $\hat{\hat{\xi}}_{B_1,\alpha}$ *(see Algorithm (2.1)).*

Remark 10 (Didactic-ness of Loss Tilting)**.** *Our first proposal provides no guarantee whatsoever regarding the center of the tilted distribution, except that the Bootstrap distribution of the linearised statistic* $\hat{T}_L^*$ *will be closer to* θ_0 *w.r.t the Ordinary (no tilting) Bootsrap distribution.*
Therefore, we state Loss Tilting is didactic in the sense it is a very general algorithm, with a loss function ℓ *to be chosen according to the specific problem.*
Our purpose with such algorithm is to show that when the assumptions of Exponential Tilting fail, other options that can effectively reduce the MC variance are available (see section 2.3).

We now formally write the algorithm for **Loss Tilting** for Importance Resampling.

Algorithm 2.3 Importance Resampling with Loss Tilting

```
Set $B_1$ and $B_2$ for the pilot run and the Importance run, respectively.
for $b \in \{1, \dots, B_1\}$ do
    Obtain Bootstrap sample $\mathbf{X}^{*\mathbf{b}}$ by sampling with replacement from original sample $\mathbf{X}$
    Set $T_b^* \leftarrow \hat{t}(X^{*b})$
end for
Obtain the empirical influence values $l_j^*,\ j \in \{1, \dots, N\}$ through OLS.
Obtain an estimate of the $\alpha$ quantile $\hat{\hat{\xi}}_{B_1,\alpha}$ using $T_b^*,\ b = 1, \dots, B_1$
Calculate the probabilities of resampling each statistical unit $p_i$ by following Algorithm 2.2, yielding importance distribution $H$ of shape of Equation 1.15 with probabilities $g_i\ \forall i \in \{1, \dots, N\}$.
for $b \in \{1, \dots, B_2\}$ do
    Obtain Bootstrap sample $\mathbf{X}^{*\mathbf{b}}$ by sampling with replacement from original sample $\mathbf{X}$ with probabilities $g_i\ \forall i \in \{1, \dots, N\}$
    Set $T_b^* \leftarrow \hat{t}(X^{*b})$
    Compute the likelihood ratio as $w(\mathbf{X}^{*\mathbf{b}}) \leftarrow \frac{dF(\mathbf{X}^{*\mathbf{b}})}{dH(\mathbf{X}^{*\mathbf{b}})}$
end for
if $\alpha < 0.5$ then
    Estimate the $\alpha$ quantile using Equation (2.10)
else
    Estimate the $\alpha$ quantile using Equation (2.11)
end if
```

2.2.3. Our second proposal: Contribution Tilted Mixture (CTM)

This Importance Resampling technique we propose at the present work is specifically designed for the case in which the task of interest is the construction of Bootstrap SCBs. As a matter of fact, it is only applicable with a statistic which is the *sup* or *inf* of an element-wise statistic, such as (1.39) and (1.34). As shown in Remark (8), the additional (apart from the one inherent to Student's t) nonlinearity induced by the *sup* (*inf*) is a problem, and what is more it renders the statistics non-smooth, so if we try to obtain their derivatives to obtain influence values as in 1.23 may be cumbersome, apart from the departure from Assumption (3)

Our second proposal, **Contribution Tilted Mixture**, has been escogitated for this special case in which we have a non-smooth operator such as *sup* or inf, and working with high-dimensional data. In particular, with $\mathcal{S}$ being a discrete set of the form $\mathcal{S} = \{1, ..., p\}$ with p being a possibly large integer (see Section 1.4). Note this is the case even for the functional setting (*i.e.* $p = \infty$), where a discretisation of the curve is utilised, see for *e.g.* Pini and Vantini (2017).

The intuition is the following: since both statistics (1.34) and (1.39) (the same goes for their inf versions) are the *supremum* ($infimum$) of an estimated element-wise statistic, for *e.g.* (1.39): $\sup_{s\in\mathcal{S}} \hat{\mu}^*(s) - \hat{\mu}(s)$, then their value necessarily corresponds to the value of one of its components $\tilde{s}$, *i.e.* $\sup_{s\in\mathcal{S}} \hat{\mu}^*(s) - \hat{\mu}(s) = \hat{\mu}^*(\tilde{s}) - \hat{\mu}(\tilde{s})$.
Therefore, if such statistics take only the value of their component say $\tilde{s}$, we can "forget" about the fact it is a *sup* (inf) and apply Exponential (or any other) tilting to the quantity $\hat{\mu}^*(\tilde{s}) - \hat{\mu}(\tilde{s})$ in the case of statistic (1.41) or $\frac{\sqrt{N}(\hat{\mu}^*(\tilde{s})-\hat{\mu}(\tilde{s}))}{\hat{\sigma}^*(\tilde{s})}$ in the case of (1.36).
In particular, if we choose Exponential Tilting, then the deviation from Assumption (2) would not be violated as badly as for statistic (1.41) (see Remark 8) and would only contain the Student's t nonlinearity in (1.36), which does not hinder the variance reduction of Importance Resampling (see Johns (1988), Do and Hall (1991), Davison (1988)).
Nonetheless, usually it will not be the case that the *sup* (inf) takes the value of a single component. Different components $s \in \mathcal{S}$ may be the ones whose value is the one taken by the statistic with the *sup* (inf) .
Thus, we consider as a "contributor" each element $s \in \mathcal{S}$ of the multivariate (functional) statistic. We make the following reasoning: the more frequent the value of element $s \in \mathcal{S}$ is the one taken by the statistic with the *sup* (inf), the bigger the weight we give to the (exponential) tilting done on the (univariate) quantity of the statistic at component s.
Exploiting the fact that in Algorithm 2.1 runs a pilot run, it would be possible to count, for each element $s \in \mathcal{S}$, how many times it was such element whose value became the value of the *sup* (inf) statistics (1.41) and (1.36).
Therefore, the idea is to use as an Importance distribution a weighted Mixture of the individual weights $p_j^{(s)}$, $j \in \{1, \dots, N\}$, $s \in \mathcal{S}$ obtain through the point-wise (Exponential) Tiltings to re-center at the desired quantile α.
Thus, we denoting with $\hat{m}^*(s)$, $s \in \mathcal{S}$ either statistic $\sqrt{N}\frac{\hat{\mu}^*(s)-\hat{\mu}(s)}{\hat{\sigma}^*(s)}$ or statistic $\hat{\mu}^*(s)-\hat{\mu}(s)$; $T^* = \hat{t}(\mathbf{X}^*) = \sup_{s\in\mathcal{S}} \hat{m}^*(s)$, we define the estimate of the contribution of the element $s \in \mathcal{S}$ with:

$$\hat{c}_s := \frac{\sum_{b=1}^{B_1} \mathbb{1}\{\hat{T}_b^* = \hat{m}_b^*(s)\}}{B_1}, , \; s \in \mathcal{S} \tag{2.22}$$

where B_1 is the number of Monte Carlo iterations in the pilot run, $\hat{T}_b^*$ and $\hat{m}_b^*(s)$ the values of Bootstrap $\hat{T}^*$ and $\hat{m}^*(s)$ statistics at the bth iteration.

Set as resampling probability for the jth statistical unit:

$$\tilde{p}_j := \sum_{s \in \mathcal{S}} p_i^{(s)} \hat{c}_s \tag{2.23}$$

where $p_i^{(s)}$ is the probability of resampling the ith statistical unit after applying (Exponential) Tilting to $\hat{m}^*(s)$ at a fixed $s \in \mathcal{S}$ so that it is re-centered at the same order of the quantile of interest for $\hat{T}^*$. Note it yields an Importance distribution (*i.e.* $\sum_{i=1}^{N} \tilde{p}_j = 1$) since it is a convex combination ($\sum_{s \in \mathcal{S}} \hat{c}_s = 1;\ 0 \leq c_s \leq 1,\ \forall s \in \mathcal{S}$) of the element-wise importance distributions.
Therefore, **Contribution Tilting Mixture (CTM)** can be summarised in the following:

Algorithm 2.4 Contribution Tilting Mixture

1: Given the results of the Pilot run in Algorithm 2.1, that is:
- element-wise estimate of quantile of interest $\hat{\hat{\xi}}_{B_1,\alpha}(s)$ of $\hat{m}^*(s)\,, s \in \mathcal{S}$;
- estimate contribution of each element $\hat{c}_s\,, s \in \mathcal{S}$ as in (2.22)

2: Compute through (Exponential) Tilting the elemente-wise Importance Distribution, yielding $p_j^{(s)}$, $j \in \{1, \ldots, N\}$, $s \in \mathcal{S}$

3: Set $\tilde{p}_j := \sum_{s \in \mathcal{S}} p_i^{(s)} \hat{c}_s \quad i \in \{1, \ldots, N\}$ as the Importance Resampling probabilities.

---

to be inserted in Algorithm 2.1.

2.3. Simulation study

We devote this section to test the algorithms. Different experiments focusing on statistics (1.41) and (1.36) and the MC (under both ordinary and Importance Resampling) estimate $\hat{\hat{\xi}}_{MC,\alpha}$ of the Bootstrap estimate of their αth order quantile $\hat{\xi}_\alpha$. We use **Exponential Tilting**, **Huber tilting** and **Contribution Tilting Mixture (CTM)** to obtain the importance distribution and compare their results. We denote, as before, $\hat{T}^*$ the Bootstrap distribution of the sample estimator $\hat{T}$ of statistic T of interest (whether (1.39) or (1.34)), and with $\hat{T}^*_L$ its linear approximation. All samples were taken from an underlying Gaussian Process defined on domain $\mathcal{I} = [0, 1]$, with mean $\mu(t) = sin(\pi t) + sin(2\pi t)$, $t \in \mathcal{I}$, with exponential covariance of parameters $\alpha = 1$ and $\beta = 2$ (see Ieva et al. (2019)):

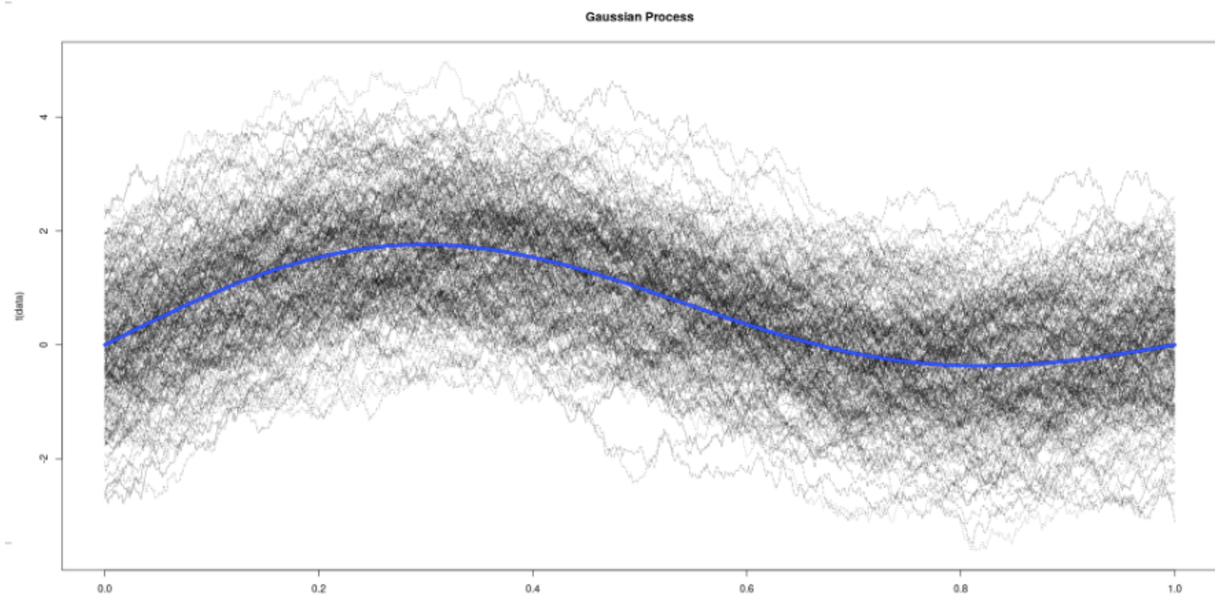

Figure 2.3: Gaussian process of which samples of different N and p (number of dimensions) are drawn.

with and both N, the sample size and p the number of random variables (elements) per statistical unit, were varied.
With the covariance kernel being exponential, the closer two points are in the domain, the higher their covariance is. Whenever $p > 1$, we decided to keep a mild covariance between the elements that conformed the statistical units in random samples $\mathbf{X}$. To accomplish such situation, for each chosen p, an equally spaced grid of p points over domain $\mathcal{I}$ was utilised. Hence, with $p = 10$, each statistical unit $\mathbf{x}_i$ (for some integer i between 1 and N of random sample $\mathbf{X}$ would have 10 elements: *i.e.* $\mathcal{S} = \{1, ..., p\}$ (following notation in Section 1.4) , where the ith element of a statistical unit corresponds to the following point in the domain $\mathcal{I}$: $t_s = \frac{s}{p+1}$, $\forall s \in \mathcal{S} = \{1, ..., p\}$. The different chosen values of N are 16, 32, 64, whereas for p they are 1, 10, 40, 100, where the last value is a is used in the Functional setting of Bootstrap SCBs (see for *e.g.* Degras (2016)).
For ease of reference, we denote the following:

- *sup-bias* as the *sup*-like statistic (1.39) with Bootstrap estimator (1.41)
- *sup-t-student* for (1.34) with Bootstrasp estimator (1.36)

We note that since samples are taken from a symmetric stochastic process, we only study the *sup*-like statistics, since the same behavior is expected from their inf-like counterparts. The only difference, as noted in Remark 4 is that estimator (2.11) is used instead of (2.10), since the quantile of interest will be $1 - \alpha$, with the chosen α for the simulations being

$\alpha = 0.5$
What is more, for all the simulations we utilised $B_1 = 100$ MC iterations for the pilot runs, $B_2 = 1000$ for Importance Resampling runs.
Regarding the choices for the Importance distribution in Importance Resampling:

- We used Davison's (Alg. 2.1) method for Exponential Tilting
- We utilised $\ell(x) = \frac{1}{2}x^2$ as loss function for Loss Tilting (see (2.18))
- For the element-wise importance distribution in CTM, we used Exponential Tilting (see Algorithm (2.4)).

2.3.1. Comparison of the statistics and their linearised approximations

In this subsection, our aim was to see the potential violation of Assumption (2) (the approximation $\hat{T}_L^*$ for $\hat{T}^*$ is accurate for both statistics, by varying both N and p.
Thus, for each statistic, we simulated $B = 1000$ MC iterations (or resamples) to analyse their distribution. We utilised Ordinary Resampling (*i.e.* with $p_j = N^{-1}$, see (1.15)), since the purpose was only to see how good the linear approximation of the Bootstrap statistics was.
Of course, when $p = 1$, *sup-bias* is a linear statistic, and this is perfectly shown in Figure (2.4). This was not the case for $sup - t - student$, which is nonlinear even when $p = 1$ (see fig. 2.5). As p increases, the more nonlinear the statistic becomes due to the *sup* operator, which results in the linear approximation missing higher-order effects and thus shows the left-shift in the linearised distribution w.r.t the actual one. In other words, an increase in the dimensionality means we deviate farther from Assumption 2.
Moreover, the contribution to non-linearity from the *sup* operator is much accentuated than the one of Student's t statistic. Indeed, by looking at Fig. (2.4), we notice that when $p = 1$ and only Student's t statistic's non-linearity is present, the linear approximation is accurate, which is coherent with what we have mentioned, *id est*, that Exponential Tilting leads to significant efficiency gains with such statistic despite of its non-linearity (see Johns (1988)). The *sup* operator, on the contrary, makes the linear approximation inaccurate while increasing p, which is noticeable from Figures (2.4) and (2.5).

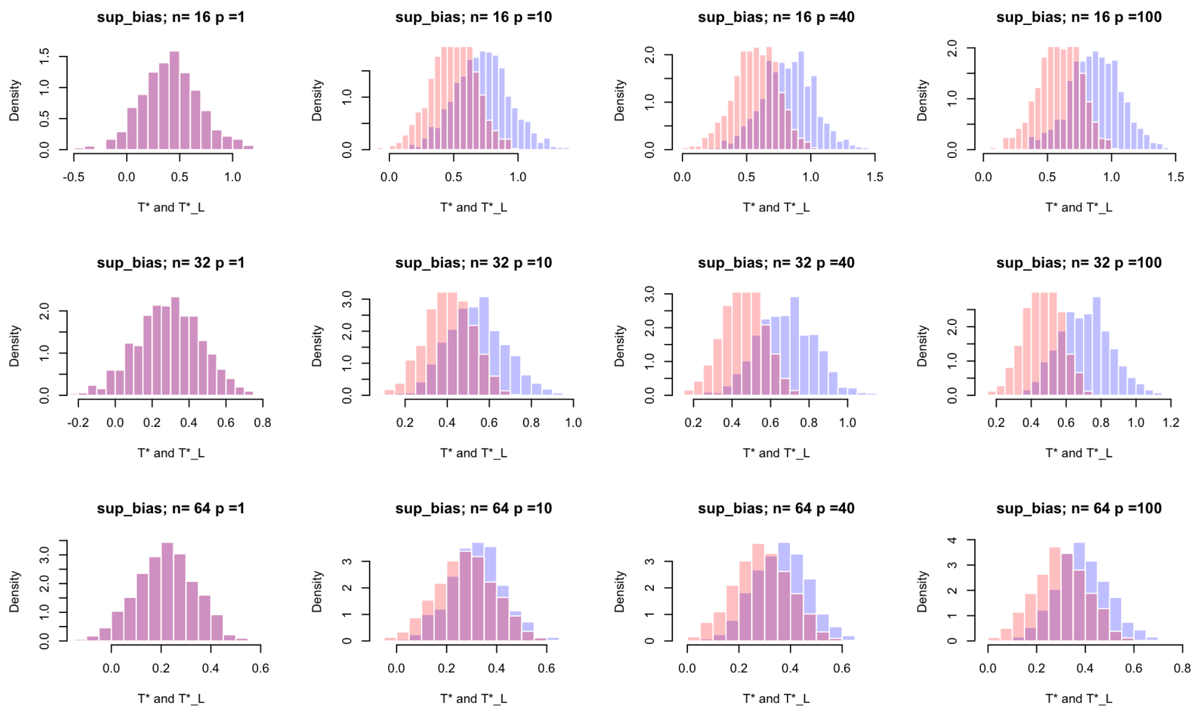

Figure 2.4: A thousand samples of the Boostrap distribution T^* (blue) and its linearised version T^*_L (pink) when using **sup-bias**

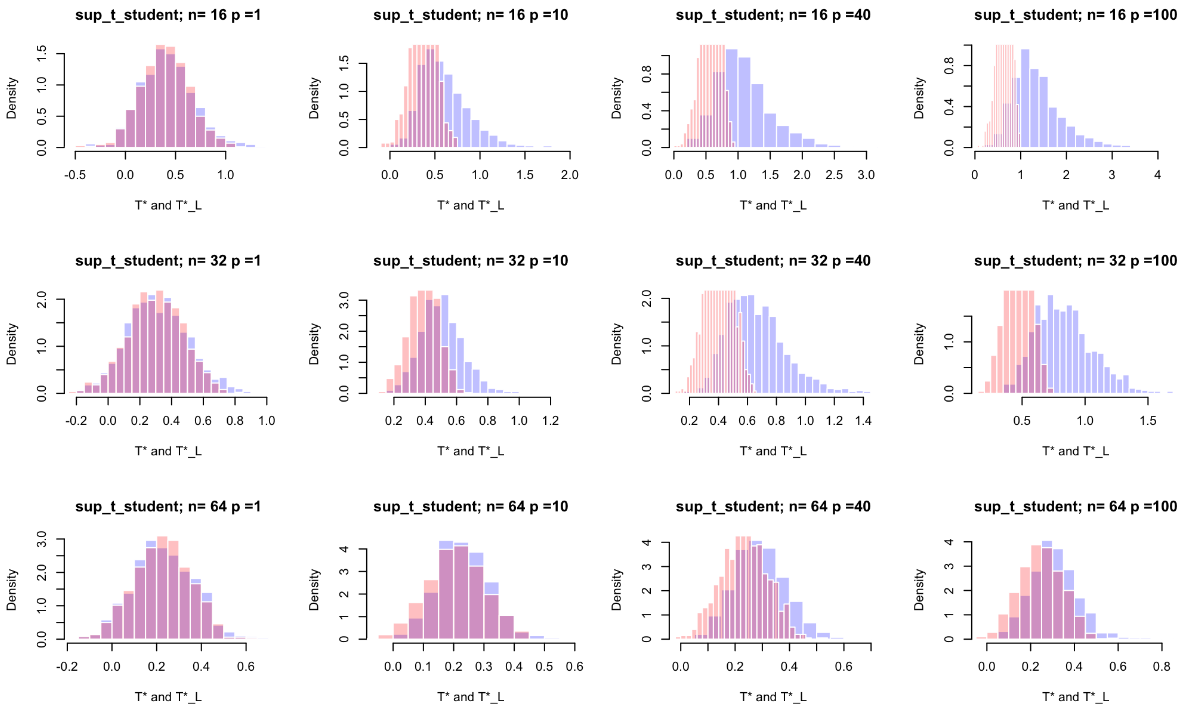

Figure 2.5: A thousand samples of the Boostrap distribution T^* (blue) and its linearised version T^*_L (pink) when using **sup-t-student**

2.3.2. Comparison of statistics under Ordinary Resampling and under Importance Resampling

In this subsection, our aim is to show simulations of the Bootstrap statistics T^* and their distributions under Importance Resampling, wherein the Importance distributions were outputted by Exponential Tilting and by our proposals, namely Loss Tilting and Contribution Tilted Mixture.

What is more, we indagated the true (given by an MC simulation of Ordinary Resampling with 50000 replications such that its variance was negligible) quantile $1-\alpha$ of the Bootstrap estimator of each statistic, i.e. $\hat{\xi}_{1-\alpha}$; its estimate after the pilot run $\hat{\hat{\xi}}_{B_1,1-\alpha}$, and its estimate under Importance Resampling ($\hat{\hat{\xi}}_{IR,1-\alpha}$).

Whereas in Section 2.3.1 it seemed that the *sup* of the element-wise Student's t had more aggressive departures with respect to the *sup* of the mean bias as p increased, we see very similar plots for both statistics under Exponential Tilting. As a matter of fact, except for the case of $p = 1$, the tilted distributions are not centered at all on the desired quantile, which spoils the fact it would not increase (if not decrease!) efficiency.

Loss tilting with the abovementioned loss function choice seems very conservative: whereas it does tilt the distribution of both statistics to the right, it is not enough to be centered at the desired quantile.

CTM seems the most promising: for both statistics, across all values of N and p it is both approximately centered around the desired point and with a significant overlap with the original distribution, which will prevent the appearance of likelihood ratio weigths (equation 1.17) that go high.

According to the results seen on this section, CTM is definitely the best method of the three.

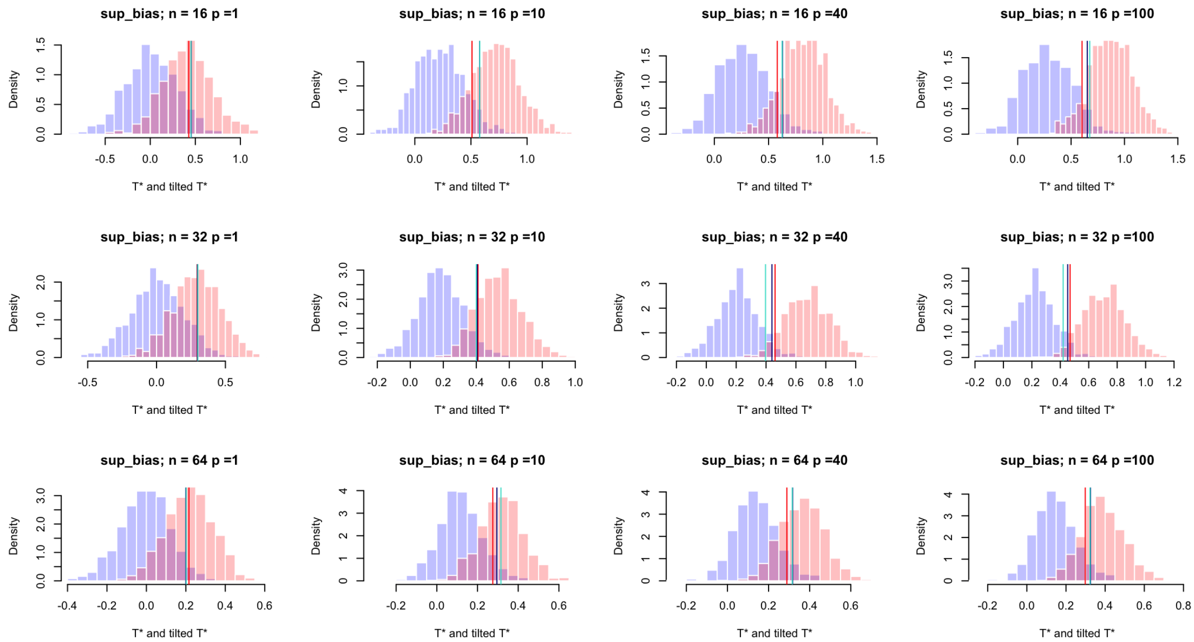

Figure 2.6: T^* (blue) and its tilted version (pink) with *sup-bias* and Exponential Tilting. The vertical red, blue and turquoise lines are $\hat{\xi}_{1-\alpha}$, $\hat{\hat{\xi}}_{B_1,1-\alpha}$ and $\hat{\hat{\xi}}_{IR,1-\alpha}$, respectively.

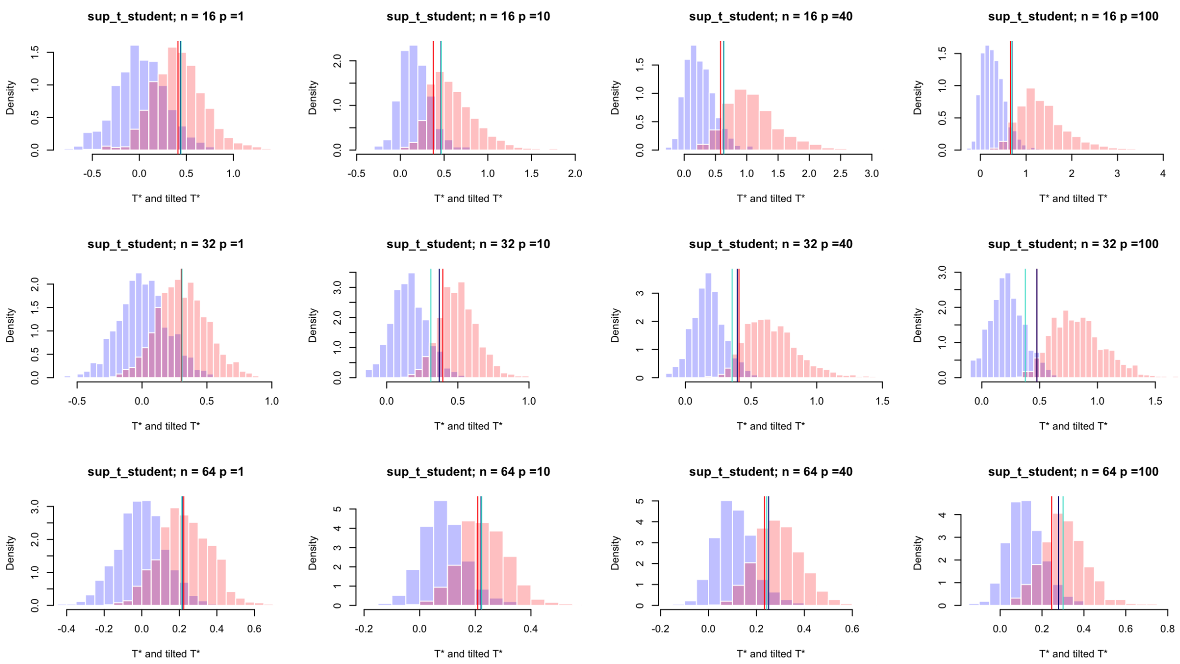

Figure 2.7: T^* (blue) and its tilted version (pink) with *sup-t-student* and Exponential Tilting.The vertical red, blue and turquoise lines are $\hat{\xi}_{1-\alpha}$, $\hat{\hat{\xi}}_{B_1,1-\alpha}$ and $\hat{\hat{\xi}}_{IR,1-\alpha}$, respectively.

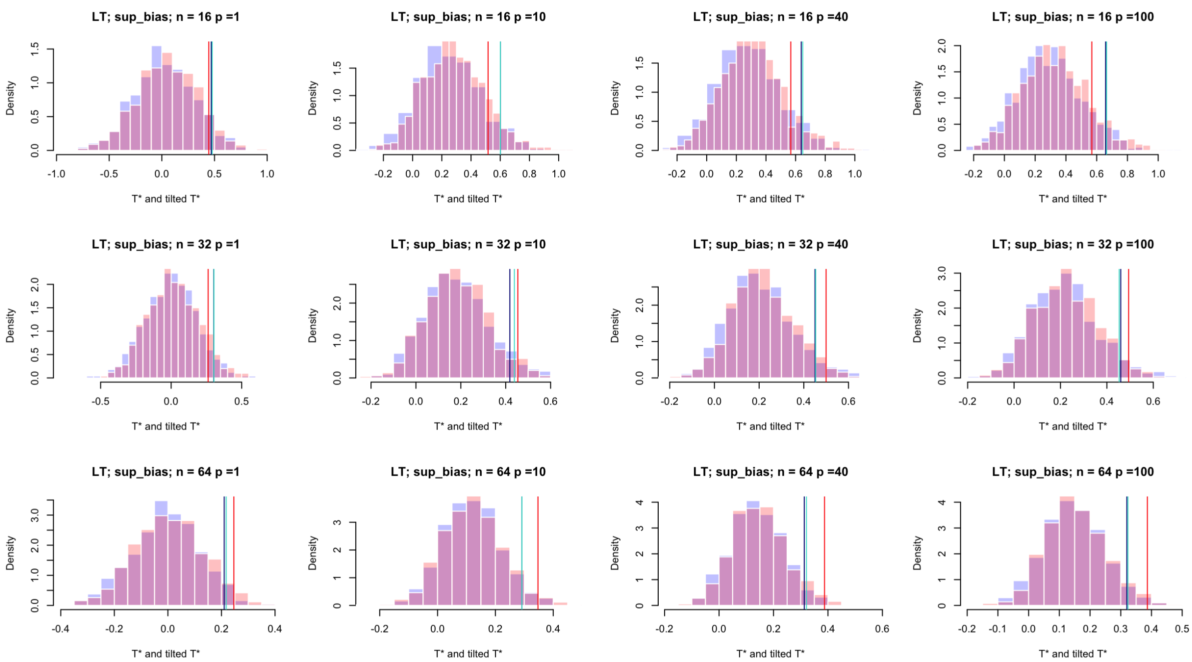

Figure 2.8: T^* (blue) and its tilted version (pink) with *sup-bias* and Loss Tilting.The vertical red, blue and turquoise lines are $\hat{\xi}_{1-\alpha}$, $\hat{\hat{\xi}}_{B_1,1-\alpha}$ and $\hat{\hat{\xi}}_{IR,1-\alpha}$, respectively.

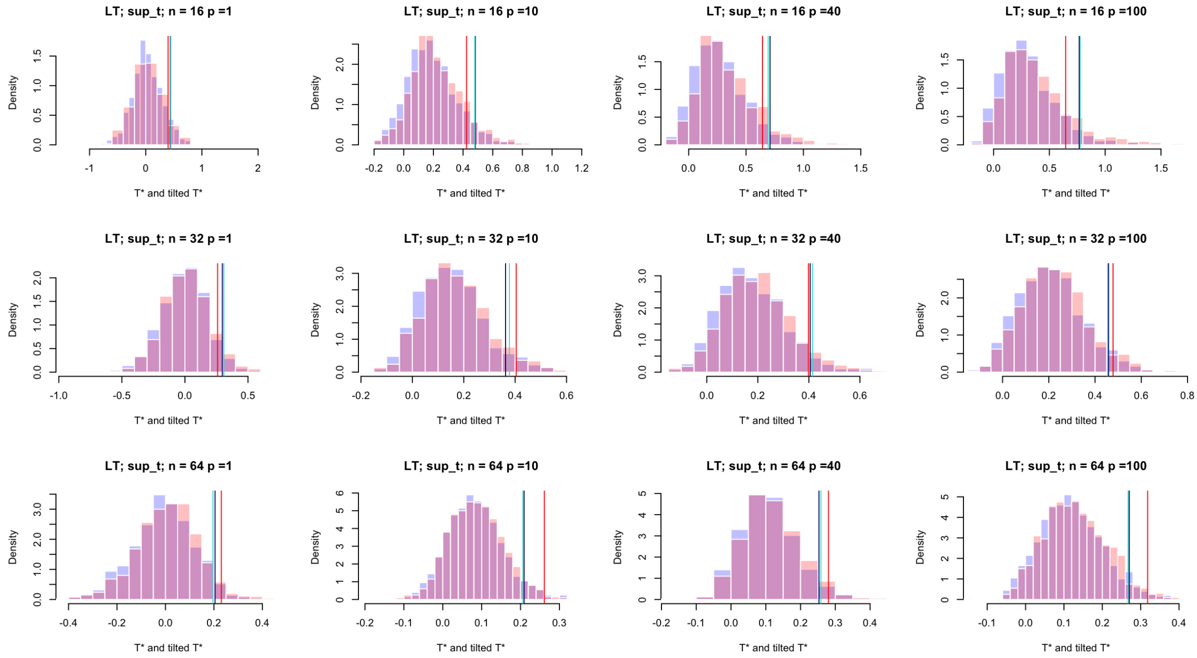

Figure 2.9: T^* (blue) and its tilted version (pink) with *sup-t-student* and Loss Tilting.The vertical red, blue and turquoise lines are $\hat{\xi}_{1-\alpha}$, $\hat{\hat{\xi}}_{B_1,1-\alpha}$ and $\hat{\hat{\xi}}_{IR,1-\alpha}$, respectively.

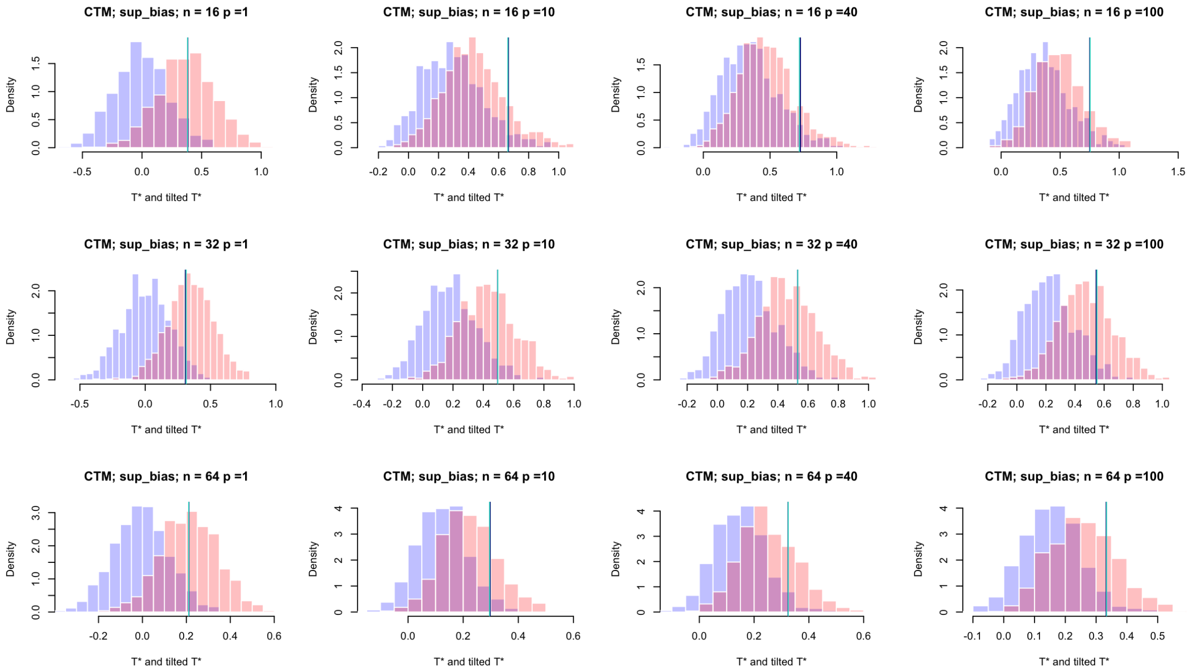

Figure 2.10: T^* (blue) and its tilted version (pink) with *sup-bias* under CTM. The vertical red, blue and turquoise lines are $\hat{\xi}_{1-\alpha}$, $\hat{\hat{\xi}}_{B_1,1-\alpha}$ and $\hat{\hat{\xi}}_{IR,1-\alpha}$, respectively.

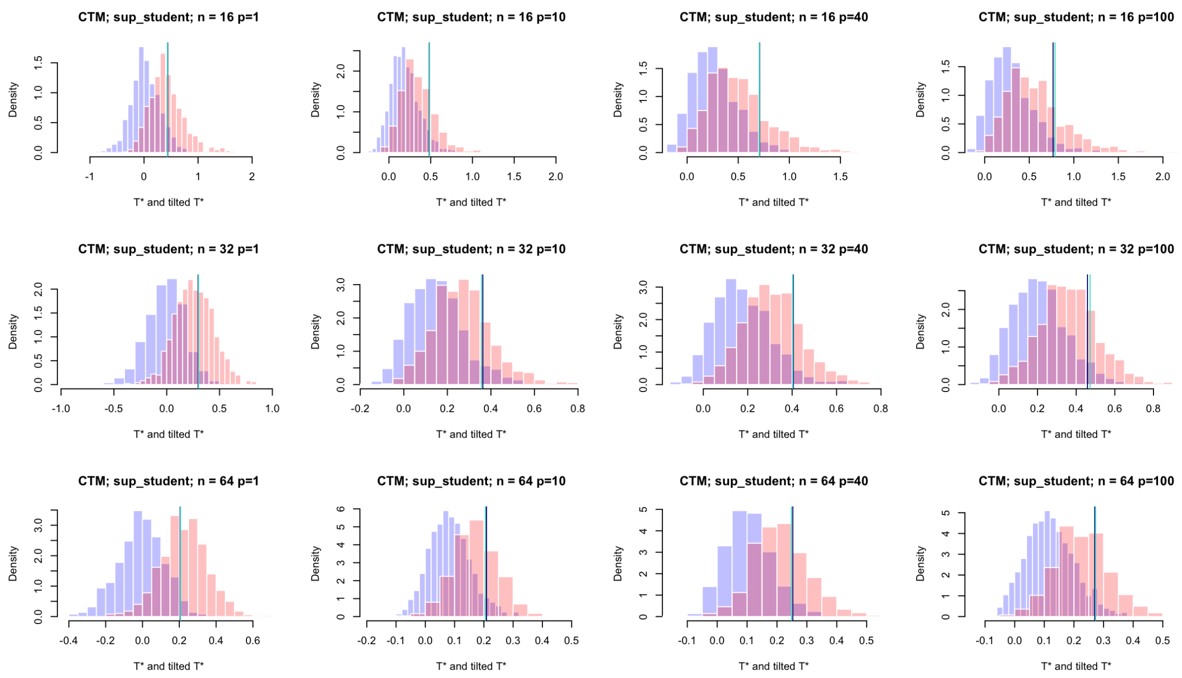

Figure 2.11: T^* (blue) and its tilted version (pink) with *sup-t-student* under CTM. The vertical red, blue and turquoise lines are $\hat{\xi}_{1-\alpha}$, $\hat{\hat{\xi}}_{B_1,1-\alpha}$ and $\hat{\hat{\xi}}_{IR,1-\alpha}$, respectively.

2.3.3. Plots of the Importance Resampling Probabilities

Another aspect of interest to our simulation was naturally the Importance distribution of shape (1.18) under the different algorithms that yield it. For both statistics and all three algorithms, we plotted the re-sampling probabilities g_j *versus* the empirical (hence estimated) influence values of each statistical unit. Again, we conducted the experiments for the different N and p mentioned above.
The idea was to confirm the shape these resampling probabilities would have: *id est* Exponential for Exponential Tilting, quadratic for Loss Tilting under our choice $\ell(x) = 2^{-1}x^2$ and without knowing *a priori* what the shape of the *CTM* probabilities g_j *versus* $\hat{l}_j$, $j \in \{1, ..., N\}$ would be.
That was indeed the case for the first two algorithms. When $p = 1$, of course Exponential Tilting and CTM coincided, given that we used the first as an element-wise Importance distribution, and so do the resampling probabilities they provided. For CTM, when p > 1, the weights seem rather a noisy version of exponential tilting, which speaks of its mixture nature. Moreover, they are less unequally distributed along the different influemce values with respect to Exponential tilting, which is good since as seen on section 2.3.2 Exponential Tilting over-shifts the distribution.

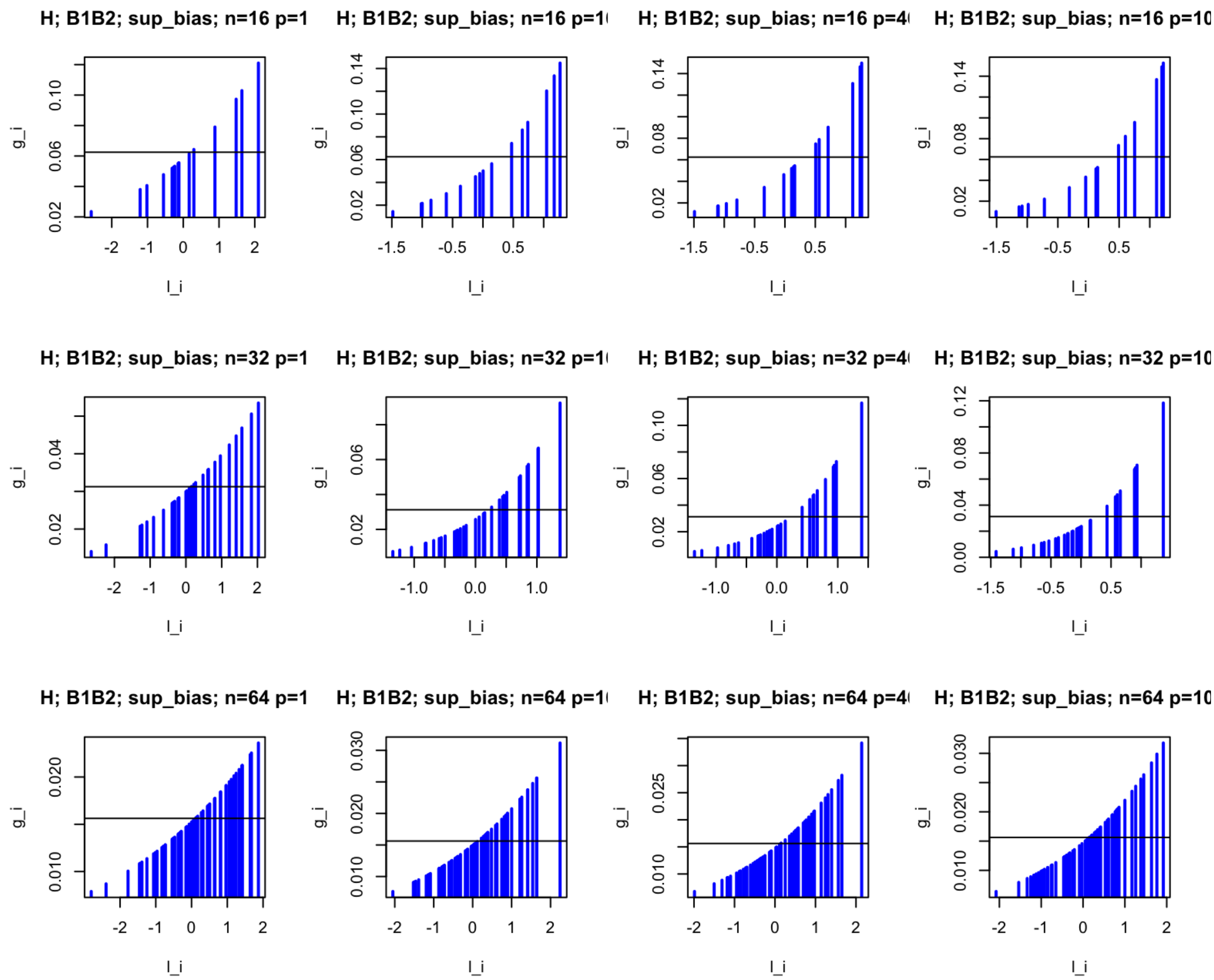

Figure 2.12: Importance resampling probabilities for statistic *sup-bias* with Exponential Tilting as a function of the empirical influence values. Horizontal line is N^{-1}

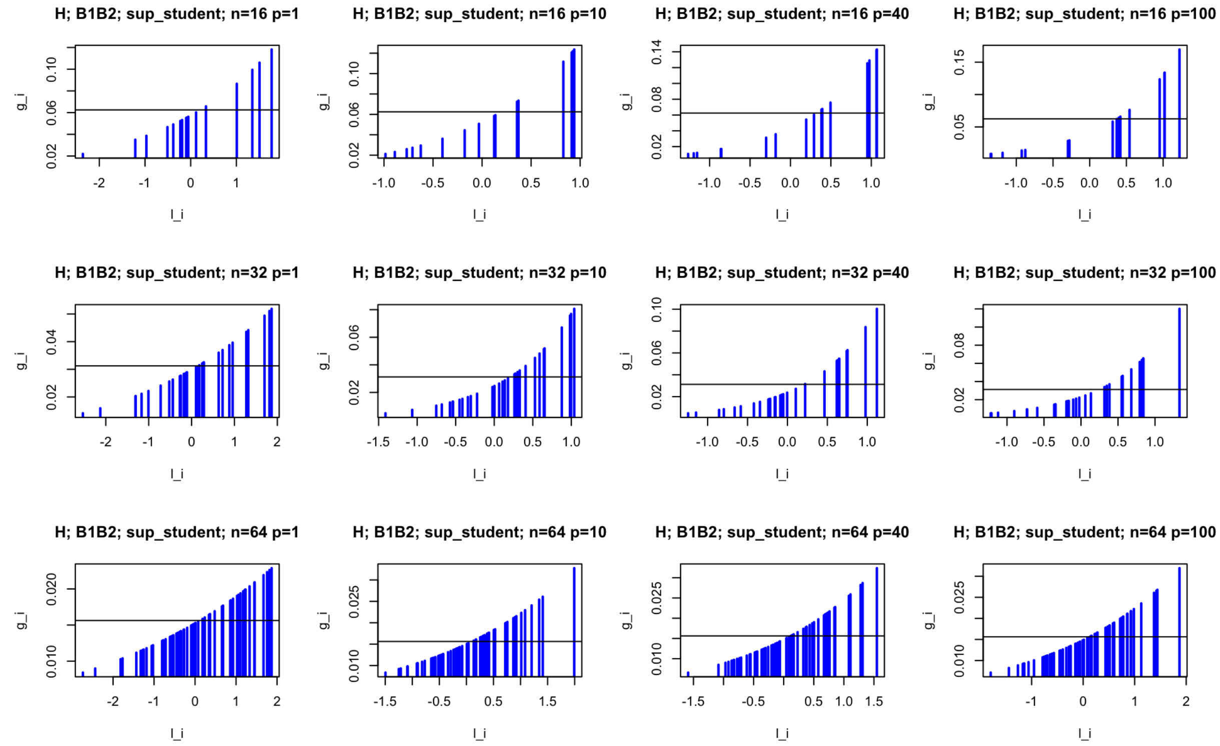

Figure 2.13: Importance resampling probabilities for statistic *sup-t-student* with Exponential Tilting as a function of the empirical influence values.. Horizontal line is N^{-1}

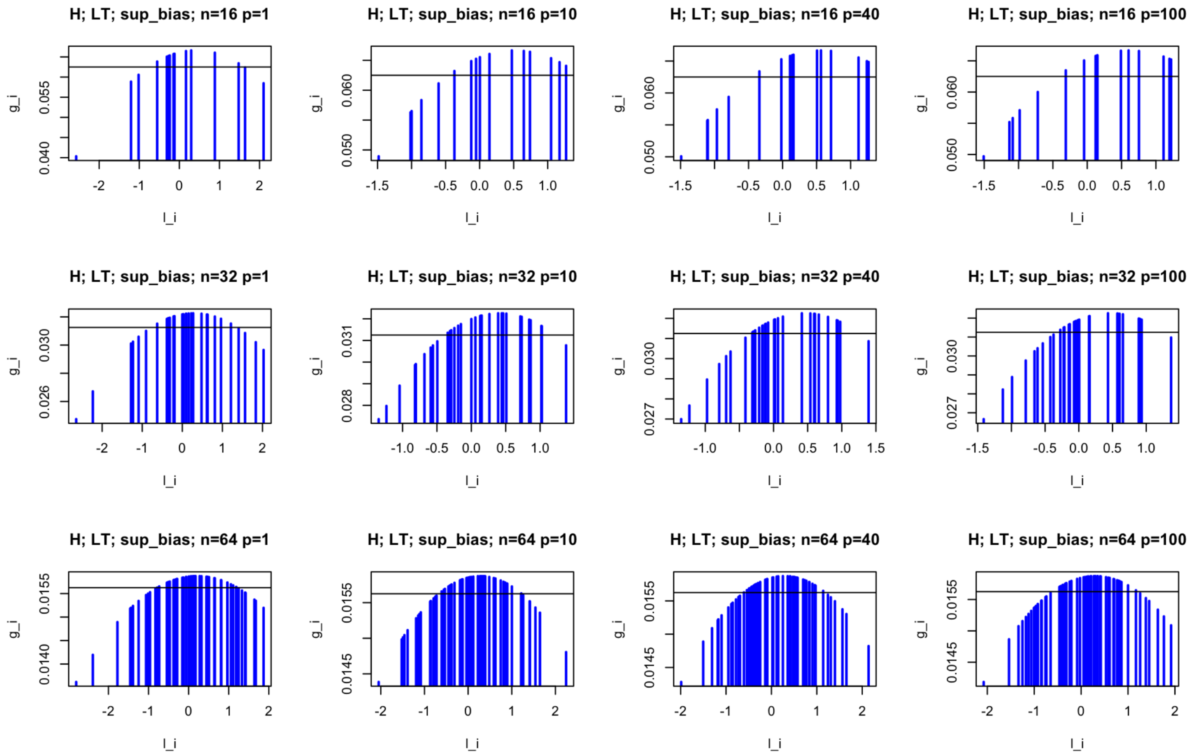

Figure 2.14: Importance resampling probabilities for statistic *sup-bias* with Loss Tilting as a function of the empirical influence values. Horizontal line is N^{-1}

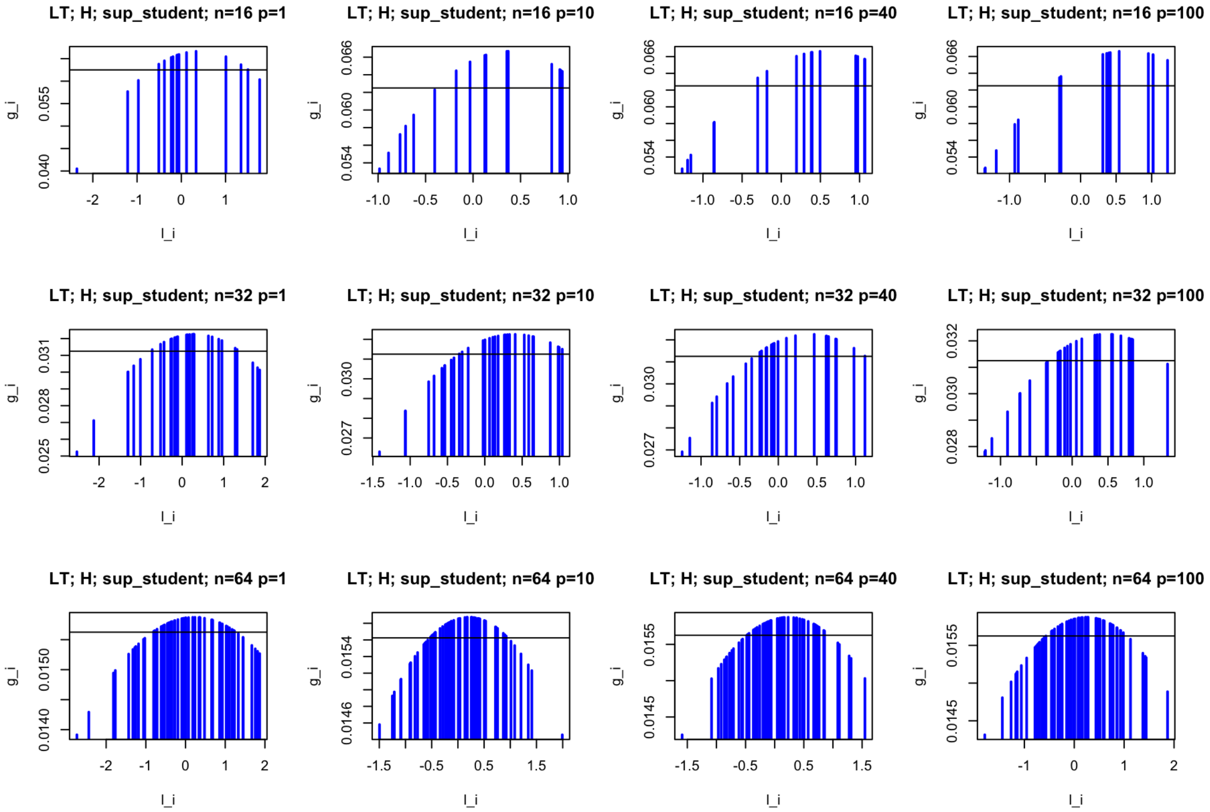

Figure 2.15: Importance resampling probabilities for statistic *sup-t-student* with Loss Tilting as a function of the empirical influence values. Horizontal line is N^{-1}

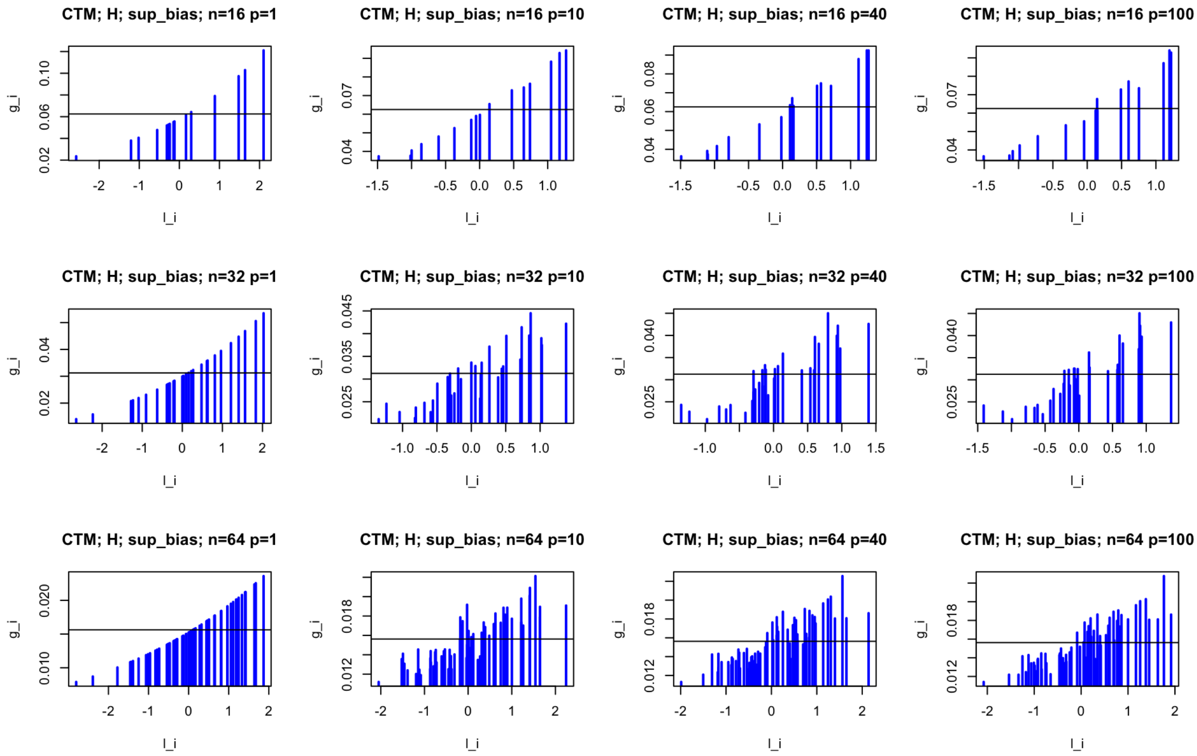

Figure 2.16: Importance resampling probabilities for statistic *sup-bias* with CTM as a function of the empirical influence values. Horizontal line is N^{-1}

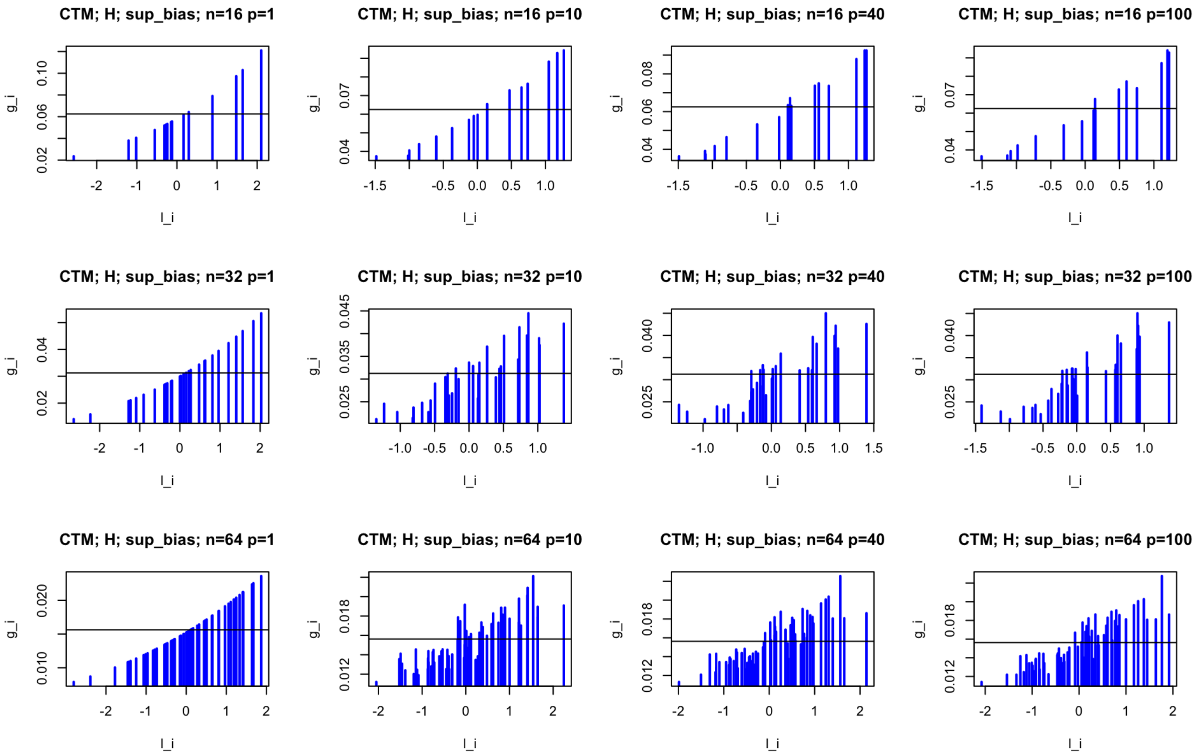

Figure 2.17: Importance resampling probabilities for statistic *sup-t-student* with CTM as a function of the empirical influence values. Horizontal line is N^{-1}

2.3.4. Comparison of the likelihood ratios

The focus of this subsection is to analyse the visual output of the ordered weights, i.e. the likelihood ratios that make up for the fact that we resampled from a distribution with unequal probabilities (see (1.17)), on the specific case of the statistic *sup-t-student*. What is more, we wish to comprehend better what is asserted in Remark (4), whence a different method is used when estimating the quantiles of order higher than 0.5.
As pointed out in Remark (4), the weights associated to lower values of the statistic $\hat{T}^*$ did explode, which justifies estimating the right tail if a variance reduction is desired. Indeed, if the weights associated to the lowest values of the MC approximation of the Bootstrap distribution $\hat{T}^*$ were used, as it can be inferred from Equation (1.13) the variance would indeed increase. In addition, as p increased, in Exponential Tilting the likelihood ratio associated to the lowest values of $\hat{T}^*$ increased, which is coherent with what we mentioned in Section 2.3.2: the importance distribution over-tilted the statistic so that there was little overlap between the distribution of $\hat{T}^*$ under ordinary resampling and tilted $\hat{T}^*$, whence these exploding weights were not a surprise. We also comment that the weights in CTM

tended to be the lowest ones, which is of course desirable for the variance reduction task, and distiguishes such algorithm as the best one under this criterion.

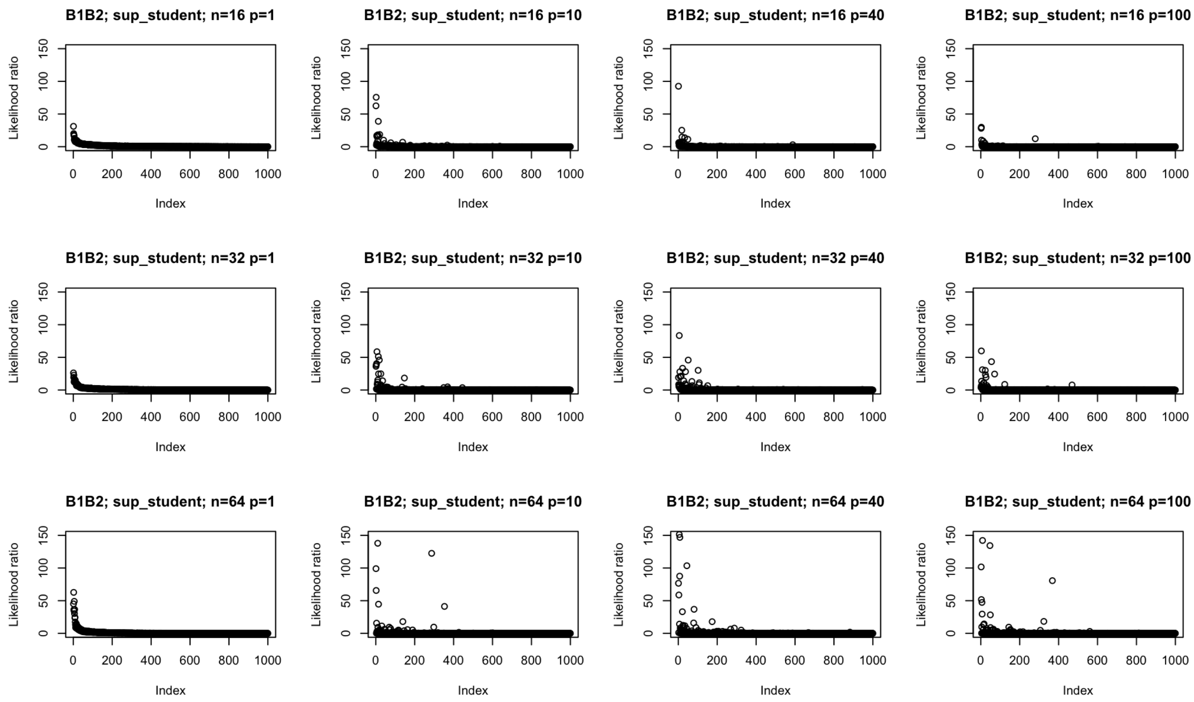

Figure 2.18: Importance resampling weights for *sup-t-student* with Exponential Tilting. The index corresponds to the order of the obtained $\hat{T}^*$ at each iteration, so that the left-most weights correspond to the lowest values of $\hat{T}^*$.

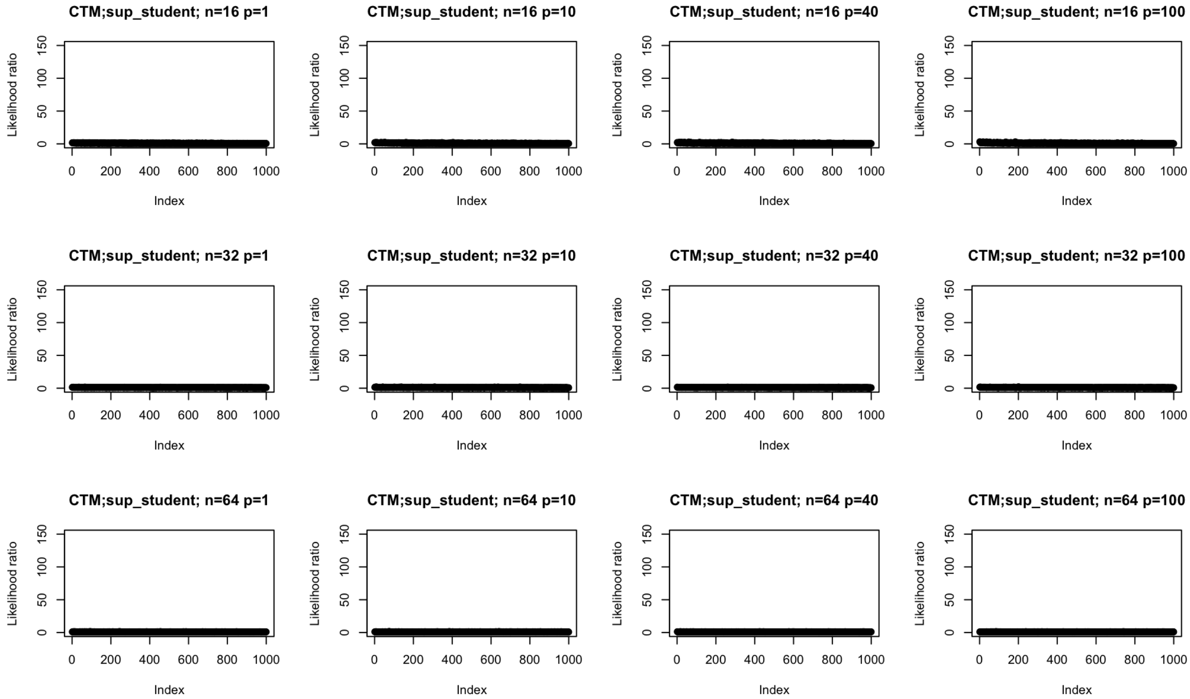

Figure 2.19: Importance resampling weights for *sup-t-student* with Loss Tilting. The index corresponds to the order of the obtained $\hat{T}^*$ at each iteration, so that the left-most weights correspond to the lowest values of $\hat{T}^*$.

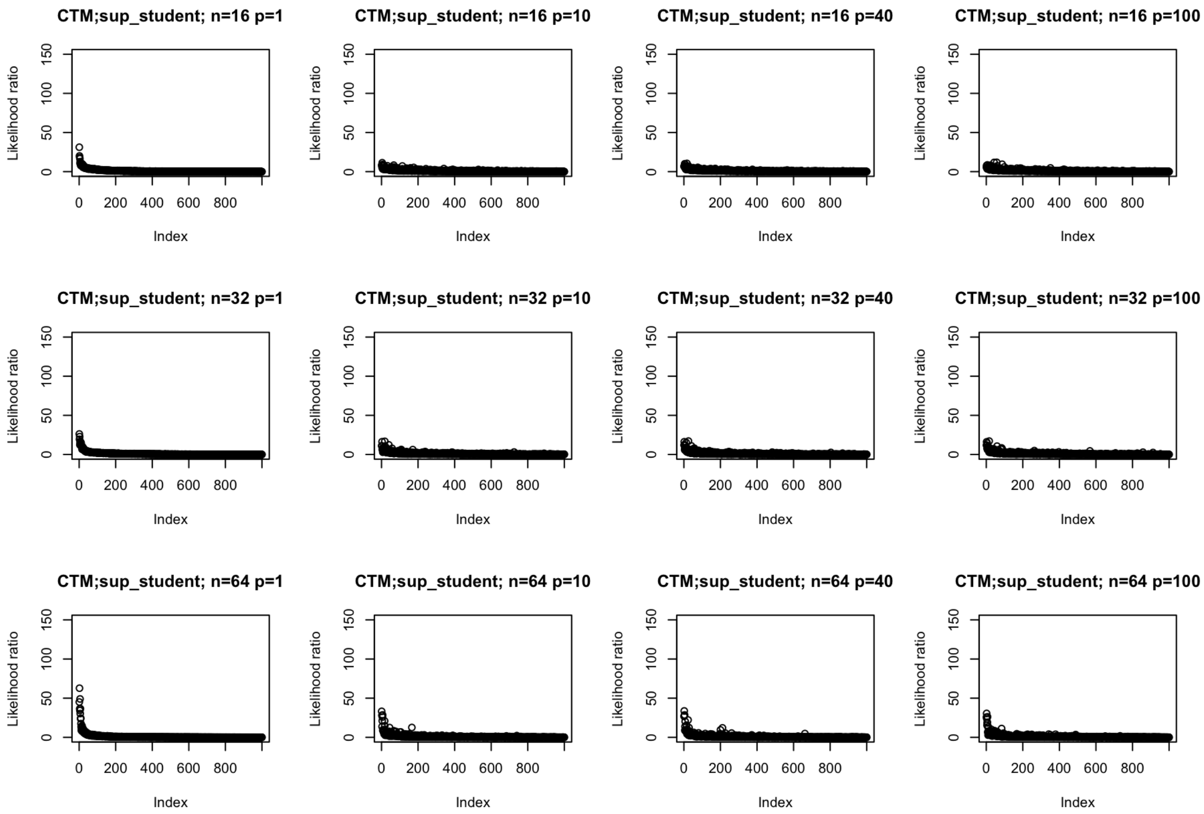

Figure 2.20: Importance resampling weights for *sup-t-student* with CTM. The index corresponds to the order of the obtained $\hat{T}^*$ at each iteration, so that the left-most weights correspond to the lowest values of $\hat{T}^*$.

2.3.5. Comparison of the Bootstrap versus the sampling distributions

In this subsection, we explored the distributions of the Bootstrap distribution $\hat{T}^*$ and the sampling distribution $\hat{T}$, for different N and p, as well as their $1-\alpha$th quantiles $\xi_{1-\alpha}$ and $\hat{\xi}_{1-\alpha}$, respectively. We did so for both statistics to have a visual of how well the Bootstrap estimate was w.r.t the true sampling distribution. We used $B = 50000$ replications in both cases, in order to render their MC error insignificant. We note that not only the quantile estimated under Bootstrap but the whole Bootstrap distribution is were close to the sampling one. This is an indication that that the Bootstrap error tends to be small, and thus that focusing on reducing the MC eror (see Remark 1) is a worthwhile task. These results are coherent with what is stated in Hesterberg (2014), *id est*, the Bootstrap does a fine job for approximating the sampling distribution for studentised statistics.

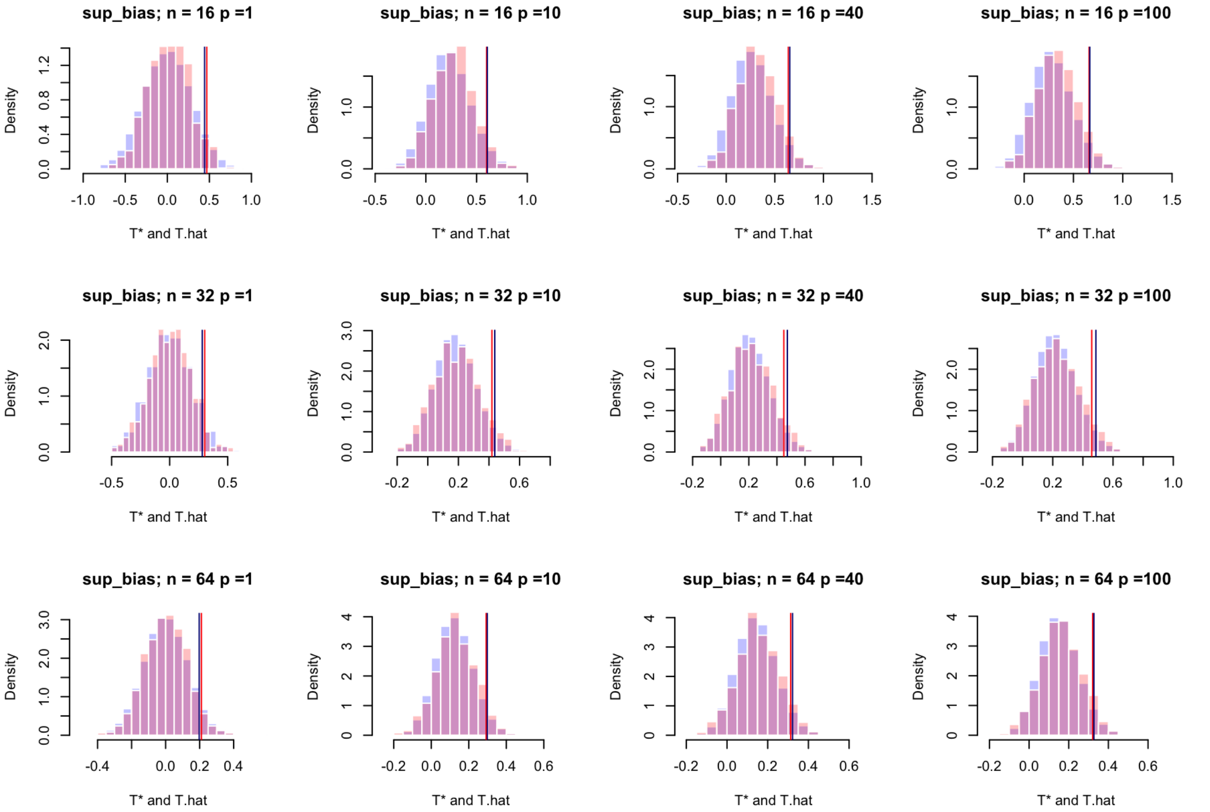

Figure 2.21: Bootstrap (blue) and actual sampling distribution (pink) of *sup-bias*, as well as their $1-\alpha$th quantiles, respectively.

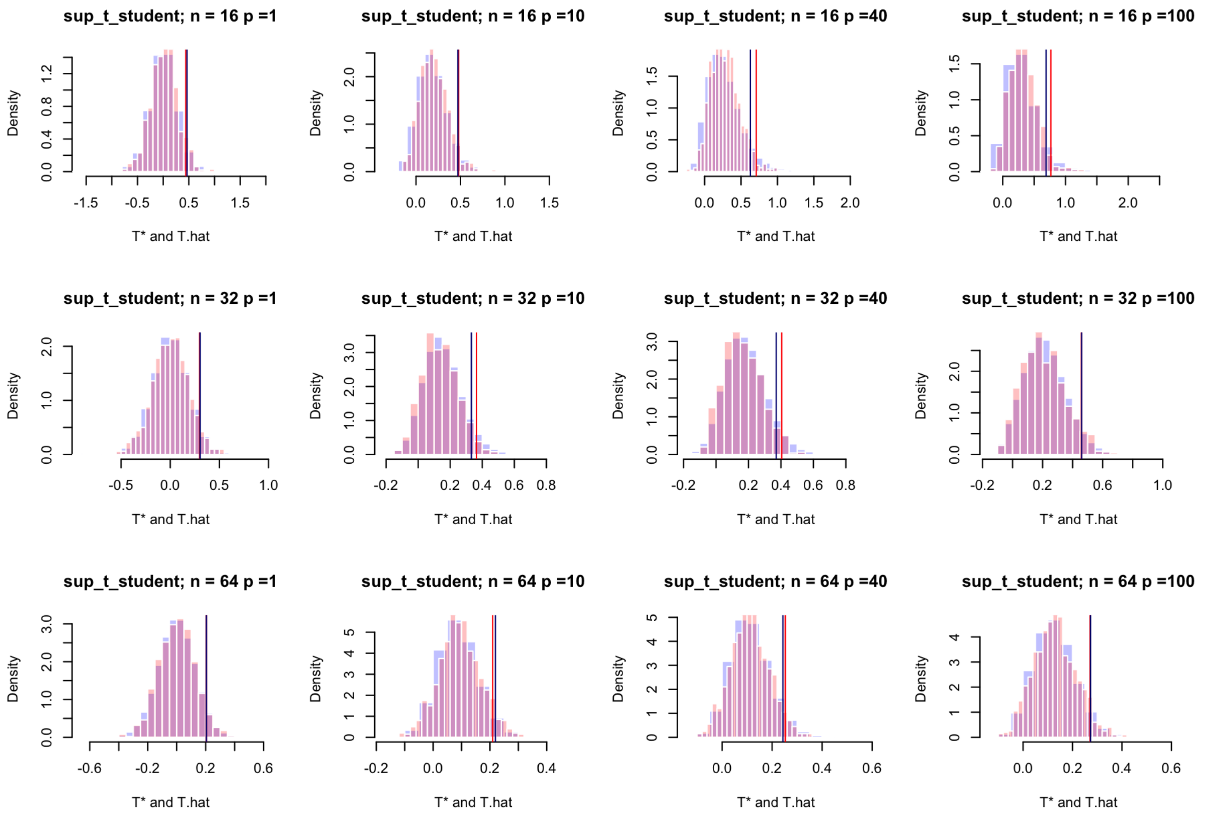

Figure 2.22: Bootstrap (blue) and actual sampling distribution (pink) of *sup-t-student*, as well as their $1-\alpha$th quantiles, respectively.

2.3.6. Efficiency comparison

Last but not least, we evaluated how good the variance reduction achieved with Importance Resampling was with respect to ordinary Resampling, which is the core interest of this thesis.
We measure the efficiency as the variance of Ordinary Bootstrarp divided by the variance under Importance Resampling; using 1000 replications from the sample original sample. Following Davison and Hinkley (1997), we estimated the efficiency in the following way:

$$e\hat{f}f = \frac{\hat{\sigma}^2_{MC}}{\hat{\sigma}^2_{IS}} \tag{2.24}$$

where $\sigma^{\hat{2}}_{MC}$ is the estimate of the variance of the quantile estimate under ordinary resampling, and $\sigma^{\hat{2}}_{IS}$ under Importance Resampling.
To obtain such estimates, for 1000 times, for each N and each p, we drew samples from

the underlying data-generating-processes (characterised by different p), performed the simulation of $B = 1000$ replications with Ordinary Resampling and Importance Resampling, with Exponential Tilting, Loss Tilting and CTM as before for the choice of the importance distribution in the latter case.

Remark 11 (Definition of efficiency)**.** *Since we are dealing with MC estimates of Bootstrap estimates, another possibility was to look at the Mean Square Error (MSE) of the MC estimate of the Bootstrap estimate with respect to the true quantile of the underlying sampling distribution, taking into account both MC and Bootstrap error.*
However, as pointed out in Davison and Hinkley (1997), since the expected value of the MC estimate is the same for both Ordinary and Importance Resampling, the bias term in the bias-variance decomposition of the MSE is negligible.
We confirmed it with our own simulations, so we decided show analyse the ratio of the variance of both MC estimates as a proxy for efficiency, analogously to Davison and Hinkley (1997).

Moreover, from such 1000 iterations, we derived histograms of the ordinary and importance resampling quantile estimates, to have a visual of their variance. The results were coherent with the previous subsections of the Simulation study. Across statistics, the results were similar. Exponential tilting **failed** when $p > 1$, which had been anticipated in 2.3.2: indeed since the linearisation was not a good approximation, Assumption (2) was violated and the classical method to yield an importance distribution for Importance Resampling led to significant losses of efficiency.
Loss Tilting, the most conservative, was always slightly above 1 meaning it is worthwile. Given that we chose a very generic loss function, what we mentioned before about it being case specific is confirmed: one should choose a loss function according to the needs of the particular problem to get a signficant increase in the efficiency. **CTM** is undoubtedly is the best algorithm! Whereas efficiencies close to 13 (as reported by Davison when both assumptions 2 & 3 hold strictly), it still manages to have very high efficiencies when the Assumptions behind under Exponential Tilting are clearlyviolated (which happens when $p \geq 10$) which speaks of great power. The fact it is able to yield a higher efficiency for $p = 100$ demonstrates its strong affinity to the task of building Bootstrap SCBs for functional data.

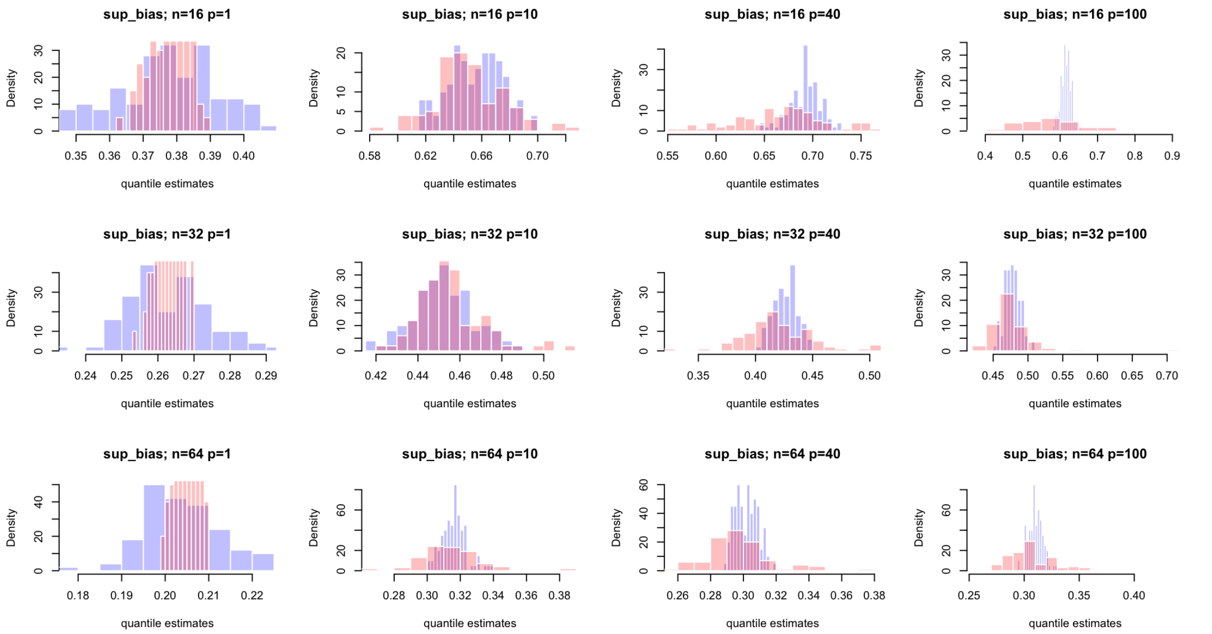

Figure 2.23: For 1000 different samples, values of the MC estimate of the Bootstrap estimate's quantile of order $1-\alpha$ for *sup-bias* under ordinary (blue) and importance (red) resampling with Exponential Tilting

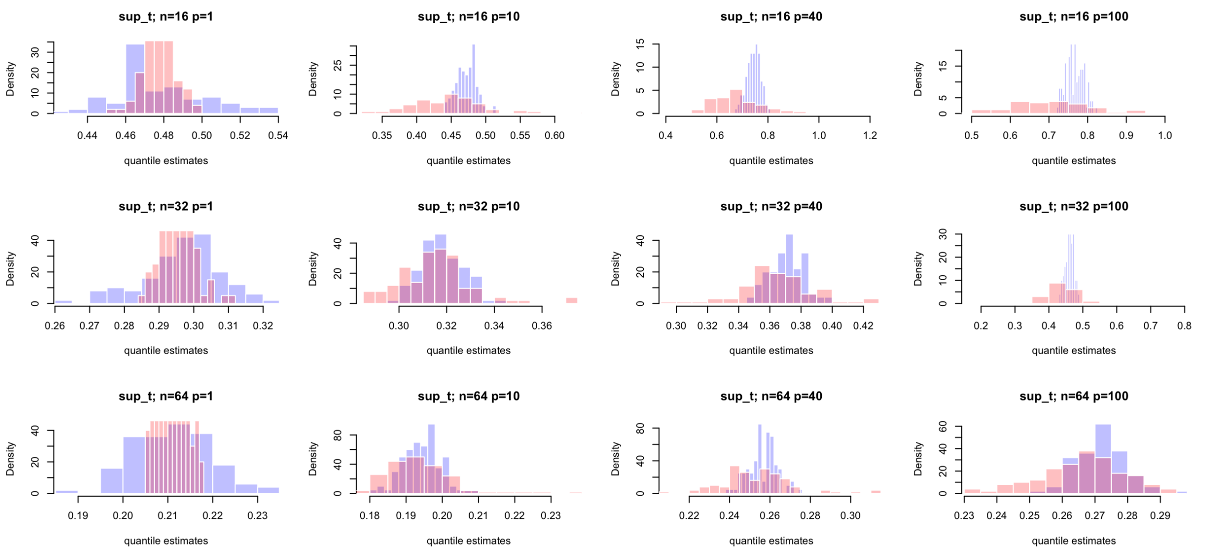

Figure 2.24: For 1000 different samples, values of the MC estimate of the Bootstrap estimate's quantile of order $1-\alpha$ for *sup-t-student* under ordinary (blue) and importance (red) resampling with Exponential Tilting

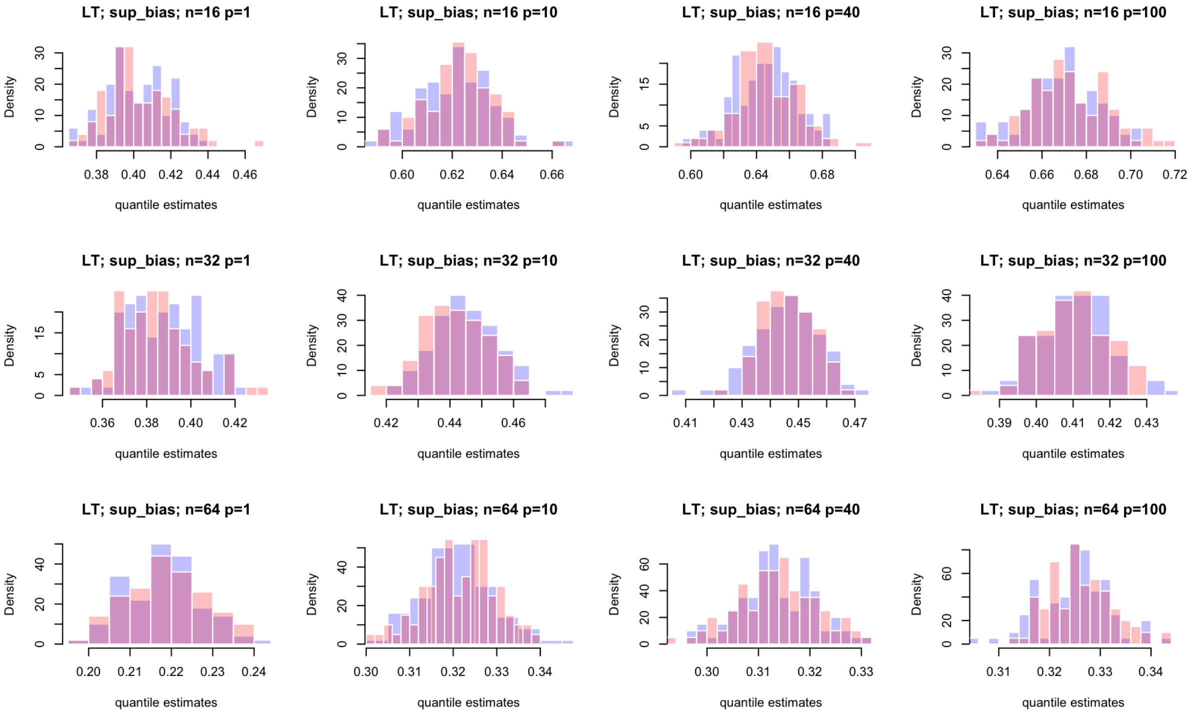

Figure 2.25: For 1000 different samples, values of the MC estimate of the Bootstrap estimate's quantile of order $1-\alpha$ for *sup-bias* under ordinary (blue) and importance (red) resampling with Loss Tilting

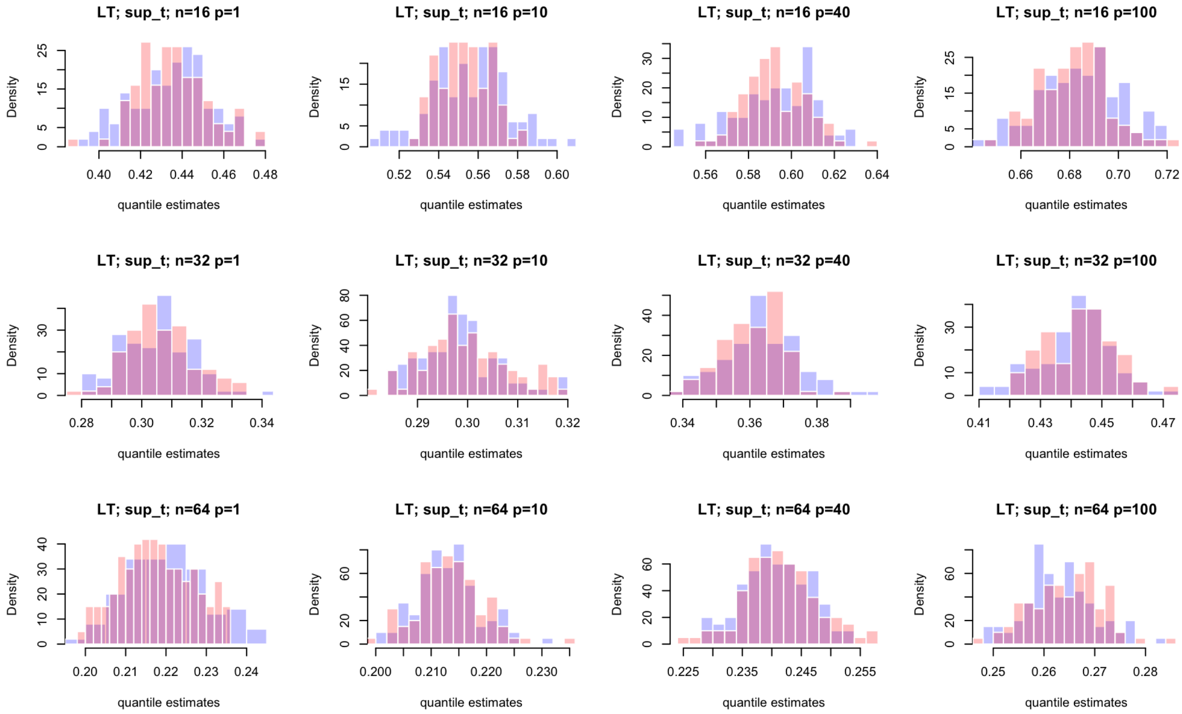

Figure 2.26: For 1000 different samples, values of the MC estimate of the Bootstrap estimate's quantile of order $1-\alpha$ for *sup-t-student* under ordinary (blue) and importance (red) resampling with Loss Tilting

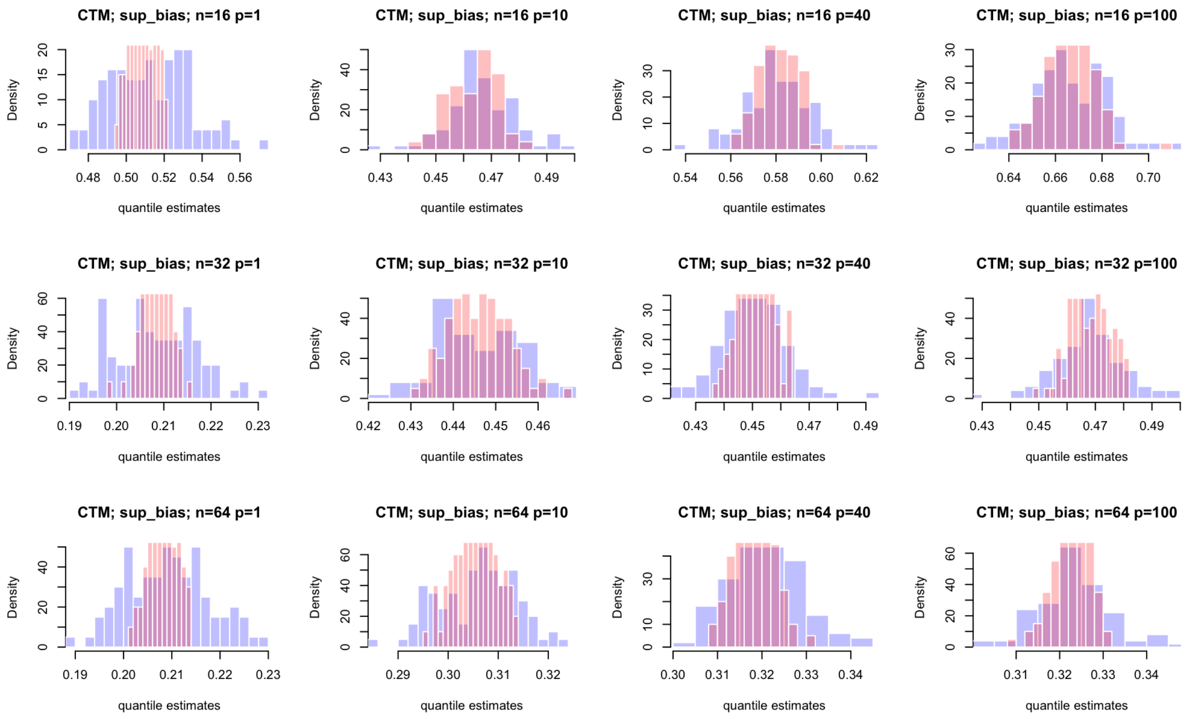

Figure 2.27: For 1000 different samples, values of the MC estimate of the Bootstrap estimate's quantile of order $1-\alpha$ for *sup-bias* under ordinary (blue) and importance (red) resampling with CTM

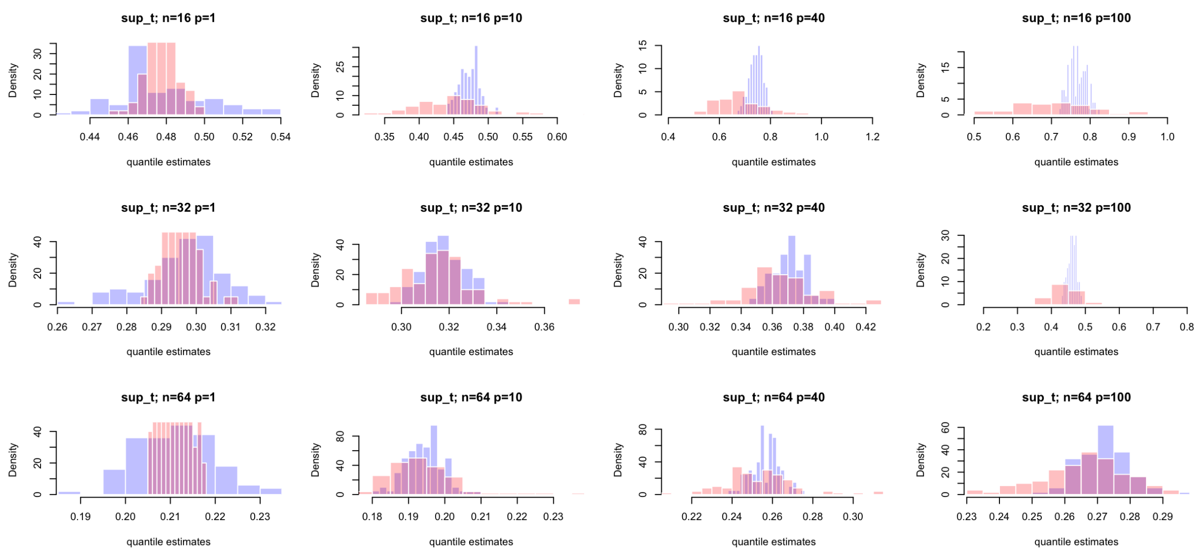

Figure 2.28: For 1000 different samples, values of the MC estimate of the Bootstrap estimate's quantile of order $1-\alpha$ for *sup-t-student* under ordinary (blue) and importance (red) resampling with CTM

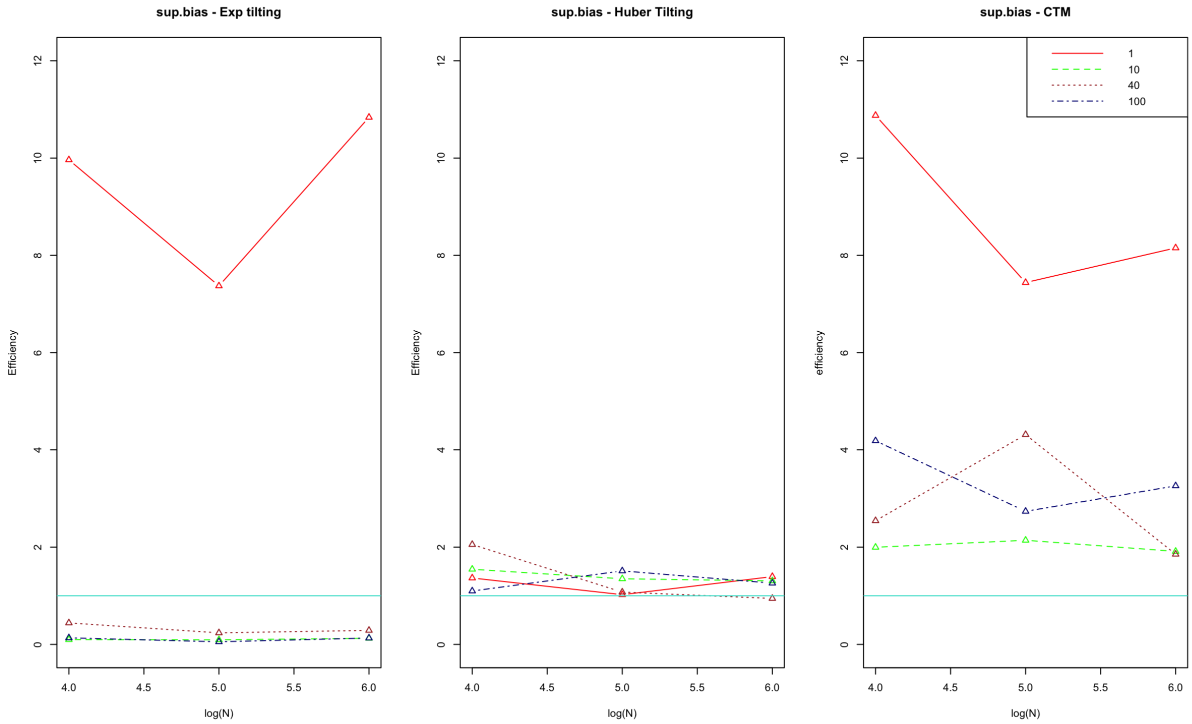

Figure 2.29: Efficiency curves for statistic *sup-bias*. Each curve represents a value of p

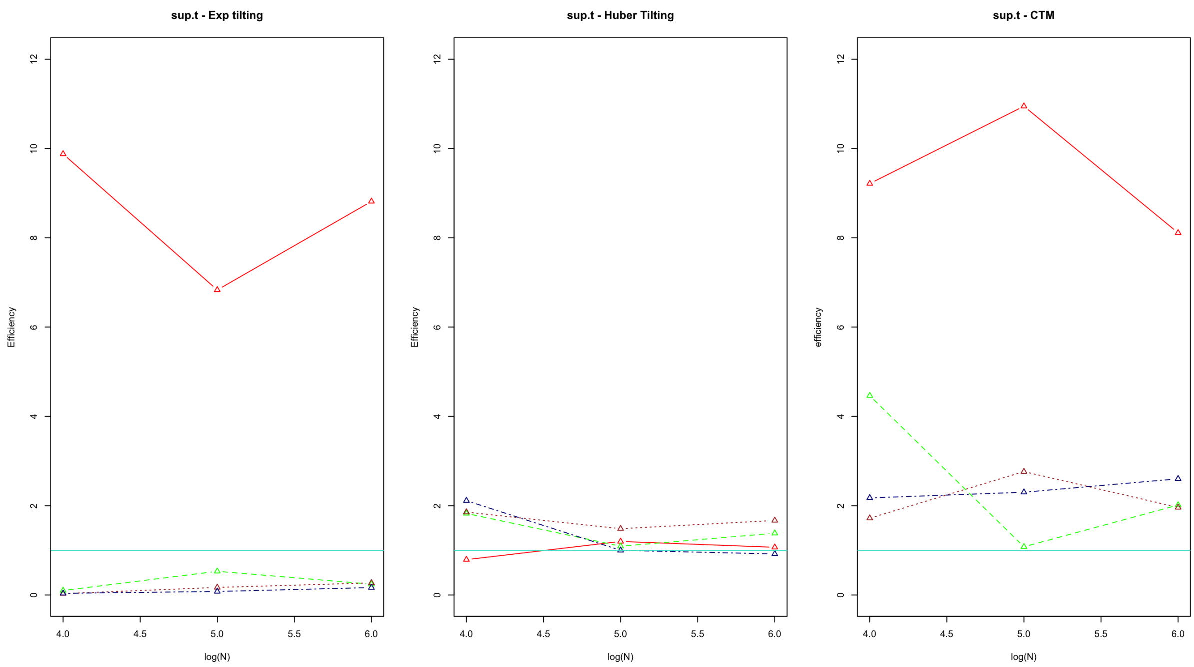

Figure 2.30: Efficiency curves for statistic *sup-t-student*. Each curve represents a value of p

Remark 12. *(Inherent variance of the Bootstrap efficiences) Even if* 1000 *replications were made on the same sample to estimate Bootstrap efficiency, values would oscillate present significant variance*
Since the sample is random as well, this is due to the fact we are dealing with the sampling distribution.

3 | Simultaneous Confidence Bands for Functional Data

3.1. Overview of the Boostrap for building SCBs

As mentioned before, Degras (2011) is the reference paper on building SCBs through the Bootstrap, in the sense it is one of the most cited works, and other researches that have been published afterwards are very similar in their methodology. Prior to such paper, Cuevas and Fraiman (2004) provided an overview of different variations of the Bootstrap to build SCBs. Indeed, in this thesis we have used the so-called naïve Bootstrap: the re-samples are taken from the original sample, and Importance Resampling, where we re-sample with different weights while correcting with the likelihood ratio. In Cuevas and Fraiman (2004), an example of a variation that is used is smooth Bootstrap (not to be confused with Smoothing in Appendix A), wherein at each iteration each re-sampled statistical unit is corrupted by adding a small random gaussian noise. This is necessary when using depth-like estimators for the location of a stochastic process (see also López-Pintado and Romo (2009)). Another option is parametric Bootstrap, where the underlying distribution F of the available observations is to be assumed of a given parametric form, and (re)samples are taken from such distribution, whose parameters are estimated from the original sample.

Cuevas and Fraiman (2004)'s paper, however, uses the concept of functional depth (López-Pintado and Romo (2009)) to build SCBs: a percentage (corresponding to the desired coverage) of the deepest Bootstrapped curves are utilised to generate them, avoiding entirely the use of a statistic of the form of (1.29) which to our knowledge would render Importance Resampling inapplicable. Another example which prescinds such shape for the SCBs is (Staszewska-Bystrova and Winker (2013)), where the problem is posed from an an operations research point of view.

Nevertheless, the latest contributions in Bootstrap SCBs have proposed bands in the form of Equation (1.29), which makes them idoneous the use of Importance Resampling. Even (Narisetty and Nair (2016)), who uses functional depths as well, recurs to a formula of such kind. Indeed, such approach is the state of the art for Bootstrap SCBs. The main aspects in which the latest contributions have varied to achieve a more accurate coverage (*i.e* for a given confidence level $1 - \alpha$, the actual coverage for the true mean of the underlying data generation process is close to such value) could be classified into two:

- **Variations of the Bootstrap**. We have already mentioned smooth and parametric Bootstrap with the latter also being seen also seen on Antoniadis et al. (2016), Paparoditis and Shang (2021), Goldsmith et al. (2013), amongst others, who assume, although not necessarily directly, some parametric model behind the data generating process and produce a Bootstrap distribution by sampling from it.
  Another variation is the Wild Bootstrap: instead of adding a white noise to the statistical units, these are directly multiplied by a random variable whose distribution is such that the obtained Bootstrap distribution is richer (in terms of a larger support, for *e.g.*) whilst not violating the (minimal) assumptions behind the Data Generating Process.
  Indeed, one of the best algorithms in terms of coverage nowadays is proposed in Telschow and Schwartzman (2022), which uses multipliers of a Rademacher distribution to estimate the Bootstrap distribution of a Degras (2011)-like statistic. It assumes symmetry of the departures of each statistical unit from the mean, though.
  One last approach we mention is the one in (Chuang et al. (2013)), which is a version of the Bootstrap for dependent (thus non i.i.d) statistical units called the Block Bootstrap.

- **What is being Bootstrapped**. Whereas in most papers re-samples are taken from the available functions (or their for *e.g.* smoothing estimates when necessary) in an naïve Bootstrap fashion, other works have considered other options. Goldsmith et al. (2013), Wang et al. (2020), Wang et al. (2020) (among others) recur to the so-called Functional Principal Component Analysis (FPCA) to decompose the statistical units (which are functions), and each FPC is multiplied by a unit variance gaussian random noise (hence a wild Bootstrap).

Remark 13 (Improvements for the Bootstrap)**.** *Whereas most of the latest works focus on increasing the empirical coverage (especially when N) is small, in this thesis we alter ordinary Bootstrap to reduce the variance of the estimate.*

In view of Remark (13), we performed a short simulation study to evaluate the potential need for variance reduction in the construction of Bootstrap SCBs.

3.2. Brief simulation study: the need for variance reduction

In this section we show the experiment we carried out to evaluate the need for Importance Resampling when building Bootstrap SCBs.
The idea was to see, for different values of B, *id est* the number of MC iterations to estimate the Bootstrap distribution of a statistic, the variance in the quantile estimation of a statistic of interest.
We chose Degras (2011)'s statistic, and for varying N and p, we drew a sample and performed a 4 Bootstrap simulations at different values of B to see its impact on the estimate across simulations that approximated the Bootstrap distribution of the statistic under the same initial random sample. The quantile of the statistic we estimated was $1 - \alpha = 0.5$. We utilised the Gaussian Process in Section 2.3 and chose a an accurate enough discretisation for the functional data.
We use the statistic just as in Degras (2011), that is, the *sup* of the element-wise absolute value of Student's t statistic (see Equation (1.31). In Chapter 2 we had used its signed version (1.34), since with the absolute value the optimisation run in Exponential Tilting would not converge using the absolute value. This makes sense since if the support of the statistic had been $\mathbb{R}^+$, automatically Assumption (3) would have been violated.
The results are very significant. Whilst in the literature values of $B = 500$ (Cuevas and Fraiman (2004)) or $B = 2500$ (Telschow and Schwartzman (2022)) are used, the estimates had a stabilised from $B = 5000$ (little variance) and completely in $B = 10000$. In other words, with the usual values for used in state-of-the-art algorithms, the MC error is still significant, which calls for the use of Importance Resampling.

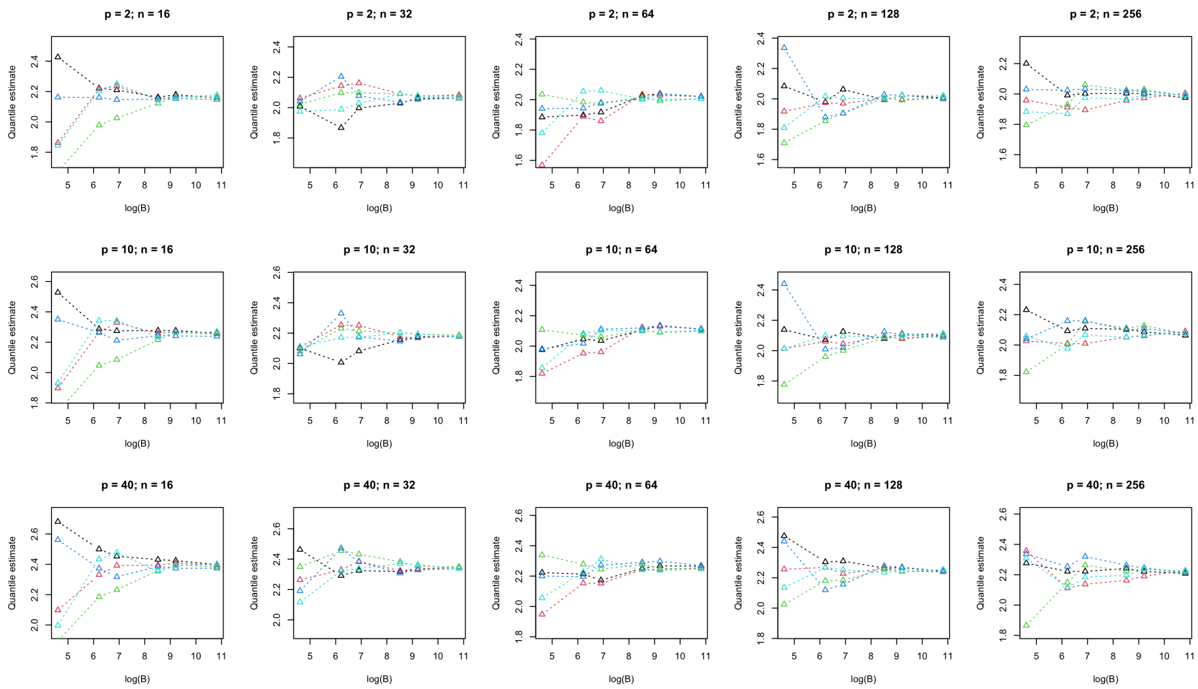

Figure 3.1: Curves of the MC quantile estimate of statistic (1.31) for four different random seeds, starting from the same sample at the different n and p.

4 | Conclusions and future developments

In this thesis, we have applied Importance Resampling for the quantile estimation for statistics such as the one Degras (2011), targeted to build SCBs for high-dimensional data. In particular, we have tested Exponential Tilting, a methodology present in the classic works on the subject matter, namely Johns (1988), Do and Hall (1991) and Davison (1988), used in Importance Resampling to yield the importance distribution. Such procedure relies on two major assumptions that regard the accuracy of the linear approximation of the statistic of interest and the normality of such proxy. The statistical functions that we have analysed, used for the Bootstrap SCBs, care the composition of some function and the sup (inf) operator, which increases nonlinearity, violating both assumptions.
Through a simulation study, we have seen that Exponential Tilting indeed fails in the setting of SCB construction with the Bootstrap, leading to increased variance with respect to Ordinary Resampling when $p > 1$. This implies that in such setting, we can (and should) do much better than Exponential Tilting to reduce the MC variance. We have thus provided two algorithms, namely Loss Tilting (LT) and Contribution Tilted Mixture (CTM). In the simulation study, we have shown that both increase the efficiency and conseuqnelty work better than Exponential Tilting, even for large values of p, when the non-linearity increases significantly. LT was chosen with a very simple loss function and managed to get consistent yet small increases in variance reduction. CTM, on the other hand, has led to significant variance reduction across all explored values of p, proving its power and potential applicability.

Finally, we have also carried out a brief experiment that confirms the need for Importance Resampling in the setting of SCB estimation through the Bootstrap.
Regarding future research, we remark that different possibilities could be explored for Exponential Tilting. Instead of a pilot run and an importance resampling run, a sequence of importance resampling runs could be carried out, updating the importance distribution accordingly. On the other hand, LT could be exploited to adjust the importance distribu-

tion at a certain desired variance, so that it is tuned according to the degree of certainty on a given initial quantile estimate.

Of course, Importance Resampling is not the only variance reduction technique for the Bootstrap, and given the proven need for such task in the construction, we conjecture that both a modification of the Bootstrap as seen in Chapter 3 as well an adaptation of a more general MC variance reduction technique would lead to better results across a (hopefully) wider range of statistics.

Bibliography

A. Antoniadis, X. Brossat, J. Cugliari, and J.-M. Poggi. A prediction interval for a function-valued forecast model: Application to load forecasting. *International Journal of Forecasting*, 32(3):939–947, 2016.

F. Bunea, A. E. Ivanescu, and M. H. Wegkamp. Adaptive inference for the mean of a gaussian process in functional data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):531–558, 2011.

A. Canty and B. D. Ripley. *boot: Bootstrap R (S-Plus) Functions*, 2022. R package version 1.3-28.1.

S.-C. Chuang, Y.-C. Hung, W.-C. Tsai, and S.-F. Yang. A framework for nonparametric profile monitoring. *Computers & Industrial Engineering*, 64(1):482–491, 2013.

A. Cuevas and R. Fraiman. On the bootstrap methodology for functional data. In J. Antoch, editor, *COMPSTAT 2004 — Proceedings in Computational Statistics*, pages 127–135, Heidelberg, 2004. Physica-Verlag HD. ISBN 978-3-7908-2656-2.

A. Davison. Discussion of the royal statistical society meeting on the bootstrap. *Journal of the Royal Statistical Society, Series B*, 50:356–357, 1988.

A. C. Davison and D. V. Hinkley. Saddlepoint approximations in resampling methods. *Biometrika*, 75(3):417–431, 1988.

A. C. Davison and D. V. Hinkley. *Bootstrap Methods and their Application*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1997. doi: 10.1017/CBO9780511802843.

D. Degras. Simultaneous confidence bands for nonparametric regression with functional data. *arXiv: Methodology*, 2009.

D. Degras. *SCBmeanfd: Simultaneous Confidence Bands for the Mean of Functional Data*, 2016. URL `https://CRAN.R-project.org/package=SCBmeanfd`. R package version 1.2.2.

D. A. Degras. Simultaneous confidence bands for nonparametric regression with functional data. *Statistica Sinica*, pages 1735–1765, 2011.

T. J. DiCiccio and B. Efron. Bootstrap confidence intervals. *Statistical science*, 11(3): 189–228, 1996.

K.-A. Do and P. Hall. On importance resampling for the bootstrap. *Biometrika*, 78(1): 161–167, 1991.

B. Efron. Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, 7 (1):1 – 26, 1979. doi: 10.1214/aos/1176344552. URL `https://doi.org/10.1214/aos/1176344552`.

B. Efron. More efficient bootstrap computations. *Journal of the American Statistical Association*, 85(409):79–89, 1990. ISSN 01621459. URL `http://www.jstor.org/stable/2289528`.

B. Efron and R. J. Tibshirani. *An Introduction to the Bootstrap*. Number 57 in Monographs on Statistics and Applied Probability. Chapman & Hall/CRC, Boca Raton, Florida, USA, 1993.

J. Goldsmith, S. Greven, and C. Crainiceanu. Corrected confidence bands for functional data using principal components. *Biometrics*, 69(1):41–51, 2013.

P. Hall. Antithetic resampling for the bootstrap. *Biometrika*, 76(4):713–724, 1989. ISSN 00063444. URL `http://www.jstor.org/stable/2336631`.

P. Hall. *The Bootstrap and Edgeworth Expansion*. Springer Series in Statistics. Springer New York, 1992. ISBN 9780387977201. URL `https://books.google.it/books?id=ReTRngEACAAJ`.

T. Hesterberg. What teachers should know about the bootstrap: Resampling in the undergraduate statistics curriculum, 2014.

D. V. Hinkley and S. Shi. Importance sampling and the nested bootstrap. *Biometrika*, 76(3):435–446, 1989. ISSN 00063444. URL `http://www.jstor.org/stable/2336110`.

F. Ieva, A. M. Paganoni, J. Romo, and N. Tarabelloni. roahd Package: Robust Analysis of High Dimensional Data. *The R Journal*, 11(2):291–307, 2019. URL `https://doi.org/10.32614/RJ-2019-032`.

M. V. Johns. Importance sampling for bootstrap confidence intervals. *Journal of the American Statistical Association*, 83(403):709–714, 1988. ISSN 01621459. URL `http://www.jstor.org/stable/2289294`.

S. Lee and I. Wong. A hybrid approach based on saddlepoint and importance sampling methods for bootstrap tail probability estimation. *Statistics and Computing*, 12:209–217, 07 2002. doi: 10.1023/A:1020742625430.

S. López-Pintado and J. Romo. On the concept of depth for functional data. *Journal of the American statistical Association*, 104(486):718–734, 2009.

A. Manzoni. Computational statistics class notes, 3rd edition. WeBeep, September 2022.

N. N. Narisetty and V. N. Nair. Extremal depth for functional data and applications. *Journal of the American Statistical Association*, 111(516):1705–1714, 2016.

E. Paparoditis and H. L. Shang. Bootstrap prediction bands for functional time series. *Journal of the American Statistical Association*, pages 1–15, 2021.

A. Pini and S. Vantini. Interval-wise testing for functional data. *Journal of Nonparametric Statistics*, 29(2):407–424, 2017.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2023. URL `https://www.R-project.org/`.

J. O. Ramsay and B. W. Silverman. *Functional Data Analysis*. Springer, 2005. ISBN 9780387400808. URL `http://www.worldcat.org/isbn/9780387400808`.

P. Secchi, A. Stamm, and S. Vantini. Inference for the mean of large p small n data: A finite-sample high-dimensional generalization of hotelling's theorem. 2013.

K. Singh. On the asymptotic accuracy of efron's bootstrap. *The annals of statistics*, pages 1187–1195, 1981.

A. Staszewska-Bystrova and P. Winker. Constructing narrowest pathwise bootstrap prediction bands using threshold accepting. *International Journal of Forecasting*, 29(2): 221–233, 2013.

F. J. Telschow and A. Schwartzman. Simultaneous confidence bands for functional data using the gaussian kinematic formula. *Journal of statistical planning and inference*, 216:70–94, 2022.

Y. Wang, G. Wang, L. Wang, and R. T. Ogden. Simultaneous confidence corridors for mean functions in functional data analysis of imaging data. *Biometrics*, 76(2):427–437, 2020.

L. Wasserman. *All of Nonparametric Statistics (Springer Texts in Statistics)*. Springer-Verlag, Berlin, Heidelberg, 2006. ISBN 0387251456.

E. Zio. *The Monte Carlo Simulation Method for System Reliability and Risk Analysis*. Springer, Jan. 2013. doi: 10.1007/978-1-4471-4588-2. URL `https://hal-centralesupelec.archives-ouvertes.fr/hal-00777141`. ISBN 978-1-4471-4588-2.

A | Appendix A: Variance Reduction methods for the Bootstrap

Several of the methods which belong to the general Monte Carlo world have been applied and researched in the Bootstrap world:

- **Antithetic resampling**. Escogitated by Hall (1989), it exploits antithethic permutations to ensure the negative covariance between couples of resamples. Each MC iteration implies retrieving two-resamples such that their estimates are negatively correlated, so that the MC expected variance is reduced.

- **Stratified sampling**. Davison and Hinkley (1997) proposes it in the case of importance resampling. Sometimes the importance distribution is a mixture of several distributions (see the Gravity example in Davison and Hinkley (1997)), or, in order not to risk an exploding variance due to a likelihood ratio between the sampling distributions skyrocketing, a **defensive mixture distribution** is used. *Id est*, the importance distribution is a mixture of the proposed one and another *defense* distribution that bounds the likelihood ratio. In this case, in a Monte Carlo experiment with total B runs, $B_0 < B$ are done sampling from the proposal distribution, and the rest from a safer distribution, such as $\hat{F}$.

- **Balanced bootstrap**. The idea is that the empirical multinomial distribution yielded by the B samples matches the one given in Equation (1.15). As mentioned both by Hall (1992) and Davison and Hinkley (1997), the intuition of this method is closely related to Latin Hypercube Sampling (Manzoni (2022)).

- **Control methods**. Davison Davison and Hinkley (1997) proposes the use of a linear approximation of the statistic both in a control variate setting and in a Multi Level Monte Carlo scheme for the bootstrap. Such linear approximation is based on the **delta method** (see Manzoni (2022), Hall (1992) and Davison and Hinkley (1997)), which is shown later in this paper when obtaining the importance distribution for estimate (1.19).

The efficiencies in the particular case of quantile estimation, which will be the case study in this article, can be visualised in the image from Hall (1992). Note that if the tail probability estimates are more efficient, provided that the importance distribution is correctly centered very similar results are expected for the quantile efficiencies.[1]

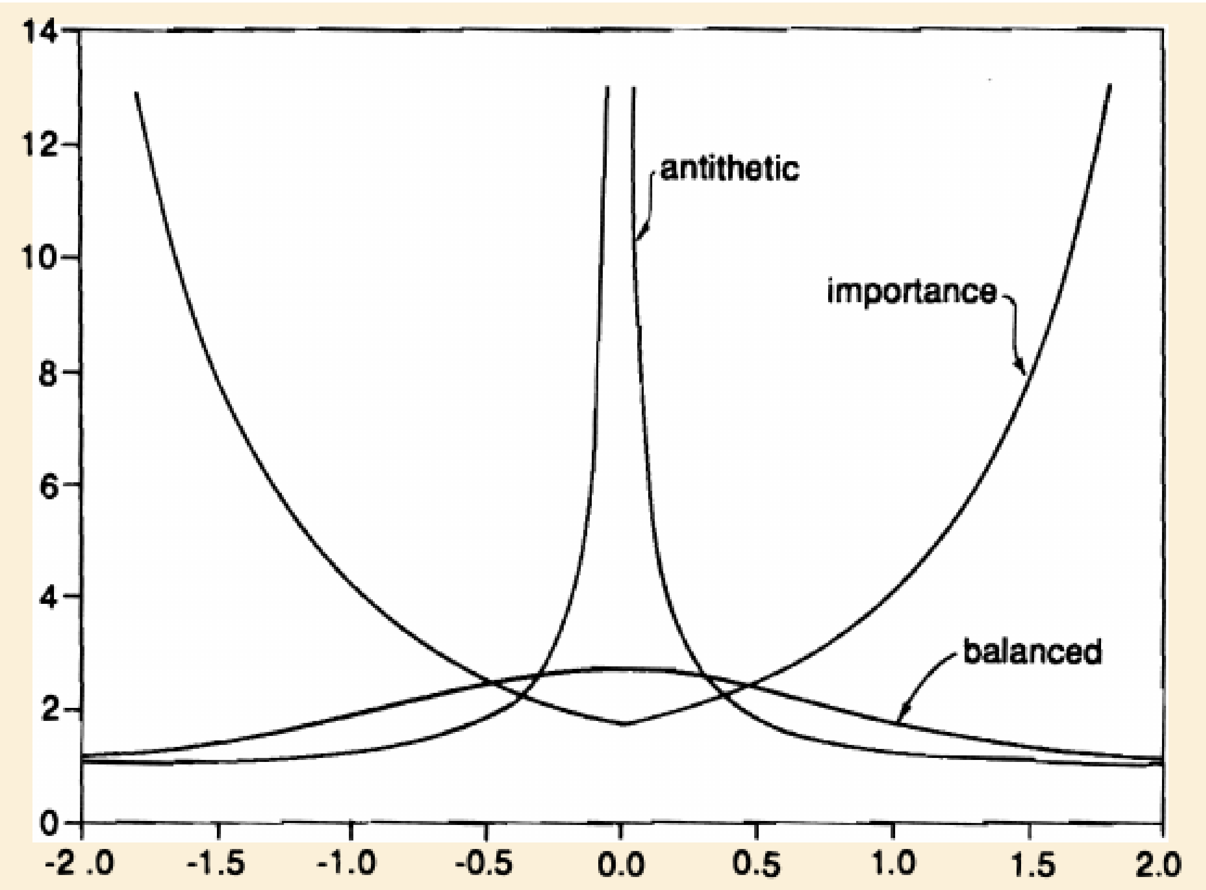

Figure A.1: Asymptotic Bootstrap efficiencies for CDF estimation of a normally distributed statistic

On the other hand, there are methods that belong to the **Bootstrap world**, *id est* they are not available in the general Monte Carlo setting. Namely,

- **Smoothing**[2]: the idea is to smooth the frequencies f_j in (1.15) through kernel smoothing to obtain a more accurate distribution of the statistic of interest. In other words, instead of sampling n_j, $\sum_j n_j = N$; $n_j \in \{0, ..., N\}$ times each statistical

[1]Whereas the book Hall (1992) does not specify how the efficiency curves were yielded, we infer it was a normal statistic based on the results presented in Davison and Hinkley (1997): indeed these efficiency values happen in the case of a linear statistic on a sample with an underlying normal distribution.

[2]Not to be confused with smooth Bootstrap

unit, now n_j are such that $\sum_j n_j = N \;; n_j \in [0, \dots, N)$, which yields a smoother and of course, depending the use case, a more accurate (in terms of the actual distribution F) Bootstrap distribution of the statistic of interest m.

- *A posteriori* **balance** corrections may also be made, so the sampling is made as in the original distribution (1.15), yet the results are modified based on the obtained empirical distribution for accuracy improvements, see Efron (1990) as well as Davison and Hinkley (1997). Some cases are named **centring methods** in Hall (1992).

- **Saddlepoint methods**. If at each Bootstrap sample we have to compute a statistic which is linear on the statistical units belonging to $\mathbf{X}^{*\mathbf{b}}$, denoted by $U^* = \sum_j^N a_j \mathbf{x}_j$, the so-called cumulant-generating function can be derived. Then, such function is used in the Saddlepoint equation, whose result is exploited to get a reliable approximation both for the CDF and the PDF of statistic U. Note that this procedure avoids completely the Monte Carlo estimate for the Bootstrap integral; for more details the reader is referred to Davison and Hinkley (1988) and Section 9.5 of Davison and Hinkley (1997).

- **Edgeworth expansions**, every statistic is a normal distribution plus a series of other terms, which is exploited in the estimation of the cumulants of a statistic as in Saddlepoint methods, seeHall (1992) and Section 9.5 of Davison and Hinkley (1997).

Naturally, both worlds can be merged. An example is Saddlepoint-Importance-Resampling (see Lee and Wong (2002)), although it is limited in the sense it cannot be applied directly to nonlinear functionals such as studententised statistics, which makes it in practice worthless to our seeting (as we will see in Chapter 3, the construction of nonparametric SCBs relies on nonlinear functionals).

The literature's variance reduction techniques are summarised in the following image:

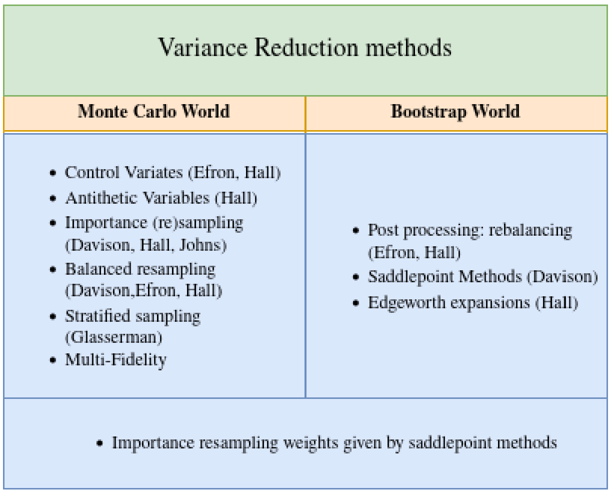

Figure A.2: Variance reduction: Monte Carlo world *versus* Bootstrap world

List of Figures

Acknowledgements

I would like to thank all the Professors at Politecnico di Milano that along these years have prepared me to be able to write this work, especially Professor Vantini, who has patiently listened to my ideas and guided me towards a fruitful work.
I would also like to thank Professor Cuevas and Professor Davison for kindly answering our questions about their papers written in the last century.
Agradezco especialmente a mi abuela Elena por haberme acompañado, entendiendo mi pasión por la epistemología en lenguaje matemático: la estadística.

Code

All the analyses are implemented using the R Programming Language (R Core Team (2023)), with packages *boot* (Canty and Ripley (2022)) and *roahd* (Ieva et al. (2019)). Codes are so far not publicly available, but the authors are at disposal for any clarification on the implementation details.