POLITECNICO MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Executive Summary of the Thesis

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

Laurea Magistrale in Mathematical Engineering - Ingegneria Matematica

Author: Alfredo Gimenez Zapiola

Advisor: Prof. Simone Vantini

Co-advisor:

Academic year: 2022-2023

1. Introduction

This thesis concerns a noticeably useful device in Nonparametric Statistics: the Bootstrap.
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 (MC) error, which is the scope of this thesis, has been of great interest for the contributors of the Bootstrap, and several variance reduction techniques have been researched. Amongst these methods, we focus on Importance Resampling, the adaptation of importance sampling to the Bootstrap world, which was proposed by Johns (1988), Do and Hall (1991) and Davison (1988), with a technique called Exponential Tilting to yield an importance distribution.
In addition, the statistics we are interested in are those that are utilised in the construction of the extension of confidence intervals to high-dimensional data, namely Simultaneous Confidence Bands (SCBs). In such context, Exponential Tilting fails in the sense it increases the MC error. We thus **propose two new algorithms**: **Loss Tilting** (LT) and **Contribution Tilted Mixture** (CTM) who effectively reduce variance for such statistics, even when the number of dimensions is high.
The thesis is organised as follows:

- Chapter 2 provides a brief explanation of the statistical devices used in this work, namely the Boostrap, Importance Resampling, the nonparametric delta method and high-dimensional data.
- In Chapter 3 we outline Importance Resampling with Exponential Tilting, as well as with our proposals **Loss Tilting** (LT) and **Contribution Tilted Mixture** (CTM). We show through a simulation study they perform better than Exponential Tilting and effectively reduce variance when estimating the required quantile of the statistics used in SCBs construction.
- In Chapter 4 we briefly overview the state of the art on SCBs and carry out an experiment to demonstrate the need for Importance Resampling in such setting.

2. Theoretical Background

2.1. The Bootstrap

It is a nonparametric method in Statistics, *i.e.* it makes minimal assumptions on the distribution of the data generating process to produce estimates. Its 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. Then, we make the following assumption:

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 $\hat{T}$ of a statistic T can be estimated (by $\hat{T}^*$) by sampling with replacement from $\hat{F}$, using the plug-in principle.*

In practice, such Bootstrap estimate is approximated by a Monte Carlo (MC) simulation, wherein at each iteration a re-sample from the given sample $\mathbf{X}$ is drawn. Hence we distinguish, for a statistic T defined by function t:

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:*

$$\mathbb{E}_F[T] = \mu \tag{1}$$

*with sample estimator function $\hat{t}$. 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*

$$\epsilon_B = \mu - \hat{\mu} \tag{2}$$

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

$$\epsilon_{MC} = \hat{\mu} - \hat{\hat{\mu}}_{MC} \tag{3}$$

$$\epsilon_{MC} = \int_{\Omega(\hat{F})} \hat{t}(\mathbf{X}^*)\, d\hat{F}(\mathbf{X}^*) \; - \frac{1}{B}\sum_{b=1}^{B} \hat{t}(\mathbf{X}^{*(\mathbf{b})}) \tag{4}$$

where B is the number of MC iterations and $\mathbf{X}^*$ the bth re-sample; and with $d\hat{F}(\mathbf{X}^*)$ being:

$$d\hat{F}(\mathbf{X}^*) = \prod_{j=1}^{N} p_j^{f_j^*} \tag{5}$$

; 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

in this thesis we focus on reducing the MC error.

2.2. Defintion of Importance Resampling

In such method, instead of sampling from distribution (5) with $p_j = \frac{1}{N}$, the idea is to (re)sample utilising some other resampling probabilities g_j provided by some importance distribution.
Id est, the integral in (4) becomes:

$$\int_{\Omega(\hat{F})} \hat{t}(\mathbf{X}^*)\, d\hat{F}(\mathbf{X}^*) = \int_{\Omega(\hat{F})} \hat{t}(\mathbf{X}^*) \frac{d\hat{F}(\mathbf{X}^*)}{dH(\mathbf{X}^*)}\, dH(\mathbf{X}^*) \tag{6}$$

where H is the so-called importance distribution, characterised by its re-sampling probabilities g_j which differ from N^{-1}, and is chosen such that the expected MC variance under Importance Resampling is diminished.
The right-hand integral is approximated with MC estimate $\hat{\hat{\mu}}_{IR}$ (where IR stands for Importance Resampling):

$$\hat{\hat{\mu}}_{IR} := \frac{1}{B}\sum_{b=1}^{B} \hat{m}(\mathbf{X}^{*\mathbf{b}}) w(\mathbf{X}^{*\mathbf{b}}) \tag{7}$$

with $w(\mathbf{X}^{*\mathbf{b}}) = \frac{d\hat{F}(\mathbf{X}^*)}{dH(\mathbf{X}^*)}$ being the likelihood ratio, necessary to ensure unbiasedness. Its variance is:

$$Var_H[\hat{\hat{\mu}}_{IR}] = B^{-1}\Big\{ \int_{\Omega(\hat{F})} (\hat{t})^2(\mathbf{X}^*) w^2(\mathbf{X}^*)\, dH(\mathbf{X}^*) - \mu^2 \Big\} \tag{8}$$

2.3. Nonparametric delta method

As it will be seen later, the importance distribution H can be provided by Exponential Tilting (Johns (1988), Davison (1988), Do and

Hall (1991)) or by our proposals LT and CTM. The first relies on the use of a statistical device known as the nonparametric delta method, which is the extension of the Taylor series expansion to statistical functions. Applying it in the Bootstrap section, we can have the linear approximation of a statistic:

$$\hat{t}_L(\mathbf{X}^*) := \hat{t}(\mathbf{X}) + n^{-1}\sum_{i=1}^{N} l_j^* \tag{9}$$

where l_j^* is the influence value (derivative of the statistical function w.r.t statistical unit) of the jth statistical unit of the Bootstrap sample.
The **nonparametric delta method result** states that for a smooth functional, the error of such approximation is asymptotically normal, which will not be the case with the statistics used in the simulation study.

2.4. Data with High Dimensions

We denote by N the sample size and p the number of random elements per statistical unit. 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 $\mathcal{S} = (a, b)$. When p is large, we deal with high-dimensional data.
In this thesis we deal with the multivariate extension of confidence intervals for large p, namely Simultaneous Confidence Bands (SCBs). We use Degras (2011) as a reference and propose two statistics pertinent to that task, namely the sup of the element-wise Student's t statistic

$$\sup_{s\in\mathcal{S}} \sqrt{N}\frac{\hat{\mu}(s)-\mu(s)}{\hat{\sigma}(s)} \tag{10}$$

with Bootstrap estimator $\sup_{s\in\mathcal{S}} \sqrt{N}\frac{\hat{\mu}^*(s)-\hat{\mu}(s)}{\hat{\sigma}^*(s)}$, and the *sup* of the element-wise so-called bias:

$$\sup_{s\in\mathcal{S}} \hat{\mu}(s)-\mu(s) \tag{11}$$

with Bootstrap estimator $\sup_{s\in\mathcal{S}} \hat{\mu}^*(s)-\hat{\mu}(s)$. The critical quantile $1-\alpha$ of these statistics is estimated such that a coverage of confidence level $1-\alpha$ is obtained. We remark that the *sup* operator is necessary for the coverage to be across all elements $s \in \mathcal{S}$ (*i.e.* simultaneous), in a similar fashion to Pini and Vantini (2017).

3. Importance Resampling

3.1. Importance Resampling for Bootstrap Quantile Estimation

As mentioned before, the estimands of interest are the critical quantiles of the statistics in Section 2.4. Thus, for a sample estimator $\hat{T}$ the quantity of interest, we are interested in quantile $\xi_{1-\alpha}$, which is the solution to:

$$\mathbb{P}_F(\hat{T} \le \xi_{1-\alpha}) = 1-\alpha \tag{12}$$

where F is the distribution of the data, its Bootstrap estimate is:

$$\mathbb{P}_{\hat{F}}(\hat{T}^* \le \hat{\xi}_{1-\alpha}) = 1-\alpha \tag{13}$$

Through an MC simulation of B iterations, an estimate $\hat{\hat{Q}}(y)$ of the ECDF (empirical cumulative distribution function) of $\hat{T}^*$ ($\hat{Q}(y)$) is obtained, namely:

$$\hat{\hat{Q}}(y_b) = \mathbb{P}_{\hat{F}}(\hat{T}^* \le y_b),\ b = 1, ..., B \tag{14}$$

whence we obtain MC estimate $\hat{\hat{\xi}}_{MC,1-\alpha}$ of $\hat{\xi}_{1-\alpha}$:

$$\hat{\hat{\xi}}_{MC,\alpha} = \inf\{y^{(b)} : \hat{\hat{Q}}_{MC}(y^{(b)}) \ge \alpha\ ,\ b = 1, \dots, B\} \tag{15}$$

(see Hall (1992)). In the case of Importance Resampling, the weights have to be taken into account too. Thus, the values of Bootstrap statistic at each iteration are ordered: $\hat{T}_1^* < \ ... \ < \hat{T}_B^*$ and their corresponding weights are $w_1^* < \ ... \ < w_B^*$. Then, if the order of the quantile is $< .5$, the MC estimate of the Bootstrap estimator for the desired quantile is $\hat{T}_M^*$ with M such that:

$$\frac{1}{B}\sum_{b=1}^{M} w_b^* \le \alpha < \frac{1}{B}\sum_{b=1}^{M+1} w_b^* \tag{16}$$

and when the order of the quantile is $> .5$, $\hat{T}_M^*$ such that

$$\frac{1}{B}\sum_{b=M}^{B} w_b^* \le 1-\alpha < \frac{1}{B}\sum_{b=M+1}^{B} w_b^* \tag{17}$$

where the idea is just to use the highest values in order to avoid potentially exploding weights (and thus increases in variance), since by tilting the importance distribution to the right, the smallest values of $\hat{T}^*$ have a high $d\hat{F}(\mathbf{X}^*)$ and low $dH(\mathbf{X}^*)$, which makes $w(\mathbf{X}^{*\mathbf{b}}) = \frac{d\hat{F}(\mathbf{X}^*)}{dH(\mathbf{X}^*)}$ skyrocket, and also the variance (8) (see Davison and Hinkley (1997) and Do and Hall (1991)). We furtherly remark that the farther the order of the quantile is from 0.5, the larger the variance of Ordinary MC is, which makes it an idoneous task for the application of Importance Reasampling (see Davison and Hinkley (1997) and Hall (1992)).

3.2. Choice of the Importance Distribution for Importance Resampling

3.2.1 Exponential Tilting

The three classic papers rely this method, namely Johns (1988), Davison (1988) and Do and Hall (1991), yet they differ slightly on this procedure. We 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 $\mathbf{X}$, exploiting Exponential Tilting:

1. We approximate the statistic $\hat{T}^*$ by its linearised version $\hat{T}^*_L$ (as seen on Equation (9)) which is an accurate approximation of itself.
2. Such statistic $\hat{T}^*_L$ follows approximately a normal distribution, which is the case asymptotically, as long as the nonparametric delta method reusult holds (which is true if the statistical function $\hat{t}$ of statistic $\hat{T}$ is smooth)
3. Exponential tilting is used to define the g_j in (5) **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.

which makes 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.*

Moving to Exponential Tilting, the idea is to assign:

$$g_j \propto (\lambda l_j),\ j = 1, ..., N \tag{18}$$

where l_j is the influence value of the jth statistical unit present in the sample, λ is a variable to tune such that the distribution is centered in the desired value. This is accomplished through a Newton solver for the problem:

$$\arg\min_{\lambda\in\mathbb{R}} \Big(\frac{\sum_{i=1}^N l_i \exp(\frac{\lambda l_i}{N})}{\sum_{i=1}^N \exp(\frac{\lambda l_i}{N})} - \theta_0 \Big)^2 \tag{19}$$

where θ_0 is the desired center of the importance distribution.

Remark 2 (Why exponential tilting?)**.** *There are two main reasons:*

- *It allows to set the Importance distribution (g_j in Equation (5)) 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 ($p_j = \frac{1}{N}$).*

Despite the influence values are derivatives, sometimes thay may be cumbersome to compute. Hence, they are estimated through regression. Moreover, the initial, rough estimate of the quantile of interest must be provided to know where to re-center the Importance distribution of the Bootstrap statistic $\hat{T}^*$. Therefore, a pilot MC run of B_1 iterations is performed. This yields:

- $\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_{B_1} \end{bmatrix}$ where $y_b = \hat{t}(\mathbf{X}^{*b})$, $b = 1, ..., B_1$, 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 $\mathbf{X}$) $\mathbf{Z} = \begin{bmatrix} f_1^{*1} & f_2^{*1} & \dots & f_N^{*1} \\ \vdots & \vdots & \ddots & \vdots \\ f_1^{*B_1} & f_2^{*B_1} & \dots & f_N^{*B_1} \end{bmatrix}$ where f_j^{*b}, $b = 1, ..., B_1$, $j = 1, ..., N$ is the frequency of the jth statistical unit on the bth Bootstrap sample. Note that $\sum_{j=1}^N f_j = N$, $\forall b \in \{1, ..., B_1\}$

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: $\hat{\mathbf{b}} = \begin{bmatrix} \hat{l}_1 \\ \vdots \\ \hat{l}_N \end{bmatrix}$. The pilot run gives also $\hat{\hat{\xi}}_{MC,\alpha}$ using (17).
We present Davison's procedure in the following Algorithm:

Algorithm 1 Importance resampling (Davison)

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 $\hat{l}_j^*$, $j \in \{1, \dots, N\}$ through regression.
Obtain an estimate of the α quantile $\hat{\hat{\xi}}_{B_1,\alpha}$ using T_b^*, $b = 1, \dots, B_1$
Calculate the probabilities of resampling each statistical unit g_j by solving Problem (19), yielding importance distribution H of shape of Equation 5 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^* as $\hat{t}(X^{*b})$
  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 αth quantile using Equation (16)
else
  Estimate the αth quantile using Equation (17)
end if

We remark that Assumption 2 is not respected when using statistics (11) and (10), due to the non-linearity of the *sup* operator. What is more, the same goes for Assumption 3, since it is not a smooth function and the nonparametric delta method result does not apply. This calls for the need of other algorithms to provide an importance distribution.

3.2.2 Loss Tilting (LT)

This is our first proposal. 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:

$$d := \theta_0 - \hat{t}(\mathbf{X}) \tag{20}$$

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 (20)
2. For **each** statistical unit j, compute the Loss function ℓ, to be provided by the statistician, of the difference between d and the (empirical if not derived analytically) influence value l_j ($\hat{l}_j$ when estimated empirically): $h_j = \ell(d - l_j)$
3. Since we want a probability distribution, we normalise h_i: $\tilde{h}_j = \frac{h_j}{\sum_{i=1}^N h_i}$
4. Since we want to give less probability the higher the loss is, we compute the complement of each $\tilde{h}_i$: $\tilde{h}_j^c = 1 - \tilde{h}_j$
5. And normalise them to get the importance distribution: $g_j = \frac{\tilde{h}_j^c}{\sum_{i=1}^N \tilde{h}_i^c}$, $j \in \{1, \dots, N\}$

3.2.3 Contribution Tilted Mixture (CTM)

This second proposal of ours is aimed at the scenario with high p and a *sup*-like operator. The intuition is the following: since both statistics (10) and (11) are the *sup* of an estimated element-wise statistic, then their value necessarily corresponds to the value of one of its components $\tilde{s}$, for *e.g.* $\sup_{s \in \mathcal{S}} \hat{\mu}^*(s) - \hat{\mu}(s) = \hat{\mu}^*(\tilde{s}) - \hat{\mu}(\tilde{s})$. Therefore, if such statistics take only the value of their component say $\tilde{s}$, we can "forget" about the fact it is a *sup* and apply (Exponential) tilting to the quantity for *e.g.* $\hat{\mu}^*(\tilde{s}) - \hat{\mu}(\tilde{s})$ in the case of statistic (11). In particular, if we chose Exponential Tilting, then the deviation from Assumption (2) would not be violated as badly as with the *sup* operator. In

practice, different components $s \in \mathcal{S}$ may be the ones whose value is the one taken by the statistic with the *sup*. We consider as a "contributor" each element $s \in \mathcal{S}$ of the multivariate 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* 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 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* statistics. The idea is to use as an Importance distribution a weighted Mixture of the individual weights $p_j^{(s)}$, $j \in \{1, \ldots, N\}$, $s \in \mathcal{S}$ obtain through the point-wise (Exponential) Tiltings to re-center at the desired quantile of order α.
Thus, we denoting with $\hat{m}^*(s)$, $s \in \mathcal{S}$ the Bootstrap estimate of either statistic (11) or (10), $T^* = \hat{t}(\mathbf{X}^*) = \sup_{s \in \mathcal{S}} \hat{m}^*(s)$, we define the estimate of the contribution of the element $s \in \mathcal{S}$ with:

$$\hat{c}_s := \frac{\sum_{b=1}^{B_1} \mathbb{1}\{\hat{T}_b^* = \hat{m}_b^*(s)\}}{B_1}, \; s \in \mathcal{S} \tag{21}$$

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:

$$\tilde{p}_j := \sum_{s \in \mathcal{S}} p_i^{(s)} \hat{c}_s \tag{22}$$

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 ($\sum_{s \in \mathcal{S}} \hat{c}_s = 1$; $0 \le c_s \le 1$, $\forall s \in \mathcal{S}$) of the element-wise importance distributions.
Therefore, **Contribution Tilting Mixture (CTM)** can be summarised in the following:

Algorithm 2 Contribution Tilting Mixture

1: Given the results of the Pilot run in Algorithm 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 (21)

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

to be inserted in Algorithm 1.

3.3. Simulation study

We made simulations by sampling with different N and p of the same data generating process. We used statistics (11) and (10), with loss function $\ell(x) = \frac{1}{2}x^2$ for LT and Exponential Tilting as the element-wise tilting in CTM.
The main result was that Exponential Tilting failed at $p > 1$, due to the violations of Assumptions 2 and 3, which we confirmed with different experiments. LT with a very generic loss function, although conservative, still managed to reduce variance w.r.t ordinary Monte Carlo. CTM, the best algorithm, consistently provided variance reductions relative to Ordinary resampling, even at the highest values of p, proving its affinity for data with high dimensions. We measured the efficiency as the variance of Ordinary Bootstrarp divided by the variance under Importance Resampling; using 1000 replications from the sample original sample.

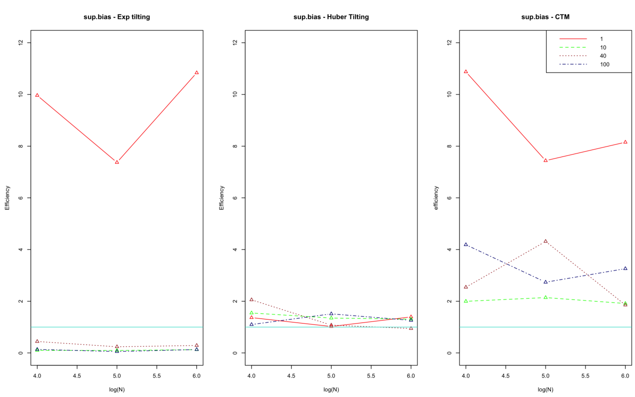

Figure 1: Efficiency curves for statistic (11)). Each curve represents a value of p

4. Simultaneous Confidence Bands for High Dimensional Data

4.1. Overview of the Boostrap for building SCBs

We declare Degras (2011) as the reference paper on this subject, and statistics similar to the one *ibi* present were explored in this thesis. State of the art works have not dealt so far (to our knowledge) with variance reduction of a Bootstrap estimate in high dimensional data, yet different variations of the Bootstrap have been proposed to improve the actual coverage level *i.e.* percentage of times the mean was within the SCBs, relative to the desired coverage level, especially at small sample sizes. Such variations are Smooth, Parametric, Block, Wild versions of the Bootstrap, among others.

4.2. Brief simulation study: the need for variance reduction

For different values of N and p of the same underlying data generating process, we performed MC simulations for the critical (.95)th quantile of Degras (2011)'s statistic, with different values of B, the total number of MC iterations. Whereas values of 500 (Cuevas and Fraiman (2004)) or 2500 (Telschow and Schwartzman (2022)) are commonly used nowadays, we noticed at least $B = 5000$ iterations are needed, which calls for the need of Importance Resampling in this setting.

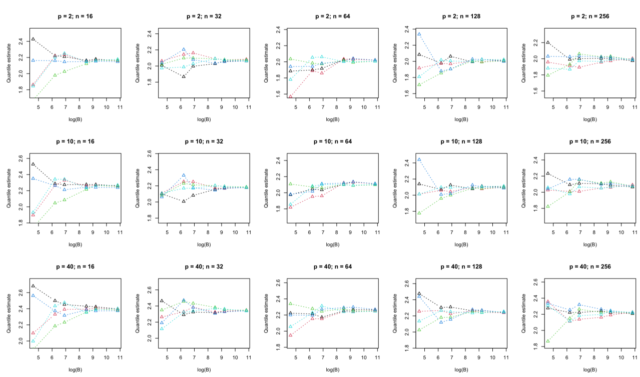

Figure 2: Curves of the MC quantile estimate of Degras (2011)'s statistic for four different random seeds, starting from the same sample at the different n and p.

5. Conclusions

In this thesis, we have applied Importance Resampling, a variance reduction techinque, for the estimation of quantiles of statistics used in the construction of SCBs. Exponential Tilting, the classical technique proposed by Johns (1988), Do and Hall (1991) and Davison (1988) fails in these circumstances, especially when the dimensions p are many. This opens the possibility for other ways to obtain an importance distribution. We have proposed Loss Tilting and Contribution Tilted Mixture, which we have proven to be effective for variance reduction through a simulation study. We have also performed a brief experiment in which we show that Importance Resampling is pertinent nowadays, as the MC error with the number of iterations used in state-of-the-art works is still signficant. We believe future research should be directed towards combining both variance reduction methods and implementing variations of the Bootstrap in order to hopefully reduce both MC and Bootstrap errors.

References

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. *Bootstrap Methods and their Application.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1997. doi: 10.1017/CBO9780511802843.

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

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

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`.

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`.

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

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.