REVSTAT – Statistical Journal
Volume 12, Number 2, June 2014, 169–197

ROBUST BOOTSTRAP: AN ALTERNATIVE TO BOOTSTRAPPING ROBUST ESTIMATORS

Authors: Conceição Amado
– Department of Mathematics and CEMAT,
Instituto Superior Técnico, Universidade de Lisboa, Portugal
conceicao.amado@tecnico.ulisboa.pt

Ana M. Bianco
– Facultad de Ciencias Exactas y Naturales,
Universidad de Buenos Aires and CONICET, Argentina
abianco@dm.uba.ar

Graciela Boente
– Facultad de Ciencias Exactas y Naturales,
Universidad de Buenos Aires and CONICET, Argentina
gboente@dm.uba.ar

Ana M. Pires
– Department of Mathematics and CEMAT,
Instituto Superior Técnico, Universidade de Lisboa, Portugal
apires@math.ist.utl.pt

Received: May 2012 Revised: April 2013 Accepted: July 2013

Abstract:

- There is a vast literature on robust estimators, but in some situations it is still not easy to make inferences, such as confidence regions and hypothesis testing. This is mainly due to the following facts. On one hand, in most situations, it is difficult to derive the exact distribution of the estimator. On the other one, even if its asymptotic behaviour is known, in many cases, the convergence to the limiting distribution may be rather slow, so bootstrap methods are preferable since they often give better small sample results. However, resampling methods have several disadvantages including the propagation of anomalous data all along the new samples. In this paper, we discuss the problems arising in the bootstrap when outlying observations are present. We argue that it is preferable to use a robust bootstrap rather than to bootstrap robust estimators and we discuss a robust bootstrap method, the Influence Function Bootstrap denoted IFB. We illustrate the performance of the IFB intervals in the univariate location case and in the logistic regression model. We derive some asymptotic properties of the IFB. Finally, we introduce a generalization of the Influence Function Bootstrap in order to improve the IFB behaviour.

Key-Words:

- *influence function; resampling methods; robust inference.*

AMS Subject Classification:

- 62F25, 62F40, 62G35.

1. INTRODUCTION

It is well known, that outliers or contamination have often an undesirable effect on statistical procedures. For this reason, robust methods provide more reliable inferences. However, in most situations, it is difficult to derive the exact distribution of robust estimators. On the other hand, even when its asymptotic distribution may be derived, the convergence to it may be rather slow. This suggests the use of bootstrap methods which are preferable since they can give even better small sample results. It is easy to understand that the outliers' effect increases when bootstrapping. Indeed, due to propagation effects, many bootstrap samples may have a higher contamination level than the original one. For that reason, the breakdown point for the whole procedure decreases and may become very small, even when based on an estimator with a high breakdown point. Besides, bootstrapping a robust estimator poses other challenges since the frequency of mathematical and numerical difficulties increases and also, the computation time grows up dramatically. These facts motivates the search of robust bootstrap procedures.

To allow for a small proportion of contamination on the data, we assume that the actual distribution of the data belongs to a contamination "neighbourhood" of a certain specified "central" parametric model, $P_{\mathbf{\Omega}}$ with $\mathbf{\Omega} = (\boldsymbol{\theta}, \boldsymbol{\tau})$, where $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^q$ stands for the parameter of interest while $\boldsymbol{\tau} \in \mathbb{R}^s$ denotes the nuisance parameters. In other words, we assume that $\mathbf{X}_1, ..., \mathbf{X}_n$ are a random sample with the same distribution as $\mathbf{X} \in \mathbb{R}^p$, where $\mathbf{X} \sim P_{\mathbf{\Omega}}$. The problem is to perform robust inference for the parameter $\boldsymbol{\theta}$, but with the snag that the sampling distribution of the statistics (pivot variable) is unknown.

As far as we know, the first work related to estimating the sampling distribution of robust estimators is due to Ghosh *et al.* (1984). This author showed that it is necessary to impose a tail condition on the underlying distribution, to ensure that the bootstrap variance estimate of the sample median converges. Athreya (1987) also showed that the bootstrap fails for heavy tailed distributions, while Shao (1990) again pointed out the non-robustness of the classical bootstrap. Shao (1992) proposed a "tail truncation" in order to obtain consistency of the bootstrap variance estimators, however it is not clear how to apply this in practice. Later on, Stromberg (1997) recommended either a robust estimate of the variance (of the bootstrap distribution) or the use of the deleted-d jackknife, as alternative bootstrap estimates for the robust estimators variability. Stromberg (1997) also studied a different resampling scheme (Limited Replacement Bootstrap), but concluded that it does not perform very well. Singh (1998) suggested a robust version of the bootstrap, for certain univariate L and M-estimators, by resampling from a winsorized sample instead of the original sample. This method is denoted, from now on, WB. Salibian–Barrera and Zamar (2002) introduced a robust bootstrap, denoted RB, based on a weighted representation of MM-regression and univariate location estimates. In Willems and Van Aelst (2005) and Salibian–Barrera *et al.* (2006), these methods were extended to other families of estimators. These proposals, being fast and stable, solve most of the problems pointed out above.

Amado and Pires (2004) suggested another method, also fast and stable, which consists on forming each bootstrap sample by resampling with different probabilities so that the potentially more harmful observations have smaller probabilities of selection. This method, denoted IFB, performs robust inference for a parameter based on the influence function (at the central model) of a classical point estimator. In this paper, we investigate the performance of the IFB procedure by simulation. To adapt for the sample size, a generalized procedure will also be considered.

The paper is organized as follows. In Section 2, we review the IFB procedure. In Section 3, we give different simulation results concerning the bootstrap intervals for univariate location and for logistic regression parameters. In Section 4, we present a generalization of the method and compare the results obtained with the new proposal and with the WB and RB procedures. Conclusions are given in Section 5, while technical results are relegated to the Appendix.

2. INFLUENCE FUNCTION BOOTSTRAP

The Influence Function Bootstrap is based on three main ideas: (1) resample less frequently highly influential observations (in the sense of Hampel's influence function); (2) at the same time, resample with equal probabilities the observations belonging to the "main structure"; (3) use a classical estimator on each "robustified" resample. Let us first consider a non robust estimator of $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}^{nr}$, based on the random sample with influence function $\mathrm{IF}\left(\mathbf{x};\hat{\boldsymbol{\theta}}^{nr},P_{\boldsymbol{\Omega}}\right)$ and its Standardized Influence Function, i.e.

$$\mathbb{SIF}(\mathbf{x};\hat{\boldsymbol{\theta}}^{nr},P_{\boldsymbol{\Omega}}) = \left[\mathrm{IF}\left(\mathbf{x};\hat{\boldsymbol{\theta}}^{nr},P_{\boldsymbol{\Omega}}\right)^{\mathrm{T}} V^{-1}_{(\hat{\boldsymbol{\theta}}^{nr},P_{\boldsymbol{\Omega}})}\mathrm{IF}\left(\mathbf{x};\hat{\boldsymbol{\theta}}^{nr},P_{\boldsymbol{\Omega}}\right)\right]^{1/2},$$

with $V_{(\hat{\boldsymbol{\theta}},P_{\boldsymbol{\Omega}})} = E_{P_{\boldsymbol{\Omega}}}\left[\mathrm{IF}\left(\mathbf{x};\hat{\boldsymbol{\theta}},P_{\boldsymbol{\Omega}}\right)\mathrm{IF}\left(\mathbf{x};\hat{\boldsymbol{\theta}},P_{\boldsymbol{\Omega}}\right)^{\mathrm{T}}\right]$ stands for the asymptotic variance of the estimator $\hat{\boldsymbol{\theta}}$. Assume that, as usual, $\mathbb{SIF}(\mathbf{x};\hat{\boldsymbol{\theta}}^{nr},P_{\boldsymbol{\Omega}})$ depends on $P_{\boldsymbol{\Omega}}$ only through the vector of unknown parameters, $\boldsymbol{\Omega}=(\boldsymbol{\theta},\boldsymbol{\tau})$, and that appropriate invariance properties hold. Now, define a Robust Standardized Empirical Influence Function by plugging into the $\mathbb{SIF}$ robust estimates, $\hat{\boldsymbol{\Omega}}^r = (\hat{\boldsymbol{\theta}}^r,\hat{\boldsymbol{\tau}}^r)$, of the unknown parameters and denote this function by $\mathbb{RESIF}(\mathbf{x};\hat{\boldsymbol{\theta}}^{nr},\hat{\boldsymbol{\Omega}}^r)$.

As a simple example on the computation of the $\mathbb{RESIF}(\mathbf{x};\hat{\boldsymbol{\theta}}^{nr},\hat{\boldsymbol{\Omega}}^r)$, consider multivariate location, $\boldsymbol{\theta}$, with a multivariate normal distribution as central model.

In this case, the nuisance parameter $\boldsymbol{\tau} = \boldsymbol{\Sigma}$ is the scatter matrix, so that $\hat{\boldsymbol{\Omega}}^r = (\hat{\boldsymbol{\theta}}^r, \hat{\boldsymbol{\Sigma}}^r)$ are robust estimators of the location and scatter parameters. Thus, it is easy to verify that, when $\hat{\boldsymbol{\theta}}^{nr} = \bar{\mathbf{x}}$, $\mathbb{RESIF}(\mathbf{x}; \hat{\boldsymbol{\theta}}^{nr}, \hat{\boldsymbol{\Omega}}^r)$ is the robust Mahalanobis distance currently used for outlier detection in multivariate data sets.

We now proceed to recall the IFB procedure introduced in Amado and Pires (2004). Given $c > 0$, let $0 \leq \eta(c, \cdot) \leq 1$ be a weight function verifying

$$\left.\frac{\partial\, \eta\,(c,t)}{\partial t}\right|_{t=c} = 0 \tag{2.1}$$

$$\lim_{t\to\infty} t^2\, \eta(c,t) = 0\,, \tag{2.2}$$

for each fixed value of the tuning constant c. As pointed out in Proposition 1 in Amado and Pires (2004), the condition (2.2) protects the bootstrap distribution from the harmful effect of outliers.

The Influence Function Bootstrap (IFB) procedure is described in the following steps:

a) Obtain $\mathbb{RESIF}_i = \mathbb{RESIF}(\mathbf{x}_i; \hat{\boldsymbol{\theta}}^{nr}, \hat{\boldsymbol{\Omega}}^r)$, $i = 1, 2, ..., n$.

b) Compute weights, w_i, according to

$$w_i = I_{[0,c]}\left(|\mathbb{RESIF}_i|\right) + \eta\left(c, |\mathbb{RESIF}_i|\right) \times I_{]c,+\infty]}\left(|\mathbb{RESIF}_i|\right)\,.$$

c) Compute the resampling probabilities $\mathbf{p} = (p_1, p_2, ..., p_n)$ as $p_i = w_i / \sum_{j=1}^n w_j$.

d) Resample with replacement according to $\mathbf{p}$ and for each robustified bootstrap sample compute the non-robust version of the estimate of interest.

Remark 2.1. The tuning constant c can be calibrated so as to obtain highly efficient procedures. Effectively, it is enough to determine or simulate the distribution of the $\mathbb{SIF}$ at the central parametric model and choose for c a very high percentile of this distribution.

Remark 2.2. A flexible family of functions from where the η function can be chosen is the kernel of the p.d.f. of the t-distribution and its limiting form, the normal distribution, that is,

$$\eta_{d,\gamma}(c,x) = \begin{cases} \left[1 + \dfrac{(x-c)^2}{\gamma\, d^2}\right]^{-\frac{\gamma+1}{2}} & 0 < \gamma < \infty \\[2ex] \exp\left[-\dfrac{(x-c)^2}{2d^2}\right] & \gamma = \infty \end{cases}\,.$$

More details about the method can be found in Amado and Pires (2004).

However, this method does not provide an explicit estimator to being bootstrapped. To identify this estimator, we will consider the case of a univariate parameter, to be more precise, the simplest case of an univariate location parameter with known scale.

Let us fix some notation which will be helpful in the sequel.

At the sample level we have: the sample denoted $(x_1, x_2, \cdots, x_n)$; the empirical distribution function, $P_n = \sum_{i=1}^n \delta_{x_i}/n$ with δ_x the point mass at x; the weights, $w_i = w(x_i; P_n)$, $1 \le i \le n$, defined in b); the weighted empirical distribution function denoted $P_{w_n,n} = \sum_{i=1}^n p_i \delta_{x_i}$, with $p_i = w_i / \sum_{i=1}^n w_i$ introduced in c).

Related to the above description, at the population level we have: an univariate random variable X; its probability density function, f with related distribution function P and a random variable denoted X_w with probability density function, f_w, called the weighted density function, with related weighted distribution function, P_w defined through

$$f_w(x) = \frac{w(x;P)f(x)}{\int w(x;P)f(x)dx} \quad \text{and} \quad P_w(x) = \int_{-\infty}^{x} f_w(u)du\,.$$

Besides, we can also define the mean, $\mu\,(P_w)$, and variance, $\sigma^2\,(P_w)$, of X_w. If $\lim_{x\to\infty} x^2 w(x;\cdot) < \infty$, then both $\mu\,(P_w)$ and $\sigma^2\,(P_w)$ are well defined and finite. Moreover, $\mu\,(P_w) \equiv \mu_w\,(P)$. The IFB procedure actually bootstraps the sample mean from $P_{w_n,n}$.

Concerning the asymptotic behaviour of the bootstrap proposal, Proposition 6.1 in the Appendix states that if $\hat{\boldsymbol{\Omega}}^r \xrightarrow{a.s.} \boldsymbol{\Omega}$ and $w(x;\cdot)$ is a Lipschitz continuous function of the unknown parameters, then $P_{w_n,n}(I_{(-\infty,x]}) \xrightarrow{a.s.} P_w(I_{(-\infty,x]})$, uniformly in x. This result entails easily that if $\lim_{x\to\infty} x^2 w(x;\cdot) = 0$, the variance of the weighted empirical distribution converges to $\sigma^2\,(P_w)$. We will now show that $\sigma^2\,(P_w)$ is related to the asymptotic variance of a robust estimator with score function $u\sqrt{w(u)}$.

By the Central Limit Theorem, $\sqrt{n}\,(\mu\,(P_{w_n,n}) - \mu\,(P_w)) \xrightarrow{d} N\left(0, \sigma^2\,(P_w)\right)$ (see Proposition 6.1b) in the Appendix for a related result concerning the Influence Function Bootstrap distribution). Thus, for large n, we have that

$$\text{(2.3)} \qquad \operatorname{Var}\left(\mu(P_{w_n,n})\right) \simeq \frac{\sigma^2\,(P_w)}{n} = \frac{\int \left(x-\mu(P_w)\right)^2 w(x)f(x)dx}{n \int w(x)f(x)dx}\,.$$

Let us consider a location M-functional with score function $\psi_M(u) = u\sqrt{w(u)}$, denoted by $\mu_{\sqrt{w}}(P)$ and its related estimator, $\mu_{\sqrt{w}}(P_n)$. The asymptotic variance of $\mu_{\sqrt{w}}(P_n)$, at the central model, is given by

$$\frac{\int (x-\mu)^2\, w(x) f(x)\, dx}{n\left[\mathbb{E}\psi'_M(X-\mu)\right]^2} = \frac{\int (x-\mu)^2\, w(x) f(x) dx}{n\left[\int \left(\sqrt{w(u)} + u\left(\sqrt{w(u)}\right)'\right) dP\right]^2}\ , \tag{2.4}$$

where h' stands for the derivative of the function $h : \mathbb{R} \to \mathbb{R}$. It is worth noting that the difference between expressions (2.3) and (2.4) is the denominator which will lead to the correction term to be introduced in Section 4. Almost equivalently, we may consider a weighted estimator (W-estimator) with a fixed number of steps and weights $\sqrt{w(u)}$. As we will see in Section 4 this relation give us a initial start point to perform a generalization of IFB method.

3. NUMERICAL RESULTS

In this section, we illustrate the IFB method in two models. We first consider the problem of computing confidence intervals for the location parameter under a location-scale model. Then, we focus on the problem of providing exact inferences for the regression parameter under a logistic regression model.

3.1. Univariate location model

We now present, as an example, the results of a simulation study concerning an univariate location parameter, μ, in the framework of a location-scale model. The aim is to compute confidence intervals for the parameter μ. In this simulation study we choose the nominal confidence level equal to 90%. We considered data sets $X_1, ..., X_n$, with sample size $n = 20$ and 50. The uncontaminated observations, which we label as C_0 in the Tables, are generated from $N(0,1)$. Three contamination situations are also studied

- C_1: Under this contamination, the data are generated from a $0.75N(0{,}1)+ 0.25N(0,9)$ distribution.
- C_2: This contamination corresponds to a high pointwise contamination, where 90% of the data have a standard normal distribution, $N(0,1)$, and 10% of the points are replaced by 10.
- C_3: The observations have the same distribution as Y/U where $Y \sim N(0,1)$ and $U \sim U(0,1)$, with Y and U independent.

The estimator is $\bar{X}$ and the intervals computed are: the classical t-intervals (CI_{ML}), the classical bootstrap with uniform weights (BCI_{ML}), the robust influence function bootstrap (BCI_{IF}) and the bootstrap intervals obtained by resampling from a winsorized sample (BCI_{WIN}). For the three bootstrap procedures the bootstrap percentile method was used for obtaining the confidence intervals. For BCI_{IF} intervals, we take $\mathbb{RESIF}(x) = |x - \text{median}(X_i)|/\text{MAD}(X_i)$ and $\eta(c,\cdot) = \eta_{d,\gamma}(c,\cdot)$ with $d = c = \sqrt{\chi^2_{1;0.99}}$ and $\gamma = \infty$. The number of bootstrap samples was $B = 2000$ in all cases and the number of simulation runs was 1000. The nominal level of the confidence intervals is 0.90.

Table 1 summarizes the results obtained by reporting coverage probability estimates, as well as mean and standard deviation of the lengths of the 1000 simulated confidence intervals.

Table 1: Confidence intervals for univariate location with confidence nominal level 0.90.

Cont. Scheme	Method	Coverage		$n=20$ Length		$n=50$ Length	
		$n=20$	$n=50$	Mean	Std. Dev.	Mean	Std. Dev.
C_0	CI_{ML}	0.899	0.901	0.7646	0.1252	0.4727	0.0484
	BCI_{ML}	0.874	0.895	0.7070	0.1165	0.4583	0.0475
	BCI_{IF}	0.871	0.892	0.7061	0.1155	0.4584	0.0478
	BCI_{WIN}	0.764	0.805	0.5570	0.1150	0.3764	0.0465
C_1	CI_{ML}	0.917	0.903	1.2799	0.3621	0.8073	0.1322
	BCI_{ML}	0.873	0.883	1.1811	0.3337	0.7813	0.1283
	BCI_{IF}	0.888	0.900	1.0770	0.2644	0.7107	0.1020
	BCI_{WIN}	0.766	0.765	0.7914	0.2203	0.7813	0.1283
C_2	CI_{ML}	0.820	0.048	2.4879	0.0655	1.5049	0.0244
	BCI_{ML}	0.598	0.015	2.2919	0.0767	1.4555	0.0375
	BCI_{IF}	0.864	0.890	0.8299	0.2096	0.5410	0.0928
	BCI_{WIN}	0.673	0.426	0.7158	0.1391	0.4937	0.0659
C_3	CI_{ML}	0.951	0.939	22.748	163.06	26.478	202.24
	BCI_{ML}	0.858	0.829	20.093	141.56	24.184	181.13
	BCI_{IF}	0.880	0.883	1.8251	0.4686	1.1783	0.1883
	BCI_{WIN}	0.698	0.678	1.8900	1.1634	1.1495	0.3450

From Table 1, we conclude that in the non-contaminated setting, C_0, the bootstrap intervals BCI_{ML} and BCI_{IF} have a behaviour similar to that of the classical t intervals, even when the latter are the optimal ones. The bootstrap intervals are shorter than the exact intervals CI_{ML}, but at the cost of losing some level. As expected, the optimal intervals CI_{ML} attain the largest coverage probabilities values. Besides, the intervals BCI_{WIN} achieve the smallest coverage probability for both $n = 20$ and 50, but they also have the smallest mean length. Under C_1, all the procedures keep a similar coverage value, even when their lengths are increased. On the other hand, under both C_2 and C_3, the coverage of the classical t and classical bootstrap intervals is completely spoiled for $n = 50$. For $n = 20$, the classical intervals almost keep their coverage, while classical bootstrap intervals lose coverage, under C_2. Under C_3, the coverage preservation is made at the expense of providing larger confidence intervals than those obtained for normal samples, leading to practically non-informative intervals. Under any contamination, for both sample sizes, the BCI_{WIN} intervals achieve smaller coverage probabilities than BCI_{IF} intervals, and far away from the nominal value. On the other hand, the coverage of BCI_{IF} intervals is very stable under all the contamination patterns keeping at same time the length under control.

These results show that, for the location model, the IFB procedure achieves its aim: it is a fast, robust and efficient inference method. It has also proven to work well in other situations including inference for the correlation coefficient (Amado and Pires, 2004) and selection of variables in linear discriminant analysis (Amado, 2003).

3.2. The logistic regression model

In order to check the behaviour of the proposal in a more complex model, we consider a special case of the generalized linear model (GLM), the logistic regression model. Under a logistic regression model, the observations $(Y_i, \mathbf{X}_i)$, $1 \leq i \leq n$, $\mathbf{X}_i \in \mathbb{R}^p$, are independent with the same distribution as $(Y, \mathbf{X}) \in \mathbb{R}^{p+1}$ such that the conditional distribution of $Y|\mathbf{X} = \mathbf{x}$ is $Bi(1, \mu(\mathbf{x}))$. The mean $\mu(\mathbf{x}) = \mathbb{E}(Y|\mathbf{X} = \mathbf{x})$ is modelled linearly through a known link function, that is, $\mu(\mathbf{x}) = H\left(\beta_0 + \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}\right)$ where, for the logistic model, $H(t) = 1/(1 + \exp(-t))$. Note that in this case, the nuisance parameter $\boldsymbol{\tau}$ is not present, so we will denote the distribution of the observations $P_{\boldsymbol{\theta}}$. We consider Influence Function Bootstrap intervals based on the weighted version of the Bianco and Yohai estimators (WBY) as introduced in Croux and Haesbroeck (2003). In order to guarantee existence of solution, Croux and Haesbroeck (2003) proposed to use the score function

$$\phi(t) = \begin{cases} t\exp(-\sqrt{d}) & \text{if } t \leq d \\ -2(1+\sqrt{t})\exp(-\sqrt{t}) + \left(2(1+\sqrt{d}) + d\right)\exp(-\sqrt{d}) & \text{otherwise.} \end{cases} \tag{3.1}$$

To define the robust bootstrap, we need to compute the $\mathbb{SIF}$. The influence function of the functional $\boldsymbol{\beta}_{\text{ML}}$ related to the maximum likelihood estimator $\widehat{\boldsymbol{\beta}}_{\text{ML}}$ is given by

$$\mathrm{IF}((y, \mathbf{x}), \boldsymbol{\beta}_{\text{ML}}, P_{\boldsymbol{\theta}}) = I(\boldsymbol{\beta})^{-1}(y - H(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}))\mathbf{x}\,, \tag{3.2}$$

where $P_{\boldsymbol{\beta}}(y = 1|\mathbf{x}) = H(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta})$ and $I(\boldsymbol{\beta}) = \mathbb{E}\left(H(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta})(1 - H(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}))\mathbf{x}\mathbf{x}^{\mathrm{T}}\right)$ stands for the information matrix. Therefore,

$$\mathbb{SIF}((y,\mathbf{x}),\boldsymbol{\beta}_{\text{ML}},P_{\boldsymbol{\beta}}) = \left\{(y - H(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}))^2\mathbf{x}^{\mathrm{T}}I(\boldsymbol{\beta})^{-1}\mathbf{x}\right\}^{\frac{1}{2}} .$$

Note that the distribution of the $\mathbb{SIF}$ is not independent of the parameter and so, the tuning constant c, as defined in Amado and Pires (2004), depends on $\boldsymbol{\beta}$. A data-driven procedure to compute c can be defined considering a preliminary robust estimator of $\boldsymbol{\beta}$. For the sake of simplicity, in our simulation process we have computed a unique value c from the true value $\boldsymbol{\beta}$.

To assess the performance of the bootstrapping influence robust intervals in the logistic model, first consider uncontaminated data sets following a model similar to that presented in Croux and Haesbroeck (2003). We select a high dimension regression parameter combined with a moderate sample size, that is $p = 11$ and $n = 100$. Since the influence function (3.2) depends on the regression parameter, we consider two different values for $\boldsymbol{\beta}$. To be more precise, we generate 1000 samples with covariates $\mathbf{X}_i = (1, \mathbf{Z}_i^{\mathrm{T}})^{\mathrm{T}}$ with $\mathbf{Z}_i \sim N_{10}(0, \mathbf{I})$ and binary responses Y_i such that $Y_i|\mathbf{X}_i = \mathbf{x} \sim Bi(1, H(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}))$. In the first case, $\boldsymbol{\beta} = (0, 0, ..., 0)^{\mathrm{T}}$, while in the second one, we choose $\boldsymbol{\beta} = (1, ..., 1)^{\mathrm{T}}/3\sqrt{11}$.

We calculate the classical maximum likelihood (ML) and the robust weighted estimators introduced in Croux and Haesbroeck (2003) and denoted $\widehat{\boldsymbol{\beta}}_{\mathrm{WBY}}$. The robust estimators were computed using the loss function (3.1) with tuning constant $d = 0.5$ and weights based on the robust Mahalanobis distance $d(\mathbf{z}, \widehat{\mu}_{\mathbf{z}}, \widehat{\boldsymbol{\Sigma}}_{\mathbf{z}})$, where $(\widehat{\mu}_{\mathbf{z}}, \widehat{\boldsymbol{\Sigma}}_{\mathbf{z}})$ stand for the Minimum Covariance Determinant estimators (MCD) of multivariate location and scatter of the explanatory variables $\mathbf{Z}_i$. We compute the asymptotic intervals based on the maximum likelihood estimators, ACI_{ML}, the related bootstrap intervals BCI_{ML}, the asymptotic intervals associated to the robust estimators ACI_{ROB} and the Influence Function Bootstrap intervals, BCI_{IF}, computed using the robust weights derived from the robust estimator $\widehat{\boldsymbol{\beta}}_{\mathrm{WBY}}$. In all cases, the number of bootstrap samples is $B = 2000$.

Tables 2 and 3 summarize the results in terms of coverage, mean length and standard deviation of the length of the obtained intervals, for both values of the regression parameter, under the central model. In Tables 2 and 3, we observe that the coverage of all the computed intervals is close to the nominal confidence level 0.90 for all the components of the regression parameter. The observed confidence level of the BCI_{IF} is close to the values obtained for the classical asymptotic intervals, while the classical bootstrap intervals BCI_{ML} achieve the lowest confidence levels. Besides, as expected, the asymptotic maximum likelihood intervals ACI_{ML} are the shortest ones, showing also the smallest standard deviations of the lengths. At the same time, we observe that BCI_{ML} intervals are the longest, while the BCI_{IF} have smaller standard deviation of the lengths than ACI_{ROB} and BCI_{ML} intervals. In fact, we confirm that the performance of the BCI_{IF} intervals is the same regardless the value of the regression parameter.

Table 2: Coverage, mean length and standard deviation of the length for the non-contaminated samples from a logistic model with $\boldsymbol{\beta} = (0, ..., 0)^{\mathrm{T}}$, $p = 11$. Nominal level 0.90.

Comp.	ACI_{ML}	ACI_{ROB}	BCI_{ML}	BCI_{IF}
Coverage				
β_0	0.876	0.907	0.850	0.887
β_1	0.885	0.904	0.842	0.892
β_2	0.884	0.916	0.841	0.888
β_3	0.898	0.908	0.868	0.894
β_4	0.896	0.916	0.852	0.893
β_5	0.880	0.914	0.877	0.887
β_6	0.890	0.917	0.861	0.890
β_7	0.867	0.888	0.851	0.866
β_8	0.875	0.900	0.862	0.883
β_9	0.894	0.897	0.845	0.891
β_{10}	0.868	0.896	0.842	0.867
Mean Length				
β_0	0.743	0.848	0.977	0.929
β_1	0.755	0.898	1.014	0.965
β_2	0.758	0.904	1.016	0.967
β_3	0.757	0.908	1.012	0.968
β_4	0.758	0.907	1.018	0.967
β_5	0.755	0.906	1.018	0.966
β_6	0.756	0.903	1.015	0.966
β_7	0.757	0.900	1.011	0.965
β_8	0.755	0.909	1.011	0.966
β_9	0.756	0.904	1.018	0.968
β_{10}	0.755	0.907	1.015	0.966
Standard Deviation Length				
β_0	0.030	0.091	0.082	0.064
β_1	0.064	0.158	0.137	0.106
β_2	0.068	0.164	0.136	0.112
β_3	0.064	0.156	0.128	0.104
β_4	0.063	0.172	0.133	0.110
β_5	0.064	0.163	0.128	0.108
β_6	0.063	0.172	0.130	0.106
β_7	0.064	0.161	0.129	0.105
β_8	0.064	0.165	0.131	0.110
β_9	0.065	0.162	0.132	0.118
β_{10}	0.065	0.172	0.137	0.127

Table 3: Coverage, mean length and standard deviation of the length for the non-contaminated samples from a logistic model with $\boldsymbol{\beta} = (1, ..., 1)^{\mathrm{T}}/3\sqrt{11}$, $p = 11$. Nominal level 0.90.

Comp.	ACI_{ML}	ACI_{ROB}	BCI_{ML}	BCI_{IF}
Coverage				
β_0	0.885	0.912	0.851	0.895
β_1	0.872	0.900	0.833	0.879
β_2	0.873	0.900	0.857	0.887
β_3	0.882	0.900	0.862	0.892
β_4	0.875	0.897	0.847	0.882
β_5	0.892	0.918	0.871	0.903
β_6	0.895	0.925	0.843	0.901
β_7	0.878	0.881	0.834	0.875
β_8	0.875	0.913	0.829	0.880
β_9	0.888	0.907	0.855	0.894
β_{10}	0.876	0.907	0.853	0.888
Mean Length				
β_0	0.754	0.874	1.006	0.947
β_1	0.772	0.937	1.054	0.988
β_2	0.770	0.930	1.051	0.986
β_3	0.765	0.922	1.039	0.978
β_4	0.767	0.923	1.044	0.979
β_5	0.766	0.923	1.041	0.977
β_6	0.771	0.928	1.050	0.985
β_7	0.774	0.940	1.051	0.991
β_8	0.767	0.921	1.043	0.980
β_9	0.769	0.927	1.046	0.980
β_{10}	0.768	0.935	1.044	0.985
Standard Deviation Length				
β_0	0.034	0.110	0.103	0.069
β_1	0.069	0.179	0.164	0.117
β_2	0.067	0.180	0.153	0.117
β_3	0.067	0.185	0.154	0.113
β_4	0.067	0.182	0.148	0.113
β_5	0.066	0.181	0.152	0.114
β_6	0.068	0.185	0.149	0.114
β_7	0.068	0.188	0.154	0.116
β_8	0.067	0.175	0.144	0.112
β_9	0.069	0.180	0.154	0.114
β_{10}	0.068	0.190	0.150	0.122

In the second part of this numerical study, we evaluate the performance of the Influence Function Bootstrap intervals under non-contaminated and contaminated samples with $p = 3$. We generate 1000 samples of size $n = 100$ where $\mathbf{X} = (1, \mathbf{Z}^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^3$, corresponding to an intercept and two covariates. The explanatory variables $\mathbf{Z}_i$ are i.i.d. and such that $\mathbf{Z}_i \sim N_2(0, \mathbf{I}_2)$, while the response variables Y_i follow a logistic model $Y_i|\mathbf{X}_i = \mathbf{x} \sim Bi(1, H(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}))$ with $\boldsymbol{\beta}^{\mathrm{T}} = (0, 2, 2)$. We identify this case as the non-contaminated situation C_0 and we also consider the following contamination schemes:

- C_1: 5 misclassified observations are introduced on a hyperplane parallel to the true discriminating hyperplane $\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}$ with a shift equal to $1.5 \times \sqrt{2}$ and with the first covariate x_1 around 5.
- C_2 : similar to scheme of C_1, but with a shift equal to $5 \times \sqrt{2}$.

We computed the same intervals as for $p = 11$. In the bootstrapping procedures, the number of resamples is $B = 2000$ and the simulated samples where we detect possible non-overlapping leading to non-convergence were replaced by new ones. Table 4 sums up the simulation results. Under the central model,

Table 4: Coverage, mean length and standard deviation of the length, for non-contaminated and contaminated samples from a logistic model with $\boldsymbol{\beta} = (0, 2, 2)^{\mathrm{T}}$. Nominal level 0.90.

Method	Coverage			Mean Length			Std. Dev. Length		
	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2
C_0									
ACI_{ML}	0.902	0.890	0.901	1.010	1.624	1.629	0.101	0.357	0.354
ACI_{ROB}	0.929	0.930	0.933	1.072	1.810	1.827	0.157	0.566	0.584
BCI_{ML}	0.846	0.778	0.797	1.152	2.038	2.030	0.207	0.845	0.824
BCI_{IF}	0.908	0.827	0.860	1.124	1.924	1.924	0.158	0.582	0.579
C_1									
ACI_{ML}	0.714	0.088	0.859	0.903	0.882	1.352	0.084	0.153	0.285
ACI_{ROB}	0.882	0.860	0.844	1.003	1.647	1.632	0.134	0.504	0.506
BCI_{ML}	0.819	0.280	0.716	1.087	1.050	1.951	0.186	0.723	1.361
BCI_{IF}	0.767	0.513	0.861	0.976	1.377	1.633	0.132	0.603	0.468
C_2									
ACI_{ML}	0.629	0.000	0.001	0.708	0.547	0.749	0.027	0.030	0.070
ACI_{ROB}	0.881	0.860	0.843	1.004	1.647	1.634	0.137	0.500	0.510
BCI_{ML}	0.689	0.000	0.007	0.725	0.553	0.779	0.044	0.060	0.100
BCI_{IF}	0.820	0.824	0.798	0.982	1.801	1.692	0.154	0.410	0.480

we observe a similar behaviour to that described for $p = 11$, that is the coverage of the BCI_{ROB} is close to the values obtained with ACI_{ML}. We can observe the serious effect of the contamination on the classical asymptotic and bootstrap intervals

ACI_{ML} and BCI_{ML}. Indeed, both types of intervals are completely non-informative for β_1 under both contamination schemes, since the coverage is less than 0.30 under C_1 and 0 under C_2. On the other hand, under C_1, the intervals BCI_{IF} achieve lower coverages than the asymptotic intervals ACI_{ROB} for components β_0 and β_1, but they are also shorter than the former. Besides, the intervals BCI_{IF} obtained for β_2 have higher coverage with a similar length to that of ACI_{ROB} and the standard deviation of their length is smaller than that of the asymptotic robust intervals based on $\widehat{\boldsymbol{\beta}}_{\text{WBY}}$. Under C_2, the comparison of the BCI_{IF} intervals and the asymptotic robust ones, ACI_{ROB}, is similar to that described for C_1, but in this case the coverage values of the intervals obtained for β_0 and β_1 are closer. Unlike the previous case, for β_2 the BCI_{IF} intervals achieve a lower coverage than ACI_{ROB} and BCI_{IF} intervals for β_1 and β_2 are larger than the robust asymptotic ones. Moreover, the standard deviations of the length of the BCI_{IF} intervals for β_2 and β_3 is smaller than those of the ACI_{ROB} ones. We conclude that BCI_{IF} intervals are comparable to the asymptotic intervals based on the robust estimator, and this is more evident under C_0 and under the case of the more severe contamination C_3 for the chosen value of the parameter.

4. GENERALIZATION OF THE INFLUENCE FUNCTION BOOTSTRAP

As shown in the simulation study, a weakness of the IFB procedure is the choice of the tuning constant. Effectively, in order to avoid undercoverage of the confidence intervals (or underestimation of the variance), the constant c needs to be a very high percentile of the $\mathbb{SIF}$ which restricts the degree of robustness of the proposal.

In order to determine the needed correction, recall the discussion given in Section 2 for an univariate location parameter with known scale, regarding the M-estimator related to the bootstrap procedure. In fact, (2.3) and (2.4) give the expressions for the asymptotic variance of the mean of the bootstrap distribution and of an M-estimator with score function $\psi_M(u) = u\sqrt{w(u)}$. Now, assuming that $\mu_{\sqrt{w}}(P) \approx \mu_w(P)$, which is true if P is approximately symmetric, the bootstrap distribution of $\mu(P_{w_n,n})$ can be corrected, in order to be closer to the bootstrap distribution of $\mu_{\sqrt{w}}(P_n)$, by sampling n_{new} observations from $P_{w_n,n}$, with

$$n_{new} = \frac{\left[\int \left(\sqrt{w(u)} + u\left(\sqrt{w(u)}\right)'\right) dP\right]^2}{\int w(u)dP} \times n\,, \tag{4.1}$$

where h' stands for the derivative of the function $h : \mathbb{R} \to \mathbb{R}$. The corrected sample size n_{new} can be estimated by

$$\hat{n}_{new} = \frac{\left[\sum_{i=1}^{n} \sqrt{w(u_i)} + \sum_{i=1}^{n} u_i \left(\sqrt{w(u_i)}\right)'\right]^2}{\sum_{i=1}^{n} w(u_i)} ,$$

where u_i denotes the current standardized residuals. Another possible correction is to sample n observations from $P_{w_n,n}$ and to multiply the centred bootstrap distribution by $\sqrt{n/\hat{n}_{new}}$. Incidentally, we note that this correction is very similar to one of the corrections needed by the robust bootstrap of Salibian–Barrera (2000) denoted *RB*. The Influence Function Bootstrap with correction is denoted by IFB*.

In order to illustrate the generalization of the IFB to another univariate example, we deal now with the correlation coefficient. Let $\mathbf{X}^{\mathrm{T}} = (X_1, X_2)$ be a random vector following a bivariate distribution P with mean $\boldsymbol{\mu}$ and covariance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} ,$$

with $\sigma_{ii} = \mathrm{Var}(X_i)$ and $\sigma_{ij} = \mathrm{Cov}(X_i, X_j)$, for $i \neq j$ and $i, j = 1, 2$. The correlation coefficient between X_1 and X_2 is given by $\rho = \mathrm{corr}(X_1, X_2) = \sigma_{12}/\sqrt{\sigma_1 \sigma_2}$.

Assume that we have a a random sample $(x_{11}, x_{12}), (x_{21}, x_{22}), \cdots, (x_{1n}, x_{2n})$ with distribution P and let $\rho(P_n)$ be the Pearson sample correlation coefficient. Amado and Pires (2004) give the $\mathbb{SIF}$, the robust empirical function $\mathbb{RESIF}$ and the weights w_i for $\rho(P_n)$. To apply the generalization and obtain the IFB* corresponding to ρ, we follow analogous calculus to those derived for the univariate location parameter. In order to get IFB*, we resample in each bootstrap step n_{new} observations, where n_{new} is given in (4.1). Note that we are dealing with the distribution of $\rho_{\sqrt{w}}(P_n) - \rho_{\sqrt{w}}(P)$, where $\rho_{\sqrt{w}}(P_n)$ is the estimator that links original and weighted models given by

$$\rho_{\sqrt{w}}(P_n) = \frac{\sum_{i=1}^{n} w_i (x_{i1} - \hat{\mu}_1)(x_{i2} - \hat{\mu}_2)}{\sqrt{\sum_{i=1}^{n} w_i (x_{i1} - \hat{\mu}_1)^2 \sum_{i=1}^{n} w_i (x_{i2} - \hat{\mu}_2)^2}} ,$$

with $\hat{\mu}_j = \left(\sum_{i=1}^{n} w_i x_{ij}\right)\left(\sum_{i=1}^{n} w_i\right)^{-1}$, $j = 1, 2$.

This generalization of the IFB can be extended to more complex models with multivariate parameters such as generalized linear models, but this topic will be the subject of future work.

In the next sections, we make a comparison between the IFB* distribution and the distribution of the W-estimator for an univariate location model. We also evaluate the performance of bootstrap confidence intervals for the univariate location parameter and for the correlation coefficient.

4.1. The IFB* distribution for the univariate location case

To study the performance of the IFB* distribution, we generate 500 random samples $X_1, ..., X_n$ of size $n = 20$ and 50. In the non-contaminated situation, labelled C_0 in the Tables, the observations have a $N(0, 1)$ distribution. The contaminated model, denoted C_1, is such that $X_i \sim 0.9N(0, 1) + 0.1N(10, 0.1)$ which corresponds to a contaminated pattern where 10% of the observations have a large mean with a small variance. The compared methods are IFB and IFB* with $\mathbb{RESIF}(x) = |x - \text{median}(X_i)|/\text{MAD}(X_i)$ and $\eta(c, \cdot) = \eta_{d,\gamma}(c, \cdot)$ with $d = c = 1.5$ and $\gamma = \infty$. The number of bootstrap samples is $B = 5000$.

To compare the IFB* distribution with the distribution of the W-estimator we need a reliable estimate of the "true" distribution. For that purpose, an independent prior simulation was run as follows: 5000 samples were generated from the considered distributions and the empirical percentiles (2.5, 5, 10, 25, 50, 75, 90, 95, 97.5) were determined. The selected percentiles were used in a study to evaluate bootstrap distributions by Srivastava and Chan (1989). The previous step was repeated 100 times. The final estimate of each percentile is the median of the corresponding 100 observations.

Let P^* stand for the bootstrap distribution. Four bootstrap distributions were actually considered

- The IFB distribution (without correction), centered at $\mu_w(P_n)$,

$$R^{(\text{IF})}_{\text{BOOT}}(x) = \frac{1}{B}\sum_{b=1}^{B} I\left\{\mu(P^*_{w_n,n}) - \mu_w(P_n) \leq x\right\} ,$$

- The IFB* distribution (with correction), centered at $\mu_{\sqrt{w}}(P_n)$,

$$R^{(1)}_{\text{BOOT}}(x) = \frac{1}{B}\sum_{b=1}^{B} I\left\{\mu(P^*_{w_n,\hat{n}_{new}}) - \mu_{\sqrt{w}}(P_n) \leq x\right\} ,$$

- The IFB* distribution (with correction), centered at $\mu_w(P_n)$,

$$R^{(2)}_{\text{BOOT}}(x) = \frac{1}{B}\sum_{b=1}^{B} I\left\{\mu(P^*_{w_n,\hat{n}_{new}}) - \mu_w(P_n) \leq x\right\} ,$$

- The IFB* distribution with two corrections, the previous one and an empirical correction for asymmetry, centered at $\mu_w(P_n)$,

$$R^{(3)}_{\text{BOOT}}(x) = \frac{1}{B}\sum_{b=1}^{B} I\left\{\left(\mu(P^*_{w_n,\hat{n}_{new}}) - \mu^*_{\hat{n}_{new}}\right) \times f_c + \mu^*_{\hat{n}_{new}} - \mu_w(P_n) \leq x\right\} ,$$

with $f_c = (V_{\text{BOOT}} + 25\,D^2/n)/V_{\text{BOOT}}$, $D = \mu_w(P_n) - \mu_{\sqrt{w}}(P_n)$ and V_{BOOT} equals the bootstrap estimator of mean variance from the weighted sample.

For a given percentile, p, let $\widehat{P^{-1}_{\mu_{\sqrt{w}}}(p)}$ be the estimated percentile of the distribution of $\mu_{\sqrt{w}}(P_n)$ in the previous simulation study. For each of the 500 replications and for each p, we computed $R^{(m)}_{\text{BOOT}}\left(\widehat{P^{-1}_{\mu_{\sqrt{w}}}(p)}\right)$, with $m = \text{IF}, 1, 2, 3$. Note that if the bootstrap distribution is close to the distribution of $\mu_{\sqrt{w}}$, then $R^{(m)}_{\text{BOOT}}\left(\widehat{P^{-1}_{\mu_{\sqrt{w}}}(p)}\right)$ must be close to p. Table 5 reports the mean (ME_p) over the 500 replications, for each p. To assess a the global performance a Kolmogorov–Smirnov type statistic is also given in the last column of Table 5 and denoted $KS = \max_p |ME_p - p|$. The results for other distributions, including the Cauchy and the log-normal distribution are available in Amado (2003).

Table 5: Comparison of different bootstrap distributions with the "true" distribution of the weighted estimator for the univariate location model when $n=20$ and 50.

$C_0,\ n=20$										
p	2.5	5	10	25	50	75	90	95	97.5	KS
$R^{(\text{IF})}_{\text{BOOT}}$	1.97	3.96	8.23	22.63	49.80	77.09	91.71	96.06	98.04	2.37
$R^{(1)}_{\text{BOOT}}$	2.50	4.75	9.34	23.90	49.98	76.07	90.76	95.39	97.60	1.10
$R^{(2)}_{\text{BOOT}}$	2.46	4.68	9.26	23.77	49.93	76.15	90.80	95.41	97.60	1.23
$R^{(3)}_{\text{BOOT}}$	2.49	4.73	9.32	23.84	49.93	76.08	90.73	95.37	97.57	1.16

$C_1,\ n=20$										
p	2.5	5	10	25	50	75	90	95	97.5	KS
$R^{(\text{IF})}_{\text{BOOT}}$	3.29	5.17	8.97	22.26	49.33	78.92	94.63	98.56	99.80	4.63
$R^{(1)}_{\text{BOOT}}$	4.23	6.42	10.62	23.83	48.83	76.51	92.68	97.46	99.36	2.68
$R^{(2)}_{\text{BOOT}}$	2.53	4.31	7.95	20.38	45.26	73.86	91.16	96.58	98.93	4.74
$R^{(3)}_{\text{BOOT}}$	3.27	5.16	8.88	21.27	45.79	73.91	91.00	96.40	98.79	4.21

$C_0,\ n=50$										
p	2.5	5	10	25	50	75	90	95	97.5	KS
$R^{(\text{IF})}_{\text{BOOT}}$	2.21	4.44	9.10	23.88	49.95	76.01	90.89	95.58	97.83	1.12
$R^{(1)}_{\text{BOOT}}$	2.44	4.83	9.66	24.48	50.03	75.45	90.33	95.16	97.57	0.52
$R^{(2)}_{\text{BOOT}}$	2.44	4.82	9.64	24.46	50.05	75.52	90.38	95.21	97.60	0.54
$R^{(3)}_{\text{BOOT}}$	2.45	4.83	9.65	24.48	50.05	75.50	90.37	95.19	97.59	0.52

$C_1,\ n=50$										
p	2.5	5	10	25	50	75	90	95	97.5	KS
$R^{(\text{IF})}_{\text{BOOT}}$	2.90	5.14	9.74	24.98	52.44	79.63	93.76	97.58	99.12	4.63
$R^{(1)}_{\text{BOOT}}$	3.37	6.07	11.20	26.52	52.18	77.65	92.03	96.48	98.54	2.65
$R^{(2)}_{\text{BOOT}}$	2.35	4.63	9.18	23.58	48.98	75.31	90.78	95.76	98.13	1.42
$R^{(3)}_{\text{BOOT}}$	2.55	4.88	9.47	23.87	49.14	75.30	90.70	95.68	98.07	1.13

The main conclusions from the overall experiment are: (1) the accuracy of the bootstrap approximation increases with n, but it can be quite good even for $n = 20$; (2) the results are better for symmetric distributions; (3) $R^{(3)}_{\text{BOOT}}$ is usually the best approximation, especially for asymmetric distributions. This study was also performed for another contamination patterns and larger sample sizes ($n = 100$) leading to analogous conclusions.

4.1.1. Confidence intervals for univariate location based on IFB*

For this study, we consider the simulation design of Salibian–Barrera (2000, Section 3.6.2). We generate i.i.d. observations $X_1, ..., X_n$ with $n = 20, 30, 50$ such that $X_i \sim (1-\varepsilon)N(0,1) + \varepsilon N(-7, 0.1)$, with $\varepsilon = 0, 0.1, 0.2, 0.3$. The method chosen is the basic percentile method with IFB*, where $\mathbb{RESIF}(x) = |x - \hat{\mu}_{LTS}|/\hat{\sigma}_{LTS}$ with $\hat{\mu}_{LTS}$ and $\hat{\sigma}^2_{LTS}$ the least trimmed mean and variance estimators. We also choose $c = 1.5$ and 2 and denote the procedure IFB*(1.5) and IFB*(2), respectively. The number of bootstrap samples is $B = 5000$ and the number of simulation runs is 1000.

Table 6 reports the estimated coverage and the length of 95% confidence intervals. The results under the heading "Censored simulation" are obtained after excluding from the simulation (not from the bootstrap) samples with more than 50% contamination, since there is no equivariant method able to deal with this situation.

Table 6: Estimated coverage and length, between brackets, of nominal 95% confidence intervals for a univariate location model from contaminated distribution $(1-\varepsilon)N(0,1) + \varepsilon N(-7, 0.1)$. Results in **boldface** indicate significant difference to target.

n	ε	IFB*(2)	IFB*(1.5)	Censored simulation	
20	0.0	**0.922** (0.83)	**0.915** (0.85)	—	—
	0.1	0.944 (1.14)	**0.923** (0.95)	—	—
	0.2	0.955 (1.54)	**0.927** (1.13)	0.958 (1.58)	0.935 (1.12)
	0.3	**0.920** (2.08)	**0.890** (1.36)	0.954 (2.08)	0.938 (1.33)
30	0.0	0.939 (0.70)	0.930 (0.70)	—	—
	0.1	0.964 (0.93)	0.942 (0.79)	—	—
	0.2	0.959 (1.29)	0.934 (0.90)	—	—
	0.3	0.961 (1.78)	0.933 (1.08)	**0.975** (1.78)	0.951 (1.06)
50	0.0	0.941 (0.55)	0.943 (0.55)	—	—
	0.1	0.956 (0.70)	0.954 (0.60)	—	—
	0.2	**0.974** (0.98)	0.952 (0.71)	—	—
	0.3	**0.978** (1.41)	0.961 (0.83)	—	—

Comparing the obtained results with those reported in Salibian–Barrera (2000, page 129) for the studentized robust (SRB) and weighted (WB) Bootstrap (with the same simulation conditions, but 3000 runs) we conclude that: (1) the coverage of IFB* intervals is similar to the coverage of WB intervals in all cases, and worse than that of SRB intervals only when $n = 20$; (2) under contamination, the length of the intervals follows the following order, IFB*(1.5) $<$ WB $<$ IFB*(2) $<$ SRB.

4.2. The correlation coefficient

As in Section 4.1, we now consider the distribution of IFB* for the case of the correlation coefficient. Samples with $n = 20$ observations were generated from a non-contaminated and a contaminated model, labelled C_0 and C_1, respectively. Under C_0, $\mathbf{X}_i$ are i.i.d. $\mathbf{X}_i \sim N(0, \mathbf{\Sigma})$, where

$$\mathbf{\Sigma} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix} .$$

Under C_1, the observations are still independent and such that $\mathbf{X}_i \sim N_2\left(0, \mathbf{\Sigma}\right)$ for $1 \le i \le n - [\varepsilon n]$ while $\mathbf{X}_i \sim \delta_{\mathbf{x}}$ when $n - [\varepsilon n] + 1 \le i \le n$. We choose $\varepsilon = 0.1$ and $\mathbf{x} = (-5, 5)^{\mathrm{T}}$.

As in Section 4.1, we consider four bootstrap distributions IFB (taking $c = 5$) defined as

- $R^{(\mathrm{IF})}_{\mathrm{BOOT}}(x) = (1/B) \sum_{b=1}^{B} I\left\{\left(\rho(P^*_{w_n,n}) - \rho_w(P_n)\right) \le x\right\}$,
- $R^{(1)}_{\mathrm{BOOT}}(x) = (1/B) \sum_{b=1}^{B} I\left\{\left(\rho(P^*_{w_n,\hat{n}_{new}}) - \rho_{\sqrt{w}}(P_n)\right) \le x\right\}$,
- $R^{(2)}_{\mathrm{BOOT}}(x) = (1/B) \sum_{b=1}^{B} I\left\{\left(\rho(P^*_{w_n,\hat{n}_{new}}) - \rho_w(P_n)\right) \le x\right\}$,
- $$\begin{aligned} R^{(3)}_{\mathrm{BOOT}}(x) =& (1/B) \sum_{b=1}^{B} I\Big\{\left(\rho(P^*_{w_n,\hat{n}_{new}}) - \rho^*\right) \times \sqrt{n/\hat{n}_{new}} \times \\ & \times \sqrt{f_c} + \rho^* - \rho_w(P_n) \le x\Big\} , \end{aligned}$$

  where ρ^* is the Monte Carlo approximation of the bootstrap estimator and the correction factor, f_c, is given by

  $$f_c = \left\{V_{\mathrm{BOOT}} + n^{-1} a_3 D^2_{est}\right\} / V_{\mathrm{BOOT}}$$

  with $V_{\mathrm{BOOT}} = [\mathrm{Var}(\rho(P_{w_n,n}))]^B_{\mathrm{BOOT}}$ the bootstrap estimator of the variance of the usual estimator of the correlation coefficient in the weighted sample and $D_{est} = \rho_{\sqrt{w}}(P_n) - \rho_{\sqrt{w}}(P)$.

As above, the “true” distribution of $\rho_{\sqrt{w}}(P_n) - \rho_{\sqrt{w}}(P)$ was estimated through an independent simulation study based on 5000 samples. This sample of

5000 observations was centered using its mean. Then, the empirical percentiles were computed. The previous step, was repeated 20 times and the final estimate of each percentile is the median of the obtained values over the 20 replications.

Table 7: Comparison of different bootstrap distributions with the "true" distribution of the weighted estimator for the correlation coefficient when $n = 20$.

C_0										
p	2.5	5	10	25	50	75	90	95	97.5	KS
$R^{(\mathrm{IF})}_{\mathrm{BOOT}}$	2.58	4.98	9.72	24.91	53.98	80.39	91.93	95.17	96.82	5.39
$R^{(1)}_{\mathrm{BOOT}}$	2.64	5.14	9.83	24.88	54.09	80.46	91.87	95.14	96.84	5.46
$R^{(2)}_{\mathrm{BOOT}}$	2.49	4.96	9.61	24.63	53.92	80.43	91.88	95.13	96.82	5.43
$R^{(3)}_{\mathrm{BOOT}}$	2.55	5.03	9.68	24.69	53.89	80.31	91.77	95.04	96.74	5.31

C_1										
p	2.5	5	10	25	50	75	90	95	97.5	KS
$R^{(\mathrm{IF})}_{\mathrm{BOOT}}$	2.72	5.26	10.02	25.32	54.99	80.98	92.08	95.18	96.77	5.98
$R^{(1)}_{\mathrm{BOOT}}$	2.63	5.17	9.88	25.10	54.25	80.47	91.74	94.81	96.41	5.47
$R^{(2)}_{\mathrm{BOOT}}$	2.53	5.08	9.82	25.15	54.50	80.85	92.06	95.03	96.55	5.85
$R^{(3)}_{\mathrm{BOOT}}$	2.60	5.17	9.91	25.19	54.42	80.66	91.86	94.87	96.42	5.66

Table 7 summarizes the results obtained. We observe that the approximations are better for the extreme quantiles than for the central ones, in all cases. It is worth noting that, for inference purposes, the extreme quantiles are the relevant ones.

5. CONCLUSIONS

The IFB procedure discussed in this paper allows to use resampling methods for robust inference, computing a robust estimator only for the original sample and avoiding the problems related with bootstrapping a robust estimator. It has shown to be effective for the location model. On the other hand, for the logistic regression model it shows a performance similar to that of the asymptotic confidence intervals.

To solve some problems of the procedure including the choice of the tuning constant and the identification of the functional being bootstrapped, a generalized influence function bootstrap is introduced. The empirical studies suggest that the generalized procedure IFB* has good properties, fixing some of the drawbacks of the original IFB procedure.

6. APPENDIX: SOME ASYMPTOTIC RESULTS

6.1. Convergence of the weighted empirical distribution to the weighted distribution

In this section, we will derive asymptotic results related to the consistency properties of the proposal. Let us first introduce some notation.

Let $\mathbf{X}_1, ..., \mathbf{X}_n$ be i.i.d. observations such that $\mathbf{X}_i \in \mathbb{R}^p$ with the same distribution as $\mathbf{X}$, where $\mathbf{X} \sim P$ and $\boldsymbol{\theta}_0 \in \Theta \subset \mathbb{R}^q$. Usually, $\boldsymbol{\theta}$ is the parameter allowing to parametrize the distribution of $\mathbf{X}$. Now, assume that $\widehat{\boldsymbol{\theta}}$ is a consistent estimator of $\boldsymbol{\theta}_0$ and denote by $\mathbb{P}_n$ the empirical distribution.

Given a weight function $w_1 : \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}$ such that $w_1 \geq 0$, define the following functions

$$(6.1) \qquad H_n(\mathbf{t}, \boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^{n} w_1(\mathbf{X}_i, \boldsymbol{\theta}) I_{(-\infty, \mathbf{t}]}(\mathbf{X}_i)$$

$$(6.2) \qquad H(\mathbf{t}, \boldsymbol{\theta}) = \mathbb{E}_P w_1(\mathbf{X}, \boldsymbol{\theta}) I_{(-\infty,t]}(\mathbf{X}) = P w_1(\cdot, \boldsymbol{\theta}) I_{(-\infty,\mathbf{t}]}$$

and note that $H(t, \boldsymbol{\theta}) = \mathbb{E}_P H_n(t, \boldsymbol{\theta})$.

It is worth noticing that, in Section 2 as in Amado and Pires (2004), the weighted empirical distribution involves a weight function w_1 that equals $w_1(\mathbf{x}, \boldsymbol{\theta}) = w(\mathbf{x}, \boldsymbol{\theta}) \{\int w(\mathbf{u}, \boldsymbol{\theta}) dP(\mathbf{u})\}^{-1}$ and thus, the distribution function used therein is of the form given in (6.2).

Let us assume that $P w_1 = \mathbb{E}_P w_1(\mathbf{X}, \boldsymbol{\theta}) = 1$ and that $W_1(x) = \sup_{\boldsymbol{\theta} \in \Theta} w_1(\mathbf{x}, \boldsymbol{\theta})$ is such that $P W_1^2 < \infty$.

We consider the following family of functions

$$\begin{aligned}
\mathcal{F} &= \{f_{\boldsymbol{\theta},\mathbf{t}} : \mathbb{R}^p \to \mathbb{R} \text{ such that } f_{\boldsymbol{\theta},\mathbf{t}}(\mathbf{x}) = w_1(\mathbf{x}, \boldsymbol{\theta}) I_{(-\infty,\mathbf{t}]}(\mathbf{x}) \ , \boldsymbol{\theta} \in \Theta \text{ and } \mathbf{t} \in \mathbb{R}^p\} \\
\mathcal{F}_0 &= \{f_{\mathbf{t}} : \mathbb{R}^p \to \mathbb{R} \text{ such that } f_{\mathbf{t}}(\mathbf{x}) = w_1(\mathbf{x}, \boldsymbol{\theta}_0) I_{(-\infty,\mathbf{t}]}(\mathbf{x}) \ , \mathbf{t} \in \mathbb{R}^p\} \\
\mathcal{W} &= \{f_{\boldsymbol{\theta}} : \mathbb{R}^p \to \mathbb{R} \text{ such that } f_{\boldsymbol{\theta}}(\mathbf{x}) = w_1(\mathbf{x}, \boldsymbol{\theta}) \ , \boldsymbol{\theta} \in \Theta\} \\
\mathcal{G} &= \{g_{\mathbf{t}} : \mathbb{R}^p \to \mathbb{R} \text{ such that } g_{\mathbf{t}}(\mathbf{x}) = I_{(-\infty,\mathbf{t}]}(\mathbf{x}) \ , \mathbf{t} \in \mathbb{R}^p\} \ .
\end{aligned}$$

We have that $\mathcal{F} = \mathcal{W} \cdot \mathcal{G}$ and $H_n(\mathbf{t}, \boldsymbol{\theta}) - H(\mathbf{t}, \boldsymbol{\theta}) = (\mathbb{P}_n - P) f_{\boldsymbol{\theta},\mathbf{t}}$. Denote by $\mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - P)$.

It is worth noticing that, when w_1 is bounded, $\mathcal{G}$ and $\mathcal{F}_0$ are both P-Glivenko–Cantelli and Donsker with envelope $G(\mathbf{x}) \equiv 1$ and $F_0(\mathbf{x}) = w_1(\mathbf{x}, \boldsymbol{\theta}_0)$.

Proposition 6.1 states that $H_n(\mathbf{t}, \boldsymbol{\theta})$ is a uniformly strongly consistent estimator of $H(\mathbf{t}, \boldsymbol{\theta})$ giving also the rate of this convergence.

We will need the following assumptions

A1. $|w_1(\mathbf{x}, \boldsymbol{\theta}_1) - w_1(\mathbf{x}, \boldsymbol{\theta}_2)| \le \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\| F(\mathbf{x})$, with $PF^2 < \infty$ and Θ compact

A2. $\mathcal{W} = \psi(\mathcal{L})$ with $\mathcal{L}$ a finite-dimensional family of functions and $\psi : \mathbb{R} \to \mathbb{R}$ a bounded function with bounded variation.

A3. W_1 is bounded.

A4. $w_1(\cdot, \boldsymbol{\theta})$ is continuous in $\boldsymbol{\theta}$.

A5. H is continuously differentiable in $\boldsymbol{\theta}$ such that $H'(\mathbf{t}, \boldsymbol{\theta}) = \partial H(\mathbf{t}, \boldsymbol{\theta})/\partial \boldsymbol{\theta}$ is bounded in $\mathbb{R}^p \times \mathcal{V}$ with $\mathcal{V}$ a neighbourhood of $\boldsymbol{\theta}_0$.

Remark 6.1. W_1 provides an envelope for $\mathcal{W}$. Moreover, under mild conditions on the functions w_1, $\mathcal{W}$ is P-Glivenko–Cantelli and Donsker family. For instance, $\mathcal{W}$ is both P-Glivenko–Cantelli and Donsker if either **A1** or **A2** holds.

Proposition 6.1. *Assume $\widehat{\boldsymbol{\theta}}$ is a consistent estimator and that either **A1** or **A2** holds. Then,*

a) $\sup_{\mathbf{t} \in \mathbb{R}^p} |H_n(\mathbf{t}, \widehat{\boldsymbol{\theta}}) - H(\mathbf{t}, \boldsymbol{\theta}_0)| \xrightarrow{a.s.} 0$.

b) *If, in addition, $\widehat{\boldsymbol{\theta}}$ has a root-n order of convergence and **A3** to **A5** hold, we have that*

$$\sqrt{n} \sup_{\mathbf{t} \in \mathbb{R}^p} |H_n(\mathbf{t}, \widehat{\boldsymbol{\theta}}) - H(\mathbf{t}, \boldsymbol{\theta}_0)| = O_{\mathbb{P}}(1) \ . \tag{6.3}$$

Proof of Proposition 6.1: a) Under either **A1** or **A2**, we will have that $\mathcal{F}$ is P-Glivenko–Cantelli and so,

$$\sup_{f \in \mathcal{F}} |(\mathbb{P}_n - P) f| = \sup_{\substack{\boldsymbol{\theta} \in \Theta \\ \mathbf{t} \in \mathbb{R}^p}} |H_n(\mathbf{t}, \boldsymbol{\theta}) - H(\mathbf{t}, \boldsymbol{\theta})| \xrightarrow{a.s.} 0.$$

In particular, we have that

$$\sup_{\mathbf{t} \in \mathbb{R}^p} |H_n(\mathbf{t}, \widehat{\boldsymbol{\theta}}) - H(\mathbf{t}, \widehat{\boldsymbol{\theta}})| \xrightarrow{a.s.} 0 \ .$$

Moreover, since either **A1** or **A2** holds, we have that $M_1(\boldsymbol{\theta}) = P w_1(\cdot, \boldsymbol{\theta})$ is a continuous function. Hence, we have that the consistency of $\widehat{\boldsymbol{\theta}}$ implies that $\sup_{\mathbf{t} \in \mathbb{R}^p} |H(\mathbf{t}, \widehat{\boldsymbol{\theta}}) - H(\mathbf{t}, \boldsymbol{\theta}_0)| \xrightarrow{a.s.} 0$ and thus, we obtain that

$$\sup_{\mathbf{t} \in \mathbb{R}^p} |H_n(\mathbf{t}, \widehat{\boldsymbol{\theta}}) - H(\mathbf{t}, \boldsymbol{\theta}_0)| \xrightarrow{a.s.} 0 \ . \tag{6.4}$$

b) Using **A3**, we get that $\mathcal{F}$ is Donsker, so $\mathbb{G}_n = \sqrt{n}(P_n - P)$ converges weakly to a zero mean Gaussian process $\mathbb{G}$ in $\ell^\infty(\mathcal{F})$. Therefore, the following equicontinuity condition holds

$$\text{(6.5)} \qquad \lim_{\eta\to 0} \limsup_{n\to\infty} \mathbb{P}\left(\sup_{\rho_P(f_{\boldsymbol{\theta}_1,\mathbf{t}_1} - f_{\boldsymbol{\theta}_2,\mathbf{t}_2})<\eta} |\mathbb{G}_n(f_{\boldsymbol{\theta}_1,\mathbf{t}_1} - f_{\boldsymbol{\theta}_2,\mathbf{t}_2})| > \epsilon \right) = 0$$

with $\rho_P^2(f) = P(f - Pf)^2$. Note that, $\rho_P^2(f_{\boldsymbol{\theta}_1,\mathbf{t}} - f_{\boldsymbol{\theta}_2,\mathbf{t}}) \le \mathbb{E}_P(w_1(\mathbf{X},\boldsymbol{\theta}_1) - w_1(\mathbf{X},\boldsymbol{\theta}_2))^2 = B(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ where the function $B(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ satisfies that $\lim_{\boldsymbol{\theta}\to\boldsymbol{\theta}_0} B(\boldsymbol{\theta}, \boldsymbol{\theta}_0) = 0$, since $w_1(\cdot, \boldsymbol{\theta})$ is continuous in $\boldsymbol{\theta}$ and W_1 is bounded. Then, using that $\widehat{\boldsymbol{\theta}}$ is consistent, we obtain that $\sup_{\mathbf{t}\in\mathbb{R}^p} \rho_P^2(f_{\widehat{\boldsymbol{\theta}},\mathbf{t}} - f_{\boldsymbol{\theta}_0,\mathbf{t}}) \xrightarrow{p} 0$ which implies that

$$\sup_{\mathbf{t}\in\mathbb{R}^p} |\mathbb{G}_n(f_{\widehat{\boldsymbol{\theta}},\mathbf{t}} - f_{\boldsymbol{\theta}_0,\mathbf{t}})| \xrightarrow{p} 0 \ .$$

Therefore, $\mathbb{G}_n f_{\widehat{\boldsymbol{\theta}},\mathbf{t}}$ has the same asymptotic distribution as $\mathbb{G}_n f_{\boldsymbol{\theta}_0,\mathbf{t}}$ in $\ell^\infty(\mathcal{F}_0)$. Using that $\mathcal{F}_0$ is Donsker, we get that $\mathbb{G}_n f_{\boldsymbol{\theta}_0,\mathbf{t}}$ converges to a zero mean Gaussian process $\mathbb{G}_0$ in $\ell^\infty(\mathcal{F}_0)$ with covariances given by

$$\begin{aligned} \mathbb{E}\mathbb{G}_0 f_{\boldsymbol{\theta}_0,\mathbf{t}_1} \mathbb{G}_0 f_{\boldsymbol{\theta}_0,\mathbf{t}_1} \;&=\; \mathbb{E}_P w_1^2(\mathbf{X}, \boldsymbol{\theta}_0) I_{(-\infty,\mathbf{t}_1]}(\mathbf{X}) I_{(-\infty,\mathbf{t}_2]}(\mathbf{X}) \,- \\ &\quad - \mathbb{E}_P w_1(\mathbf{X}, \boldsymbol{\theta}_0) I_{(-\infty,\mathbf{t}_1]}(\mathbf{X})\; \mathbb{E}_P w_1(\mathbf{X}, \boldsymbol{\theta}_0) I_{(-\infty,\mathbf{t}_2]}(\mathbf{X}) \ . \end{aligned}$$

In particular, $\sqrt{n} \sup_{\mathbf{t}\in\mathbb{R}^p} |H_n(\mathbf{t}, \widehat{\boldsymbol{\theta}}) - H(\mathbf{t}, \widehat{\boldsymbol{\theta}})|$ is tight and has the same asymptotic distribution as $\sqrt{n} \sup_{\mathbf{t}\in\mathbb{R}^p} |H_n(\mathbf{t}, \boldsymbol{\theta}_0) - H(\mathbf{t}, \boldsymbol{\theta}_0)|$.

Using that $\widehat{\boldsymbol{\theta}}$ has a root-n order of convergence and the fact that **A5** implies that H is continuously differentiable with bounded first derivative in a neighbourhood of $\boldsymbol{\theta}_0$, we have that (6.3) holds concluding the proof of b). □

Remark 6.2. The asymptotic distribution of $\sqrt{n} \sup_{\mathbf{t}\in\mathbb{R}^p} |H_n(\mathbf{t}, \widehat{\boldsymbol{\theta}}) - H(\mathbf{t}, \boldsymbol{\theta}_0)|$ may depend on that of $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$. Using analogous arguments, it is possible to show that

i) If $\mathbb{E}_P W_1(\mathbf{X}) \|\mathbf{X}\| < \infty$, then

$$\sup_{\boldsymbol{\theta}\in\Theta} \left\| \frac{1}{n} \sum_{i=1}^n w_1(\mathbf{X}_i, \boldsymbol{\theta}) \mathbf{X}_i - \mathbb{E}_P w_1(\mathbf{X}, \boldsymbol{\theta}) \mathbf{X} \right\| \xrightarrow{a.s.} 0$$

and so, if $\mathbf{A}(\boldsymbol{\theta}) = \mathbb{E}_P w_1(\mathbf{X}, \boldsymbol{\theta}) \mathbf{X}$ is a continuous function of $\boldsymbol{\theta}$, we have that

$$\left\| \frac{1}{n} \sum_{i=1}^n w_1(\mathbf{X}_i, \widehat{\boldsymbol{\theta}}) \mathbf{X}_i - \mathbb{E}_P w_1(\mathbf{X}, \boldsymbol{\theta}_0) \mathbf{X} \right\| \xrightarrow{a.s.} 0 \ ,$$

ii) If $\mathbb{E}_P W_1^2(\mathbf{X}) \|\mathbf{X}\|^2 < \infty$, then $Z_n = \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n w_1(\mathbf{X}_i, \widehat{\boldsymbol{\theta}}) \mathbf{X}_i - \mathbf{A}(\widehat{\boldsymbol{\theta}}) \right)$ is tight and has the same asymptotic distribution as $Z_{n,0} =$

$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n w_1(\mathbf{X}_i,\boldsymbol{\theta}_0)\mathbf{X}_i - \mathbf{A}(\boldsymbol{\theta}_0)\right)$ since $Z_n - Z_{n,0} \xrightarrow{p} 0$. Moreover,

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n w_1(\mathbf{X}_i,\widehat{\boldsymbol{\theta}})\mathbf{X}_i - \mathbf{A}(\boldsymbol{\theta}_0)\right) = Z_n + \sqrt{n}\left(\mathbf{A}(\boldsymbol{\theta}_0) - \mathbf{A}(\widehat{\boldsymbol{\theta}})\right)$$
$$= Z_{n,0} + \sqrt{n}\left(\mathbf{A}(\boldsymbol{\theta}_0) - \mathbf{A}(\widehat{\boldsymbol{\theta}})\right) + o_{\mathbb{P}}(1)\,.$$

Assume that $\widehat{\boldsymbol{\theta}}$ has a root-n order of convergence and that $\mathbf{A}(\boldsymbol{\theta})$ is continuously differentiable in $\boldsymbol{\theta}$. Denote $\mathbf{A}_0' = \partial\mathbf{A}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ where

$$\partial\mathbf{A}(\boldsymbol{\theta})/\partial\boldsymbol{\theta} = \begin{pmatrix} \frac{\partial A_1(\boldsymbol{\theta})}{\partial\theta_1} & \cdots & \frac{\partial A_p(\boldsymbol{\theta})}{\partial\theta_1} \\ \vdots & \cdots & \vdots \\ \frac{\partial A_1(\boldsymbol{\theta})}{\partial\theta_q} & \cdots & \frac{\partial A_p(\boldsymbol{\theta})}{\partial\theta_q} \end{pmatrix}.$$

Then, we have that

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n w_1(\mathbf{X}_i,\widehat{\boldsymbol{\theta}})\mathbf{X}_i - A(\boldsymbol{\theta}_0)\right) = Z_{n,0} - (\mathbf{A}_0')^{\mathrm{T}}\sqrt{n}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) + o_{\mathbb{P}}(1)$$

and so, again depending on $\mathbf{A}_0'$, the asymptotic distribution of $\sqrt{n}\left(\sum_{i=1}^n w_1(\mathbf{X}_i,\widehat{\boldsymbol{\theta}})\mathbf{X}_i/n - A(\boldsymbol{\theta}_0)\right)$ may depend on that of $\sqrt{n}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)$.

Remark 6.3. As pointed out above, for the weighted empirical distribution considered in this paper, w_1 equals $w_1(\mathbf{x},\boldsymbol{\theta}) = w(\mathbf{x},\boldsymbol{\theta})\{\int w(\mathbf{u},\boldsymbol{\theta})dP(\mathbf{u})\}^{-1}$. Thus, the function used in practice is not H_n but $\widetilde{H}_n$ defined as

$$\begin{aligned}\widetilde{H}_n(\mathbf{t},\boldsymbol{\theta}) &= \left\{\frac{1}{n}\sum_{j=1}^n w(\mathbf{X}_j,\boldsymbol{\theta})\right\}^{-1} \frac{1}{n}\sum_{i=1}^n w(\mathbf{X}_i,\boldsymbol{\theta}) I_{(-\infty,\mathbf{t}]}(\mathbf{X}_i) \\ &= H_n(\mathbf{t},\boldsymbol{\theta}) M_n(\boldsymbol{\theta})^{-1} M(\boldsymbol{\theta})\,.\end{aligned}$$

where $M(\boldsymbol{\theta}) = Pw(\cdot,\boldsymbol{\theta}) = \int w(\mathbf{u},\boldsymbol{\theta})dP(\mathbf{u})$ and $M_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{j=1}^n w(\mathbf{X}_j,\boldsymbol{\theta})$. Note that

$$\begin{aligned}\widetilde{H}_n(\mathbf{t},\widehat{\boldsymbol{\theta}}) - H(\mathbf{t},\boldsymbol{\theta}_0) &= H_n(\mathbf{t},\widehat{\boldsymbol{\theta}}) - H(\mathbf{t},\boldsymbol{\theta}_0) + \widetilde{H}_n(\mathbf{t},\widehat{\boldsymbol{\theta}}) - H_n(\mathbf{t},\widehat{\boldsymbol{\theta}}) \\ &= H_n(\mathbf{t},\widehat{\boldsymbol{\theta}}) - H(\mathbf{t},\boldsymbol{\theta}_0) + M_n\widehat{\boldsymbol{\theta}})^{-1}\left[M(\widehat{\boldsymbol{\theta}}) - M_n(\widehat{\boldsymbol{\theta}})\right] \\ &\quad \left(H_n(\mathbf{t},\widehat{\boldsymbol{\theta}}) - H(\mathbf{t},\boldsymbol{\theta}_0)\right) + M_n(\widehat{\boldsymbol{\theta}})^{-1}\left[M(\widehat{\boldsymbol{\theta}}) - M_n(\widehat{\boldsymbol{\theta}})\right] H(\mathbf{t},\boldsymbol{\theta}_0)\,.\end{aligned}$$

Hence, if we denote by $\widehat{\Delta}_n(\mathbf{t}) = H_n(\mathbf{t},\widehat{\boldsymbol{\theta}}) - H(\mathbf{t},\boldsymbol{\theta}_0)$, we have that

$$\begin{aligned}\widetilde{H}_n(\mathbf{t},\widehat{\boldsymbol{\theta}}) - H(\mathbf{t},\boldsymbol{\theta}_0) &= \widehat{\Delta}_n(\mathbf{t})\left\{1 + M_n(\widehat{\boldsymbol{\theta}})^{-1}\left[M(\widehat{\boldsymbol{\theta}}) - M_n(\widehat{\boldsymbol{\theta}})\right]\right\} + \\ &\quad + M_n(\widehat{\boldsymbol{\theta}})^{-1}\left[M(\widehat{\boldsymbol{\theta}}) - M_n(\widehat{\boldsymbol{\theta}})\right] H(\mathbf{t},\boldsymbol{\theta}_0)\,.\end{aligned}$$

Using that $\mathcal{W}$ is Glivenko–Cantelli, we get

$$M_n(\widehat{\boldsymbol{\theta}}) - M(\widehat{\boldsymbol{\theta}}) = \frac{1}{n}\sum_{j=1}^{n} w(\mathbf{X}_j, \widehat{\boldsymbol{\theta}}) - \int w(\mathbf{u}, \widehat{\boldsymbol{\theta}}) dP(\mathbf{u}) \xrightarrow{a.s.} 0\,,$$

which together with (6.4) and the facts that $\int w(\mathbf{u}, \boldsymbol{\theta}_0) dP(\mathbf{u}) > 0$ and $M(\boldsymbol{\theta}) = Pw(\cdot, \boldsymbol{\theta})$ is a continuous function entails that

$$\sup_{\mathbf{t}\in\mathbb{R}^p} |\widetilde{H}_n(\mathbf{t}, \widehat{\boldsymbol{\theta}}) - H(\mathbf{t}, \boldsymbol{\theta}_0)| \xrightarrow{a.s.} 0\ .$$

On the other hand, (6.3) entails that $\sqrt{n}\sup_{\mathbf{t}\in\mathbb{R}^p} |\widehat{\Delta}_n(\mathbf{t})| = O_{\mathbb{P}}(1)$, hence

$$\begin{aligned}\sqrt{n}\sup_{\mathbf{t}\in\mathbb{R}^p} |\widetilde{H}_n(\mathbf{t}, \widehat{\boldsymbol{\theta}}) - H(\mathbf{t}, \boldsymbol{\theta}_0)| \ \le\ & O_{\mathbb{P}}(1)\left|1 + M_n(\widehat{\boldsymbol{\theta}})^{-1}\left[M(\widehat{\boldsymbol{\theta}}) - M_n(\widehat{\boldsymbol{\theta}})\right]\right| \\ & + |M_n(\widehat{\boldsymbol{\theta}})^{-1}|\sqrt{n}\left|M(\widehat{\boldsymbol{\theta}}) - M_n(\widehat{\boldsymbol{\theta}})\right| M(\boldsymbol{\theta}_0)\ .\end{aligned}$$

Using that $\mathcal{W}$ is Donsker, we obtain that $\sqrt{n}\left|M(\widehat{\boldsymbol{\theta}}) - M_n(\widehat{\boldsymbol{\theta}})\right| = O_{\mathbb{P}}(1)$, which implies that

$$\sqrt{n}\sup_{\mathbf{t}\in\mathbb{R}^p} |\widetilde{H}_n(\mathbf{t}, \widehat{\boldsymbol{\theta}}) - H(\mathbf{t}, \boldsymbol{\theta}_0)| = O_{\mathbb{P}}(1)\ ,$$

as desired.

Moreover, as above, we have that $\sqrt{n}\left[M(\widehat{\boldsymbol{\theta}}) - M_n(\widehat{\boldsymbol{\theta}})\right]$ has the same asymptotic distribution as $\sqrt{n}\left[M(\boldsymbol{\theta}_0) - M_n(\boldsymbol{\theta}_0)\right]$, so, using that $M(\boldsymbol{\theta}_0) \neq 0$, we have

$$\begin{aligned}\sqrt{n}\left(\widetilde{H}_n(\mathbf{t}, \widehat{\boldsymbol{\theta}}) - H(\mathbf{t}, \boldsymbol{\theta}_0)\right) =& \sqrt{n}\widehat{\Delta}_n(\mathbf{t}) - M(\boldsymbol{\theta}_0)^{-1}\times \\ & \times\sqrt{n}\left[M_n(\boldsymbol{\theta}_0) - M(\boldsymbol{\theta}_0)\right] H(\mathbf{t}, \boldsymbol{\theta}_0) + o_{\mathbb{P}}(1)\ .\end{aligned}$$

An analogous expression can be derived for the mean computed with $\widetilde{H}_n(\mathbf{t}, \widehat{\boldsymbol{\theta}})$.

6.2. Some results related with the bootstrap

In this section, we will derive some results concerning the bootstrap procedures. We will fix some notation. For the sake of simplicity denote by $p_{i,\boldsymbol{\theta}} = p_i(\mathbf{X}_i, \boldsymbol{\theta}) = w_1(\mathbf{X}_i, \boldsymbol{\theta})/n$. Then, $H_n(\mathbf{t}, \boldsymbol{\theta}) = \sum_{i=1}^n p_{i,\boldsymbol{\theta}} I_{(-\infty,\mathbf{t}]}(\mathbf{X}_i)$ and the bootstrap distribution of H_n is

$$H_n^\star(\mathbf{t}, \boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} W_{n,i,\boldsymbol{\theta}} I_{(-\infty,\mathbf{t}]}(\mathbf{X}_i)$$

where $(W_{n,1,\boldsymbol{\theta}}, ..., W_{n,n,\boldsymbol{\theta}})|\vec{\mathbf{X}} \sim \mathcal{M}(n, (p_{1,\boldsymbol{\theta}}, ..., p_{n,\boldsymbol{\theta}})$ with $\vec{\mathbf{X}} = (\mathbf{X}_1, \cdots, \mathbf{X}_n)$.

It is worth noticing that $\mathbb{E}_P W_{n,i,\boldsymbol{\theta}}|\vec{\mathbf{X}} = np_{i,\boldsymbol{\theta}}$ entails that $\mathbb{E}_P(H_n^\star(\mathbf{t}, \boldsymbol{\theta}) - H_n(\mathbf{t}, \boldsymbol{\theta})) = 0$. Define $\widehat{\boldsymbol{\mu}}_{\boldsymbol{\theta}} = \sum_{i=1}^n p_{i,\boldsymbol{\theta}}\mathbf{X}_i$ and $\widehat{\boldsymbol{\mu}}_{\boldsymbol{\theta}}^\star = \frac{1}{n}\sum_{i=1}^n W_{n,i,\boldsymbol{\theta}}\mathbf{X}_i$. The next proposition states that, conditionally on the sample, the difference between $\widehat{\boldsymbol{\mu}}_{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\mu}}^{\star}_{\boldsymbol{\theta}}$ converges to 0 in probability.

Proposition 6.2. *Assume that* ***A3*** *holds. Then,*

$$H^{\star}_n(\mathbf{t},\widehat{\boldsymbol{\theta}}) - H_n(\mathbf{t},\widehat{\boldsymbol{\theta}})|\vec{\mathbf{X}} \xrightarrow{p} 0 \ , \tag{6.6}$$

If, in addition $\sup_{\boldsymbol{\theta}\in\Theta}\sup_{\mathbf{x}}\|w_1(\mathbf{x},\theta)\mathbf{x}\|<\infty$*, we have that* $\widehat{\boldsymbol{\mu}}^{\star}_{\widehat{\boldsymbol{\theta}}} - \widehat{\boldsymbol{\mu}}_{\widehat{\boldsymbol{\theta}}}|\vec{\mathbf{X}} \xrightarrow{p} 0$.

Proof of Proposition 6.2: Let us compute $\operatorname{Var}\left(H^{\star}_n(\mathbf{t},\boldsymbol{\theta}) - H_n(\mathbf{t},\boldsymbol{\theta})\right)$. Let $f_{\mathbf{t}}(\mathbf{x}) = I_{(-\infty,\mathbf{t}]}(\mathbf{x})$, then

$$\begin{aligned}\operatorname{Var}\left(H^{\star}_n(\mathbf{t},\boldsymbol{\theta}) - H_n(\mathbf{t},\boldsymbol{\theta})\right) &= \sum_{i=1}^{n}\operatorname{Var}\left(\left(\frac{1}{n}W_{n,i,\boldsymbol{\theta}} - p_{i,\boldsymbol{\theta}}\right)f_{\mathbf{t}}(\mathbf{X}_i)\right)\\ &+ 2\sum_{i<j}\operatorname{Cov}\left(\left(\frac{1}{n}W_{n,i,\boldsymbol{\theta}} - p_{i,\boldsymbol{\theta}}\right)f_{\mathbf{t}}(\mathbf{X}_i), \left(\frac{1}{n}W_{n,j,\boldsymbol{\theta}} - p_{j,\boldsymbol{\theta}}\right)f_{\mathbf{t}}(\mathbf{X}_j)\right) \ .\end{aligned}$$

Denote $Z_i = ((1/n)W_{n,i,\boldsymbol{\theta}} - p_{i,\boldsymbol{\theta}})f_{\mathbf{t}}(\mathbf{X}_i)$. Then, using that $\mathbb{E}_P Z_i = 0$, we have that

$$\begin{aligned}\operatorname{Var}(Z_i) = \mathbb{E}_P Z_i^2 &= \mathbb{E}_P\left[f^2_{\mathbf{t}}(\mathbf{X}_i)\mathbb{E}_P\left(\left(\frac{1}{n}W_{n,i,\boldsymbol{\theta}} - p_{i,\boldsymbol{\theta}}\right)^2|\vec{\mathbf{X}}\right)\right]\\ &= \frac{1}{n}\mathbb{E}_P f^2_{\mathbf{t}}(\mathbf{X}_1)p_{1,\boldsymbol{\theta}}(1-p_{1,\boldsymbol{\theta}}) \ .\end{aligned}$$

Similarly, we get that

$$\begin{aligned}\operatorname{Cov}(Z_i,Z_j) &= \mathbb{E}_P Z_i Z_j\\ &= \mathbb{E}_P\left[f_{\mathbf{t}}(\mathbf{X}_i)f_{\mathbf{t}}(\mathbf{X}_j)\mathbb{E}_P\left(\left(\frac{1}{n}W_{n,i,\boldsymbol{\theta}} - p_{i,\boldsymbol{\theta}}\right)\left(\frac{1}{n}W_{n,j,\boldsymbol{\theta}} - p_{j,\boldsymbol{\theta}}\right)|\vec{\mathbf{X}}\right)\right]\\ &= -\frac{1}{n}\mathbb{E}_P f_{\mathbf{t}}(\mathbf{X}_1)f_{\mathbf{t}}(\mathbf{X}_2)p_{1,\boldsymbol{\theta}}p_{2,\boldsymbol{\theta}} \ .\end{aligned}$$

Thus,

$$\begin{aligned}\operatorname{Var}\left(H^{\star}_n(\mathbf{t},\boldsymbol{\theta}) - H_n(\mathbf{t},\boldsymbol{\theta})\right) &= \mathbb{E}_P f^2_{\mathbf{t}}(\mathbf{X}_1)p_{1,\boldsymbol{\theta}}(1-p_{1,\boldsymbol{\theta}}) - 2\frac{1}{n}\binom{n}{2}\mathbb{E}_P f_{\mathbf{t}}(\mathbf{X}_1)f_{\mathbf{t}}(\mathbf{X}_2)p_{1,\boldsymbol{\theta}}p_{2,\boldsymbol{\theta}}\\ &= \frac{1}{n}\mathbb{E}_P f^2_{\mathbf{t}}(\mathbf{X}_1)w_1(\mathbf{X}_1,\boldsymbol{\theta})\left(1-\frac{1}{n}w_1(\mathbf{X}_1,\boldsymbol{\theta})\right)\\ &\quad -\frac{2}{n}\binom{n}{2}\frac{1}{n^2}\mathbb{E}_P f_{\mathbf{t}}(\mathbf{X}_1)f_{\mathbf{t}}(\mathbf{X}_2)w_1(\mathbf{X}_1,\boldsymbol{\theta})w_1(\mathbf{X}_2,\boldsymbol{\theta}) \ ,\end{aligned}$$

which entails that $H^{\star}_n(\mathbf{t},\boldsymbol{\theta}) - H_n(\mathbf{t},\boldsymbol{\theta}) \xrightarrow{p} 0$ for each fixed $\boldsymbol{\theta},\mathbf{t}$.

Moreover, we have the bounds

$$\begin{aligned}\left|\mathbb{E}_P f^2_{\mathbf{t}}(\mathbf{X}_1)w_1(\mathbf{X}_1,\boldsymbol{\theta})\left(1-\frac{1}{n}w_1(\mathbf{X}_1,\boldsymbol{\theta})\right)\right| &\le \mathbb{E}_P f^2_{\mathbf{t}}(\mathbf{X}_1)W_1(\mathbf{X}_1) = A_1\\ \left|\mathbb{E}_P f_{\mathbf{t}}(\mathbf{X}_1)f_{\mathbf{t}}(\mathbf{X}_2)w_1(\mathbf{X}_1,\boldsymbol{\theta})w_1(\mathbf{X}_2,\boldsymbol{\theta})\right| &\le \mathbb{E}_P f^2_{\mathbf{t}}(\mathbf{X}_1)W_1^2(\mathbf{X}_1) = A_2\end{aligned}$$

which imply that

$$\sup_{\boldsymbol{\theta}\in\Theta} \operatorname{Var}\left(H_n^\star(\mathbf{t},\boldsymbol{\theta}) - H_n(\mathbf{t},\boldsymbol{\theta})\right) \le \frac{1}{n}\left(A_1 + A_2\right) ,$$

so,

$$\sup_{\boldsymbol{\theta}\in\Theta} \mathbb{P}\left(|H_n^\star(\mathbf{t},\boldsymbol{\theta}) - H_n(\mathbf{t},\boldsymbol{\theta})| > \epsilon\right) \le \frac{1}{\epsilon^2}\frac{1}{n}\left(A_1 + A_2\right) .$$

The fact that $\mathbb{E}_P Z_i|\vec{\mathbf{X}} = 0$, $\operatorname{Cov}(Z_i, Z_j|\vec{\mathbf{X}}) = -(1/n) f_{\mathbf{t}}(\mathbf{X}_i) f_{\mathbf{t}}(\mathbf{X}_j) p_{i,\boldsymbol{\theta}} p_{j,\boldsymbol{\theta}}$ and $\operatorname{Var}(Z_i|\vec{\mathbf{X}}) = (1/n) f_{\mathbf{t}}^2(\mathbf{X}_i) p_{i,\boldsymbol{\theta}}^2$, imply

$$\operatorname{Var}\left(H_n^\star(\mathbf{t},\boldsymbol{\theta}) - H_n(\mathbf{t},\boldsymbol{\theta})|\vec{\mathbf{X}}\right) \;=\; \frac{1}{n}\sum_{i=1}^n f_{\mathbf{t}}^2(\mathbf{X}_i) p_{i,\boldsymbol{\theta}}^2 - \frac{2}{n}\sum_{i<j} f_{\mathbf{t}}(\mathbf{X}_i) f_{\mathbf{t}}(\mathbf{X}_j) p_{i,\boldsymbol{\theta}} p_{j,\boldsymbol{\theta}} .$$

Hence, using that W_1 is a bounded function and that $p_{i,\boldsymbol{\theta}} = w_1(\mathbf{X}_i, \boldsymbol{\theta})/n$, we get the following bound

$$\begin{aligned}
&\operatorname{Var}\left(H_n^\star(\mathbf{t},\boldsymbol{\theta}) - H_n(\mathbf{t},\boldsymbol{\theta})|\vec{\mathbf{X}}\right) \;\le \\
&\le\; \frac{1}{n^2}\|W_1\|_\infty^2 \frac{1}{n}\sum_{i=1}^n f_{\mathbf{t}}^2(\mathbf{X}_i) + \frac{1}{n^2}\|W_1\|_\infty^2 \frac{1}{n}\left(\sum_{i=1}^n f_{\mathbf{t}}(\mathbf{X}_i)\right)^2 \qquad (6.7) \\
&\le\; \frac{1}{n^2}\|W_1\|_\infty^2 \frac{1}{n}\sum_{i=1}^n f_{\mathbf{t}}^2(\mathbf{X}_i) + \frac{1}{n}\|W_1\|_\infty^2 \left(\frac{1}{n}\sum_{i=1}^n f_{\mathbf{t}}(\mathbf{X}_i)\right)^2 .
\end{aligned}$$

The fact that $|f_{\mathbf{t}}^2(\mathbf{X}_i)| \le 1$ entails that

$$\sup_{\boldsymbol{\theta}\in\Theta} \mathbb{P}\left(|H_n^\star(\mathbf{t},\boldsymbol{\theta}) - H_n(\mathbf{t},\boldsymbol{\theta})| > \epsilon|\vec{\mathbf{X}}\right) \le \frac{1}{\epsilon^2}\frac{2}{n}\|W_1\|_\infty^2 .$$

Hence,

$$\mathbb{P}\left(\left|H_n^\star(\mathbf{t},\widehat{\boldsymbol{\theta}}) - H_n(\mathbf{t},\widehat{\boldsymbol{\theta}})\right| > \epsilon|\vec{\mathbf{X}}\right) \le \frac{1}{\epsilon^2}\frac{2}{n}\|W_1\|_\infty^2$$

implying (6.6).

Let us denote $\widehat{\boldsymbol{\mu}}_{\boldsymbol{\theta}} = \sum_{i=1}^n p_{i,\boldsymbol{\theta}}\mathbf{X}_i$ and $\widehat{\boldsymbol{\mu}}_{\boldsymbol{\theta}}^\star = \frac{1}{n}\sum_{i=1}^n W_{n,i,\boldsymbol{\theta}}\mathbf{X}_i$. Taking $f(\mathbf{X}_i) = \mathbf{X}_i$ in (6.7), we obtain

$$\operatorname{Var}\left(\widehat{\boldsymbol{\mu}}_{\boldsymbol{\theta}}^\star - \widehat{\boldsymbol{\mu}}_{\boldsymbol{\theta}}|\vec{\mathbf{X}}\right) \;\le\; \frac{1}{n^2}\frac{1}{n}\sum_{i=1}^n \|f(\mathbf{X}_i)\|^2 w_1^2(\mathbf{X}_i,\theta) + \frac{1}{n}\left\|\frac{1}{n}\sum_{i=1}^n f(\mathbf{X}_i) w_1(\mathbf{X}_i,\theta)\right\|^2 .$$

Hence, since $B = \sup_{\boldsymbol{\theta}\in\Theta}\sup_{\mathbf{x}} \|f(\mathbf{X}_i) w_1(\mathbf{X}_i,\theta)\| < \infty$, we get that

$$\mathbb{P}\left(|\widehat{\boldsymbol{\mu}}_{\widehat{\boldsymbol{\theta}}}^\star - \widehat{\boldsymbol{\mu}}_{\widehat{\boldsymbol{\theta}}}| > \epsilon|\vec{\mathbf{X}}\right) \le \frac{1}{\epsilon^2}\frac{2}{n}B^2$$

implying that $\widehat{\boldsymbol{\mu}}_{\widehat{\boldsymbol{\theta}}}^\star - \widehat{\boldsymbol{\mu}}_{\widehat{\boldsymbol{\theta}}}|\vec{\mathbf{X}} \xrightarrow{p} 0$. □

ACKNOWLEDGMENTS

This work received financial support from Portuguese National Funds through FCT (Fundação para a Ciência e a Tecnologia) under the scope of project PEst-OE/MAT/UI0822/2011. This research was also partially supported by Grants 112-201101-00339 from CONICET, PICT 0397 from ANPCYT and W276 from the Universidad de Buenos Aires at Buenos Aires, Argentina and by the joint cooperation program ANPCYT–FCT PO/09/05–Proc. 44100.

We wish to thank an anonymous referee for valuable comments which led to an improved version of the original paper.

REFERENCES

[1] Amado, C. and Pires, A.M. (2004). Robust bootstrap with non-random weights based on the influence function, *Communications in Statistics — Simulation and Computation*, **33**, 377–396.

[2] Amado, C. (2003). *Robust Bootstrap Based on the Influence Function*, PhD Thesis (in portuguese), Technical University of Lisbon, Lisbon, Portugal. Available on http://www.math.ist.utl.pt/∼apires/phd.html

[3] Athreya, K.B. (1987). Bootstrap of the mean in the infinite variance case, *The Annals of Statistics*, **15**, 724–731.

[4] Croux, C., and Haesbroeck, G. (2003). Implementing the Bianco and Yohai estimator for logistic regression, *Computational Statistics & Data Analysis*, **44**, 273–295.

[5] Ghosh, M.; Parr, W.C.; Singh, K. and Babu, G.J. (1984). A note on bootstrapping the sample median, *The Annals of Statistics*, **12**, 1130–1135.

[6] Salibian-Barrera, M. (2000). *Contributions to the Theory of Robust Inference*, PhD Thesis, University of British Columbia, Vancouver, Canada.

[7] Salibian-Barrera, M. (2002). Fast and stable bootstrap methods for robust estimates, *Computing Science and Statistics*, **34**, 346–359. E. Wegman and A. Braverman (editors), Interface Foundation of North America, Inc., Fairfax Station, VA.

[8] Salibian-Barrera, M.; Van Aelst, S. and Willems, G. (2006). PCA based on multivariate MM-estimators with fast and robust bootstrap, *Journal of the American Statistical Association*, **101**, 1198–1211.

[9] Salibian-Barrera, M. and Zamar, R.H. (2002). Bootstrapping robust estimates of regression, *The Annals of Statistics*, **30**, 556–582.

[10] Singh, K. (1998). Breakdown theory for bootstrap quantiles, *The Annals of Statistics*, **26**, 1719–1732.

[11] Shao, J. (1990). Bootstrap estimation of the asymptotic variances of statistical functionals, *Annals of the Institute of Statistical Mathematics*, **42**, 737–752.

[12] Shao, J. (1992). Bootstrap variance estimators with truncation, *Statistics and Probability Letters*, **15**, 95–101.

[13] Srivastava, M.S. and Chan, Y.M. (1989). A comparison of bootstrap method and Edgeworth expansion in approximating the distribution of sample variance — one sample and two sample cases, *Communications in Statistics B*, **18**, 339–361.

[14] Stromberg A.J. (1997). Robust covariance estimates based on resampling, *Journal of Statistical Planning and Inference*, **57**, 321–334.

[15] Willems, G. and Van Aelst, S. (2005). Fast and Robust Bootstrap for LTS, *Computational Statistics and Data Analysis*, **48**, 703–715.