ISSN 1645-6726

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

REVSTAT

Statistical Journal

Volume 12, No.2
June 2014

REVSTAT

Statistical Journal

Catalogação Recomendada

REVSTAT. Lisboa, 2003-
Revstat : statistical journal / ed. Instituto Nacional de Estatística. - Vol. 1, 2003- . - Lisboa I.N.E., 2003- . - 30 cm
Semestral. - Continuação de : Revista de Estatística =
ISSN 0873-4275. - edição exclusivamente em inglês
ISSN 1645-6726

CREDITS

- **PUBLISHER**
  - *Instituto Nacional de Estatística, I.P. (INE, I.P.)*
    Av. António José de Almeida, 2
    1000-043 LISBOA
    PORTUGAL
    Tel.: +351 218 426 100
    Fax: +351 218 454 084
    Web site: http://www.ine.pt
    Customer Support Service
    (National network): 808 201 808
    (Other networks): +351 218 440 695
- **COVER DESIGN**
  - *Mário Bouçadas, designed on the stain glass window at INE, I.P., by the painter Abel Manta*
- **LAYOUT AND GRAPHIC DESIGN**
  - *Carlos Perpétuo*
- **PRINTING**
  - *Instituto Nacional de Estatística, I.P.*
- **EDITION**
  - *150 copies*
- **LEGAL DEPOSIT REGISTRATION**
  - *N.º 191915/03*
- **PRICE [VAT included]**
  - *€ 11,00*

INDEX

Abstracted/indexed in: *Current Index to Statistics, DOAJ, Google Scholar, Journal Citation Reports/Science Edition, Mathematical Reviews, Science Citation Index Expanded®, SCOPUS* and *Zentralblatt für Mathematic.*

REVSTAT – Statistical Journal
Volume 12, Number 2, June 2014, 101–118

UNIFORM APPROXIMATIONS FOR DISTRIBUTIONS OF CONTINUOUS RANDOM VARIABLES WITH APPLICATION IN DUAL STATIS METHOD

Authors: João Lita da Silva
– Department of Mathematics and CMA,
Faculty of Sciences and Technology, New University of Lisbon,
Quinta da Torre, 2829-516 Caparica, Lisbon, Portugal
jfls@fct.unl.pt

Luís Pedro Ramos
– Department of Mathematics and CMA,
Faculty of Sciences and Technology, New University of Lisbon,
Quinta da Torre, 2829-516 Caparica, Lisbon, Portugal
lpcr@fct.unl.pt

Received: December 2012 Revised: March 2013 Accepted: March 2013

Abstract:

- The matrix $\mathbf{S} = \left[\operatorname{tr}(\mathbf{W}_i\mathbf{Q}\mathbf{W}_j\mathbf{Q})\right]_{i,j=1,...,k}$ where $\mathbf{Q}$ is a symmetric positive definite matrix and $\mathbf{W}_i = \mathbf{X}_i'\mathbf{D}_i\mathbf{X}_i$, $i = 1, ..., k$ is formed by data tables $\mathbf{X}_i$ and diagonal matrices of weights $\mathbf{D}_i$, plays a central role in dual STATIS method. In this paper, we approximate the distribution function of the entries of $\mathbf{S}$, assuming data tables $\mathbf{X}_i$ given by $\mathbf{U}_i + \mathbf{E}_i$, $i = 1, ..., k$ with independent random matrices $\mathbf{E}_i$ representing errors, in order to obtain (approximately) the distribution of $\mathbf{Sv}$, where $\mathbf{v}$ is the orthonormal eigenvector of $\mathbf{S}$ associated to the largest eigenvalue. To achieve this goal, we approximate uniformly the distribution of each entry of $\mathbf{S}$. In general, our technique consists in to approximate uniformly the distribution sequence $\left\{g(\boldsymbol{V}_n + \boldsymbol{\mu}_n), n \geqslant 1\right\}$, where g is some smooth function of several variables, $\{\boldsymbol{V}_n,\, n \geqslant 1\}$ is a sequence of identically distributed random vectors of continuous type and $\{\boldsymbol{\mu}_n\}$ is a non-random vector sequence.

Key-Words:

- *dual STATIS method; uniform approximations.*

AMS Subject Classification:

- 62H25, 60E05(62E20), 60F05.

1. INTRODUCTION

The dual STATIS method is an exploratory technique of multivariate data analysis used to study simultaneously multiple data tables, each with information of groups of individuals measured on the same set of variables (see [4], [8] or [9]). The purpose of this method is to analyze the relationship between data tables and combine them into a compromise matrix corresponding to an optimal agreement between the data.

In this paper, we shall consider the n_i-by-p random data tables $\mathbf{X}_i$, $i = 1, ..., k$ consisting in the measurements of k groups of n_i individuals on the same set of p variables, the n_i-by-n_i diagonal matrices $\mathbf{D}_i$ of positive weights attached to the n_i observations of each matrix $\mathbf{X}_i$ in order to define

$$\mathbf{W}_i = \mathbf{X}_i'\mathbf{D}_i\mathbf{X}_i \ , \tag{1.1}$$

where prime denotes transpose. If the columns of $\mathbf{X}_i$ are $\mathbf{D}_i$-centered then $\mathbf{W}_i$ is the covariance matrix between the p variables of $\mathbf{X}_i$ and its elements corresponds to the scalar products between the variables in $\mathbb{R}^{n_i}$. To evaluate the closeness of two data configurations in $\mathbb{R}^{n_i}$ and $\mathbb{R}^{n_j}$, the trace $\mathrm{tr}(\mathbf{W}_i\mathbf{Q}\mathbf{W}_j\mathbf{Q})$, where $\mathbf{Q}$ is a p-by-p symmetric positive definite matrix, is commonly used as scalar product between $\mathbf{W}_i$ and $\mathbf{W}_j$, known as *Hilbert–Schmidt scalar product* between $\mathbf{W}_i$ and $\mathbf{W}_j$ (see [8], page 38).

We shall set $s_{ij} = \mathrm{tr}(\mathbf{W}_i\mathbf{Q}\mathbf{W}_j\mathbf{Q})$, $i, j = 1, ..., k$ as being the entries of the k-by-k interstructure matrix $\mathbf{S}$. The vectorial correlation coefficient RV of $\mathbf{W}_i$ and $\mathbf{W}_j$ is defined as

$$\mathrm{RV}(\mathbf{W}_i, \mathbf{W}_j) = \frac{\mathrm{tr}(\mathbf{W}_i\mathbf{Q}\mathbf{W}_j\mathbf{Q})}{\sqrt{\mathrm{tr}(\mathbf{W}_i\mathbf{Q}\mathbf{W}_i\mathbf{Q})\ \mathrm{tr}(\mathbf{W}_j\mathbf{Q}\mathbf{W}_j\mathbf{Q})}} \ .$$

(see [4]), which appears as a measure of similarity between $\mathbf{W}_i$ and $\mathbf{W}_j$. The reader is referred to [13] for further details on the RV coefficient. Moreover, from Cholesky decomposition (see [5], page 229), there exists a unique upper triangular p-by-p matrix $\mathbf{T}$ with positive diagonal elements such that $\mathbf{Q} = \mathbf{T}'\mathbf{T}$ and putting $\mathbf{A}_i = \mathbf{D}_i^{1/2}\mathbf{X}_i\mathbf{T}'$ we get

$$\mathbf{A}_i'\mathbf{A}_i = \mathbf{T}\mathbf{X}_i'\mathbf{D}_i\mathbf{X}_i\mathbf{T}'$$

which implies

$$\mathrm{tr}(\mathbf{W}_i\mathbf{Q}\mathbf{W}_j\mathbf{Q}) = \mathrm{tr}(\mathbf{A}_i'\mathbf{A}_i\mathbf{A}_j'\mathbf{A}_j) = \left\|\mathbf{A}_i\mathbf{A}_j'\right\|_{\mathrm{tr}}^2 \geqslant 0$$

where $\|\mathbf{A}\|_{\mathrm{tr}} = \sqrt{\mathrm{tr}(\mathbf{A}'\mathbf{A})}$ (see [5], page 60). Denoting by $\mathbf{a}_{i\ell}$, $\ell = 1, ..., n_i$ the rows of $\mathbf{A}_i$ and $\mathbf{a}_{jm}$, $m = 1, ..., n_j$ the rows of $\mathbf{A}_j$, $\mathbf{a}_{i\ell}\mathbf{a}_{jm}'$ is the covariance between $\mathbf{a}_{i\ell}$ and $\mathbf{a}_{jm}$ so that

$$\mathrm{tr}(\mathbf{W}_i\mathbf{Q}\mathbf{W}_j\mathbf{Q}) = \sum_{\ell=1}^{n_i}\sum_{m=1}^{n_j}\left[\mathrm{cov}(\mathbf{a}_{i\ell}, \mathbf{a}_{jm})\right]^2 .$$

Consider the eigenvalues $\rho_1 \geqslant \rho_2 ... \geqslant \rho_k$ of $\mathbf{S}$ and the corresponding orthonormal eigenvectors $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_k$. From the spectral theorem for symmetric matrices (see [7], page 104) we get

$$\mathbf{S} = \rho_1 \mathbf{v}_1 \mathbf{v}_1' + ... + \rho_k \mathbf{v}_k \mathbf{v}_k' = \mathbf{P}\mathbf{\Lambda}\mathbf{P}'$$

where $\mathbf{\Lambda}$ is the diagonal matrix whose elements are the corresponding eigenvalues and $\mathbf{P}'\mathbf{P} = \mathbf{I}$, where $\mathbf{I}$ is the k-by-k identity matrix. From the above expression, we can plot the ith stage in $\mathbb{R}^k$ as a point M_i whose coordinates are the ith row of $\mathbf{P}\mathbf{\Lambda}^{1/2}$ (see [9]). Along all text, we will assume that points M_i are all near by each other lying on around first axis, which lead us to $\rho_i \approx 0$, for each $i = 2, ..., k$ and

$$\mathbf{S} \approx \rho \mathbf{v}\mathbf{v}' \tag{1.2}$$

setting $\rho = \rho_1$ and $\mathbf{v} = \mathbf{v}_1$. In our model, we will also assume random errors on the data i.e.

$$\mathbf{X}_i = \mathbf{U}_i + \mathbf{E}_i \tag{1.3}$$

where $\mathbf{E}_i$ are independent n_i-by-p random matrices representing the errors with i.i.d. continuous entries and $\mathbf{U}_i$ are n_i-by-p non-random matrices. Moreover, we shall admit that $\mathrm{E}(\mathbf{S})$ has rank one, so that the spectral theorem for symmetric matrices allow us to write

$$\mathrm{E}(\mathbf{S}) = \lambda \boldsymbol{\alpha}\boldsymbol{\alpha}' \tag{1.4}$$

where λ is the largest eigenvalue of $\mathrm{E}(\mathbf{S})$ associated to the orthonormal eigenvector $\boldsymbol{\alpha}$. Hence, we are led to consider the model

$$\mathbf{S} = \lambda \boldsymbol{\alpha}\boldsymbol{\alpha}' + \boldsymbol{\mathcal{E}} \tag{1.5}$$

for some p-by-p random matrix $\boldsymbol{\mathcal{E}}$ satisfying $\mathrm{E}(\boldsymbol{\mathcal{E}}) = \mathbf{O}$ (null matrix).

Let us start with the following question: if the sequence of matrices $\mathbf{E}_i$, $i = 1, 2, ..., k$ are independent with i.i.d. continuous entries how can we compute the distribution function of each entry s_{ij} of the matrix $\mathbf{S}$? Generally, the distribution function of s_{ij} is hard to compute, so that our proposal answer to this question will be to approximate the distribution of s_{ij} by some computable distribution. More precisely, our results will permit us to approximate uniformly the distribution function of each entry of the random matrix $\mathbf{S}$ by its linear part. The Section 2 will describe in detail all the theoretical results required to fulfill our intentions. Once the distribution of the elements of $\mathbf{S}$ is achieved, we will be able to obtain (approximately) the distribution of $\widehat{\boldsymbol{\beta}} = \mathbf{S}\mathbf{v}$, which will be taken as an estimator of $\boldsymbol{\beta} = \lambda\boldsymbol{\alpha}$. The example exhibited in last section considering the elements of $\mathbf{E}_i$ i.i.d. normal distributed with zero mean and variance σ^2, it will illustrate our inferential purposes in a very clear way.

2. UNIFORM APPROXIMATIONS

In general way, our idea to approaching the distribution function of the entries of $\mathbf{S}$, will consist in expanding asymptotically a sequence of r.v.'s (with unknown distribution) to obtain a random sequence with identifiable distribution. Thereafter, with the aid of a uniform bound, we will establish the uniform approximation results imposing on the remainder term the asymptotic condition $o_{\Pr}(1)$.[1] The driving tool in the our proof technique is to consider asymptotic Taylor expansions.

Let us introduce the following notation: $\|\boldsymbol{a}\| = \big(\sum_{i=1}^{\kappa} a_i^2\big)^{1/2}$ for the Euclidean norm of a vector $\boldsymbol{a} \in \mathbb{R}^{\kappa}$ (see [7], page 264) and $\|\!|\boldsymbol{A}|\!\| = \Big(\sum_{i,j=1}^{\kappa} a_{ij}^2\Big)^{1/2}$ for the norm of a real matrix $\boldsymbol{A} = \big[a_{ij}\big]_{i,j=1,\ldots,\kappa}$ (known as *Frobenius norm*, see [7] page 291). Given a differentiable mapping $\boldsymbol{\varphi}\colon \mathbb{R}^N \to \mathbb{R}^M$ we shall denote the jacobian matrix of $\boldsymbol{\varphi}$ at the point $\boldsymbol{x}$ by $D\boldsymbol{\varphi}(\boldsymbol{x})$. For a differentiable real-valued function $\varphi\colon \mathbb{R}^N \to \mathbb{R}$ having second partial derivatives, $D\varphi(\boldsymbol{x})$ and $D^2\varphi(\boldsymbol{x})$ will denote, respectively, the gradient vector and the Hessian matrix of φ at the point $\boldsymbol{x}$.

Lemma 2.1. *If* $\boldsymbol{\varphi} \in C^1(\mathbb{R}^N, \mathbb{R}^N)$ *satisfies* $\|\!|D\boldsymbol{\varphi}(\boldsymbol{x})|\!\| = o\big(\,\|\boldsymbol{\varphi}(\boldsymbol{x})\|\,\big)$, $\|\boldsymbol{x}\| \to \infty$ *then* $\|\!|D\boldsymbol{\varphi}\big(\boldsymbol{x}+\boldsymbol{\theta}(\boldsymbol{x})\big)|\!\| = o\big(\,\|\boldsymbol{\varphi}(\boldsymbol{x})\|\,\big)$, $\|\boldsymbol{x}\| \to \infty$ *for any bounded function* $\boldsymbol{\theta}(\boldsymbol{x})$.

Proof: We have,

$$\frac{\|\!|D\boldsymbol{\varphi}\big(\boldsymbol{x}+\boldsymbol{\theta}(\boldsymbol{x})\big)|\!\|}{\|\boldsymbol{\varphi}(\boldsymbol{x})\|} = \frac{\|\!|D\boldsymbol{\varphi}\big(\boldsymbol{x}+\boldsymbol{\theta}(\boldsymbol{x})\big)|\!\|}{\big\|\boldsymbol{\varphi}\big(\boldsymbol{x}+\boldsymbol{\theta}(\boldsymbol{x})\big)\big\|} \frac{\big\|\boldsymbol{\varphi}\big(\boldsymbol{x}+\boldsymbol{\theta}(\boldsymbol{x})\big)\big\|}{\|\boldsymbol{\varphi}(\boldsymbol{x})\|}$$

and it is sufficient to prove that $\big\|\boldsymbol{\varphi}\big(\boldsymbol{x}+\boldsymbol{\theta}(\boldsymbol{x})\big)\big\| \sim \|\boldsymbol{\varphi}(\boldsymbol{x})\|$ as $\|\boldsymbol{x}\| \to \infty$. Considering $h\colon [0,1] \to \mathbb{R}$ defined by

$$h(t) = \log\big\|\boldsymbol{\varphi}\big(\boldsymbol{x}+t\boldsymbol{\theta}(\boldsymbol{x})\big)\big\|$$

and setting $\boldsymbol{\varphi}(\boldsymbol{x}) = \big(\varphi_1(\boldsymbol{x}), \ldots, \varphi_N(\boldsymbol{x})\big)$ we get

$$\tfrac{d}{dt}h(t) = \tfrac{\varphi_1(\boldsymbol{x}+t\boldsymbol{\theta}(\boldsymbol{x}))\,\boldsymbol{\theta}(\boldsymbol{x})\cdot D\varphi_1(\boldsymbol{x}+t\boldsymbol{\theta}(\boldsymbol{x}))+\ldots+\varphi_N(\boldsymbol{x}+t\boldsymbol{\theta}(\boldsymbol{x}))\,\boldsymbol{\theta}(\boldsymbol{x})\cdot D\varphi_N(\boldsymbol{x}+t\boldsymbol{\theta}(\boldsymbol{x}))}{\|\boldsymbol{\varphi}(\boldsymbol{x}+t\boldsymbol{\theta}(\boldsymbol{x}))\|^2}.$$

Hence,

$$\begin{aligned}\left|\frac{d}{dt}h(t)\right| &\leqslant \tfrac{|\varphi_1(\boldsymbol{x}+t\boldsymbol{\theta}(\boldsymbol{x}))|\,\|\boldsymbol{\theta}(\boldsymbol{x})\|\|D\varphi_1(\boldsymbol{x}+t\boldsymbol{\theta}(\boldsymbol{x}))\|+\ldots+|\varphi_N(\boldsymbol{x}+t\boldsymbol{\theta}(\boldsymbol{x}))|\,\|\boldsymbol{\theta}(\boldsymbol{x})\|\|D\varphi_N(\boldsymbol{x}+t\boldsymbol{\theta}(\boldsymbol{x}))\|}{\|\boldsymbol{\varphi}(\boldsymbol{x}+t\boldsymbol{\theta}(\boldsymbol{x}))\|^2} \\ &\leqslant N\,\|\boldsymbol{\theta}(\boldsymbol{x})\|\;\tfrac{\|\!|D\boldsymbol{\varphi}(\boldsymbol{x}+t\boldsymbol{\theta}(\boldsymbol{x}))|\!\|}{\|\boldsymbol{\varphi}(\boldsymbol{x}+t\boldsymbol{\theta}(\boldsymbol{x}))\|}\end{aligned}$$

[1] $X_n = o_{\Pr}(1)$ means $X_n \xrightarrow{\Pr} 0$ as $n \to \infty$.

and mean value theorem lead us to

$$\left|\log\left\|\boldsymbol{\varphi}\big(\boldsymbol{x}+\boldsymbol{\theta}(\boldsymbol{x})\big)\right\| - \log\|\boldsymbol{\varphi}(\boldsymbol{x})\|\right| = |h(1)-h(0)| = \\ = \left|\frac{dh}{dt}(c)\right| \leqslant \kappa\,\|\boldsymbol{\theta}(\boldsymbol{x})\|\,\frac{\left\|\left\|D\boldsymbol{\varphi}\big(\boldsymbol{x}+c\boldsymbol{\theta}(\boldsymbol{x})\big)\right\|\right\|}{\left\|\boldsymbol{\varphi}\big(\boldsymbol{x}+c\boldsymbol{\theta}(\boldsymbol{x})\big)\right\|}\,, \qquad 0<c<1$$

which implies $\left\|\boldsymbol{\varphi}\big(\boldsymbol{x}+\boldsymbol{\theta}(\boldsymbol{x})\big)\right\| \sim \|\boldsymbol{\varphi}(\boldsymbol{x})\|$ as $\|\boldsymbol{x}\| \to \infty$. □

Next, we present the main uniform approximation result.

Theorem 2.1. *Let $\boldsymbol{V}_n = (V_{1n}, ..., V_{Nn})$ be a sequence of random vectors of continuous type such that $\{V_{in},\, n \geqslant 1\}$ $(1 \leqslant i \leqslant N)$ is identically distributed and $\sup_{n\geqslant 1}\|\boldsymbol{V}_n\| \leqslant W$ for some r.v. W. If $X_n := g(\boldsymbol{V}_n + \boldsymbol{\mu}_n)$ where $\boldsymbol{\mu}_n$ is a non-random vector sequence verifying $\|\boldsymbol{\mu}_n\| \to \infty$ and g is a $C^2(\mathbb{R}^N)$ map such that $\dfrac{Dg(\boldsymbol{t})}{\|Dg(\boldsymbol{t})\|}$ exists as $\|\boldsymbol{t}\| \to \infty$ and $\left\|\left\|D^2 g(\boldsymbol{t})\right\|\right\| = o\big(\,\|Dg(\boldsymbol{t})\|\,\big)$, $\|\boldsymbol{t}\| \to \infty$ then, with $Y_n := g(\boldsymbol{\mu}_n) + Dg(\boldsymbol{\mu}_n)\cdot\boldsymbol{V}_n$, the law of X_n is uniformly approximate by the law of Y_n for large values of n, that is,*

$$\sup_{x\in\mathbb{R}} |F_{X_n}(x) - F_{Y_n}(x)| \to 0\ .$$

Proof: Using the Taylor formula for g we get

$$X_n := g(\boldsymbol{\mu}_n + \boldsymbol{V}_n) = g(\boldsymbol{\mu}_n) + Dg(\boldsymbol{\mu}_n)\cdot\boldsymbol{V}_n + \frac{1}{2}\,D^2 g(\boldsymbol{\mu}_n + \theta_n\boldsymbol{V}_n)\cdot\boldsymbol{V}_n^2 = \\ = Y_n + \frac{1}{2}\,D^2 g(\boldsymbol{\mu}_n + \theta_n\boldsymbol{V}_n)\cdot\boldsymbol{V}_n^2\,, \qquad 0<\theta_n<1$$

where $Dg(\boldsymbol{a})\cdot\boldsymbol{V}_n = \displaystyle\sum_i \frac{\partial g}{\partial x_i}(\boldsymbol{a})V_{in}$ and $D^2 g(\boldsymbol{a})\cdot\boldsymbol{V}_n^2 = \displaystyle\sum_{i,j}\frac{\partial^2 g}{\partial x_i \partial x_j}(\boldsymbol{a})V_{in}V_{jn}$ (see [11], page 150). For $\varepsilon > 0$ fixed we have

$$\Pr\big\{X_n \leqslant x\big\} = \Pr\left\{X_n \leqslant x, \frac{|X_n - Y_n|}{\|Dg(\boldsymbol{\mu}_n)\|} \leqslant \varepsilon\right\} + \Pr\left\{X_n \leqslant x, \frac{|X_n - Y_n|}{\|Dg(\boldsymbol{\mu}_n)\|} > \varepsilon\right\} \\ \leqslant \Pr\left\{Y_n \leqslant x + \varepsilon\,\|Dg(\boldsymbol{\mu}_n)\|\right\} + \Pr\left\{\frac{|X_n - Y_n|}{\|Dg(\boldsymbol{\mu}_n)\|} > \varepsilon\right\}$$

and

$$\Pr\left\{Y_n \leqslant x - \varepsilon\,\|Dg(\boldsymbol{\mu}_n)\|\right\} = \Pr\left\{Y_n \leqslant x - \varepsilon\,\|Dg(\boldsymbol{\mu}_n)\|\,, X_n \leqslant x\right\} + \\ + \Pr\left\{Y_n \leqslant x - \varepsilon\,\|Dg(\boldsymbol{\mu}_n)\|\,, X_n > x\right\} \\ \leqslant \Pr\big\{X_n \leqslant x\big\} + \Pr\left\{\frac{|X_n - Y_n|}{\|Dg(\boldsymbol{\mu}_n)\|} > \varepsilon\right\}$$

that is,

$$-\Pr\left\{\tfrac{|X_n - Y_n|}{\|Dg(\boldsymbol{\mu}_n)\|} > \varepsilon\right\} + F_{Dg(\boldsymbol{\mu}_n)\cdot\boldsymbol{V}_1}\left(x - g(\boldsymbol{\mu}_n) - \varepsilon\,\|Dg(\boldsymbol{\mu}_n)\|\right) \leqslant F_{X_n}(x) \leqslant \\ \leqslant F_{Dg(\boldsymbol{\mu}_n)\cdot\boldsymbol{V}_1}\left(x - g(\boldsymbol{\mu}_n) + \varepsilon\,\|Dg(\boldsymbol{\mu}_n)\|\right) + \Pr\left\{\tfrac{|X_n - Y_n|}{\|Dg(\boldsymbol{\mu}_n)\|} > \varepsilon\right\}.$$

We can rewrite the above inequalities as

$$-\Pr\left\{\frac{|X_n-Y_n|}{\|Dg(\boldsymbol{\mu}_n)\|}>\varepsilon\right\}+F_{\frac{Dg(\boldsymbol{\mu}_n)\cdot\boldsymbol{V}_1}{\|g(\boldsymbol{\mu}_n)\|}}\left(\frac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}-\varepsilon\right)\leqslant F_{X_n}(x)\leqslant$$
$$\leqslant F_{\frac{Dg(\boldsymbol{\mu}_n)\cdot\boldsymbol{V}_1}{\|Dg(\boldsymbol{\mu}_n)\|}}\left(\frac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}+\varepsilon\right)+\Pr\left\{\frac{|X_n-Y_n|}{\|Dg(\boldsymbol{\mu}_n)\|}>\varepsilon\right\}.$$

Therefore,

$$|F_{X_n}(x)-F_{Y_n}(x)|\leqslant\left|F_{\frac{Dg(\boldsymbol{\mu}_n)\cdot\boldsymbol{V}_1}{\|Dg(\boldsymbol{\mu}_n)\|}}\left(\frac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}+\varepsilon\right)-F_{\frac{Dg(\boldsymbol{\mu}_n)\cdot\boldsymbol{V}_1}{\|Dg(\boldsymbol{\mu}_n)\|}}\left(\frac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}\right)\right|+$$
$$+\left|F_{\frac{Dg(\boldsymbol{\mu}_n)\cdot\boldsymbol{V}_1}{\|Dg(\boldsymbol{\mu}_n)\|}}\left(\frac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}\right)-F_{\frac{Dg(\boldsymbol{\mu}_n)\cdot\boldsymbol{V}_1}{\|Dg(\boldsymbol{\mu}_n)\|}}\left(\frac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}-\varepsilon\right)\right|+\Pr\left\{\frac{|X_n-Y_n|}{\|Dg(\boldsymbol{\mu}_n)\|}>\varepsilon\right\}$$

and we obtain the following uniform bound,

(2.1)

$$\sup_{x\in\mathbb{R}}|F_{X_n}(x)-F_{Y_n}(x)|\leqslant\sup_{x\in\mathbb{R}}\left|F_{\frac{Dg(\boldsymbol{\mu}_n)\cdot\boldsymbol{V}_1}{\|Dg(\boldsymbol{\mu}_n)\|}}\left(\frac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}+\varepsilon\right)-F_{\frac{Dg(\boldsymbol{\mu}_n)\cdot\boldsymbol{V}_1}{\|Dg(\boldsymbol{\mu}_n)\|}}\left(\frac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}\right)\right|+$$
$$+\sup_{x\in\mathbb{R}}\left|F_{\frac{Dg(\boldsymbol{\mu}_n)\cdot\boldsymbol{V}_1}{\|Dg(\boldsymbol{\mu}_n)\|}}\left(\frac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}\right)-F_{\frac{Dg(\boldsymbol{\mu}_n)\cdot\boldsymbol{V}_1}{\|Dg(\boldsymbol{\mu}_n)\|}}\left(\frac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}-\varepsilon\right)\right|+\Pr\left\{\frac{|X_n-Y_n|}{\|Dg(\boldsymbol{\mu}_n)\|}>\varepsilon\right\}$$

Since

$$\frac{Dg(\boldsymbol{\mu}_n)\cdot\boldsymbol{V}_1}{\|Dg(\boldsymbol{\mu}_n)\|}\xrightarrow{\mathrm{d}}\boldsymbol{\tau}\cdot\boldsymbol{V}_1\qquad(\boldsymbol{\tau}\neq\mathbf{0})$$

and $\boldsymbol{\tau}\cdot\boldsymbol{V}_1=\sum_i\tau_iV_{i1}$ is continuous then Polya's theorem (see [3], page 3) states

$$F_{\frac{Dg(\boldsymbol{\mu}_n)\cdot\boldsymbol{V}_1}{\|Dg(\boldsymbol{\mu}_n)\|}}\longrightarrow F_{\boldsymbol{\tau}\cdot\boldsymbol{V}_1}\ \text{ uniformly on }\mathbb{R}.$$

Hence, we can still write

$$\sup_{x\in\mathbb{R}}\left|F_{\frac{Dg(\boldsymbol{\mu}_n)\cdot\boldsymbol{V}_1}{\|Dg(\boldsymbol{\mu}_n)\|}}\left(\frac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}+\varepsilon\right)-F_{\frac{Dg(\boldsymbol{\mu}_n)\cdot\boldsymbol{V}_1}{\|Dg(\boldsymbol{\mu}_n)\|}}\left(\frac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}\right)\right|\leqslant$$
$$\leqslant\sup_{x\in\mathbb{R}}\left|F_{\frac{Dg(\boldsymbol{\mu}_n)\cdot\boldsymbol{V}_1}{\|Dg(\boldsymbol{\mu}_n)\|}}\left(\frac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}+\varepsilon\right)-F_{\boldsymbol{\tau}\cdot\boldsymbol{V}_1}\left(\frac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}+\varepsilon\right)\right|+$$
$$+\sup_{x\in\mathbb{R}}\left|F_{\boldsymbol{\tau}\cdot\boldsymbol{V}_1}\left(\frac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}+\varepsilon\right)-F_{\boldsymbol{\tau}\cdot\boldsymbol{V}_1}\left(\frac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}\right)\right|+$$
$$+\sup_{x\in\mathbb{R}}\left|F_{\boldsymbol{\tau}\cdot\boldsymbol{V}_1}\left(\frac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}\right)-F_{\frac{Dg(\boldsymbol{\mu}_n)\cdot\boldsymbol{V}_1}{\|Dg(\boldsymbol{\mu}_n)\|}}\left(\frac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}\right)\right|$$

and

$$\sup_{x\in\mathbb{R}}\left|F_{\frac{Dg(\boldsymbol{\mu}_n)\cdot\boldsymbol{V}_1}{\|Dg(\boldsymbol{\mu}_n)\|}}\left(\frac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}\right)-F_{\frac{Dg(\boldsymbol{\mu}_n)\cdot\boldsymbol{V}_1}{\|Dg(\boldsymbol{\mu}_n)\|}}\left(\frac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}-\varepsilon\right)\right|\leqslant$$
$$\leqslant\sup_{x\in\mathbb{R}}\left|F_{\frac{Dg(\boldsymbol{\mu}_n)\cdot\boldsymbol{V}_1}{\|Dg(\boldsymbol{\mu}_n)\|}}\left(\frac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}\right)-F_{\boldsymbol{\tau}\cdot\boldsymbol{V}_1}\left(\frac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}\right)\right|+$$
$$+\sup_{x\in\mathbb{R}}\left|F_{\boldsymbol{\tau}\cdot\boldsymbol{V}_1}\left(\frac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}\right)-F_{\boldsymbol{\tau}\cdot\boldsymbol{V}_1}\left(\frac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}-\varepsilon\right)\right|+$$
$$+\sup_{x\in\mathbb{R}}\left|F_{\boldsymbol{\tau}\cdot\boldsymbol{V}_1}\left(\frac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}-\varepsilon\right)-F_{\frac{Dg(\boldsymbol{\mu}_n)\cdot\boldsymbol{V}_1}{\|Dg(\boldsymbol{\mu}_n)\|}}\left(\frac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}-\varepsilon\right)\right|.$$

Choosing $\varepsilon > 0$ small enough, we get for each $\eta > 0$

$$\sup_{x\in\mathbb{R}}\left|F_{\boldsymbol{\tau}\cdot\boldsymbol{V}_1}\left(\frac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}+\varepsilon\right)-F_{\boldsymbol{\tau}\cdot\boldsymbol{V}_1}\left(\frac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}\right)\right|<\eta$$

and

$$\sup_{x\in\mathbb{R}}\left|F_{\boldsymbol{\tau}\cdot\boldsymbol{V}_1}\left(\frac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}\right)-F_{\boldsymbol{\tau}\cdot\boldsymbol{V}_1}\left(\frac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}-\varepsilon\right)\right|<\eta$$

provided that $F_{\boldsymbol{\tau}\cdot\boldsymbol{V}_1}$ is uniformly continuous. Since $\sup\limits_{n\geqslant 1}\|\theta_n\boldsymbol{V}_n\|\leqslant W$ we obtain from Lemma 2.1,

$$\frac{|X_n-Y_n|}{\|Dg(\boldsymbol{\mu}_n)\|}\leqslant\frac{\|\boldsymbol{V}_n\|^2}{2}\,\frac{\left\|\left|D^2g(\boldsymbol{\mu}_n+\theta_n\boldsymbol{V}_n)\right|\right\|}{\|Dg(\boldsymbol{\mu}_n)\|}\leqslant\frac{W^2}{2}\,\frac{\left\|\left|D^2g(\boldsymbol{\mu}_n+\theta_n\boldsymbol{V}_n)\right|\right\|}{\|Dg(\boldsymbol{\mu}_n)\|}\xrightarrow{\text{a.s.}}0$$

so that for every $\varepsilon>0$, $\Pr\left\{\dfrac{|X_n-Y_n|}{\|Dg(\boldsymbol{\mu}_n)\|}>\varepsilon\right\}=o(1)$ as $n\to\infty$. Taking $\varepsilon>0$ small enough, Polya's theorem permit us to conclude that

$$\sup_{x\in\mathbb{R}}|F_{X_n}(x)-F_{Y_n}(x)|\to 0.$$ □

Remark 2.1. Let us note that, in Theorem 2.1, when $N=1$ the condition $\dfrac{Dg(\boldsymbol{t})}{\|Dg(\boldsymbol{t})\|}$ exists as $\|\boldsymbol{t}\|\to\infty$ can be dropped since in this case the uniform bound (2.1) takes the look

$$\begin{aligned}\sup_{x\in\mathbb{R}}|F_{X_n}(x)-F_{Y_n}(x)|\leqslant&\sup_{x\in\mathbb{R}}\left|F_{V_1}\left(\frac{x-g(\mu_n)}{\frac{dg}{dx_1}(\mu_n)}+\varepsilon\right)-F_{V_1}\left(\frac{x-g(\mu_n)}{\frac{dg}{dx_1}(\mu_n)}\right)\right|+\\&+\sup_{x\in\mathbb{R}}\left|F_{V_1}\left(\frac{x-g(\mu_n)}{\frac{dg}{dx_1}(\mu_n)}\right)-F_{V_1}\left(\frac{x-g(\mu_n)}{\frac{dg}{dx_1}(\mu_n)}-\varepsilon\right)\right|+\Pr\left\{\left|\frac{X_n-Y_n}{\frac{dg}{dx_1}(\mu_n)}\right|>\varepsilon\right\}\end{aligned}$$

and F_{V_1} is uniformly continuous.

When $N=1$ we can consider functions $g\colon\mathbb{R}\longrightarrow\mathbb{R}$ defined by $g(x)=x^r$ (*power behavior*) or more generally any polynomial $\sum_{k=0}^m a_kx^k$ with real coefficients a_k. Moreover, functions g such that $\frac{d}{dx}g(x)=\exp(x^r)$, $r<1$ (*exponencial behavior*) or $\frac{d}{dx}g(x)=\log^r\left(x^2+1\right)$ (*logarithmic behavior*) can be chosen broadening the important class of polynomials. In the multidimensional case, a remarkable example occurs when $\boldsymbol{V}_1$ has multivariate normal distribution with positive-definite variance-covariance matrix and g is a polynomial function in N variables $x_1,...,x_N$ with arbitrary real coefficients, that is,

$$g(x_1,...,x_N)=\sum_{k_1=0}^{m_1}...\sum_{k_N=0}^{m_N}a_{k_1...k_N}x_1^{k_1}...x_N^{k_N}.$$

3. APPLICATION OF THE UNIFORM APPROXIMATIONS TO THE DUAL STATIS METHOD

Without loss of generality, we will assume that the p-by-p symmetric positive definite matrix $\mathbf{Q}$ and the n_i-by-n_i diagonal matrix $\mathbf{D}_i$ introduced in the first section are, respectively, the identity matrix and the diagonal matrix $\operatorname{diag}\left(\frac{1}{n_i},...,\frac{1}{n_i}\right)$. Indeed, these assumptions can be made performing the linear transformation $\mathbf{X}_i$ to $\mathbf{Y}_i = \sqrt{n_i}\,\mathbf{D}_i^{1/2}\mathbf{X}_i\mathbf{T}'$ as a preliminary step. From (1.1) and (1.3) we have

$$\begin{aligned}\mathbf{W}_i &= \mathbf{X}_i'\mathbf{D}_i\mathbf{X}_i \\ &= (\mathbf{U}_i+\mathbf{E}_i)'\mathbf{D}_i(\mathbf{U}_i+\mathbf{E}_i) \\ &= \frac{1}{n_i}\mathbf{U}_i'\mathbf{U}_i + \frac{1}{n_i}\left[\mathbf{U}_i'\mathbf{E}_i + (\mathbf{U}_i'\mathbf{E}_i)'\right] + \frac{1}{n_i}\mathbf{E}_i'\mathbf{E}_i.\end{aligned}$$

Thus,

$$\begin{aligned}\mathbf{W}_i\mathbf{W}_i = {}& \frac{1}{n_i^2}(\mathbf{U}_i'\mathbf{U}_i)^2 + \frac{1}{n_i^2}(\mathbf{U}_i'\mathbf{U}_i)\left(\mathbf{U}_i'\mathbf{E}_i + (\mathbf{U}_i'\mathbf{E}_i)'\right) + \frac{1}{n_i^2}(\mathbf{U}_i'\mathbf{U}_i)(\mathbf{E}_i'\mathbf{E}_i) + \\ &+ \frac{1}{n_i^2}\left(\mathbf{U}_i'\mathbf{E}_i + (\mathbf{U}_i'\mathbf{E}_i)'\right)(\mathbf{U}_i'\mathbf{U}_i) + \frac{1}{n_i^2}\left(\mathbf{U}_i'\mathbf{E}_i + (\mathbf{U}_i'\mathbf{E}_i)'\right)^2 + \\ &+ \frac{1}{n_i^2}\left(\mathbf{U}_i'\mathbf{E}_i + (\mathbf{U}_i'\mathbf{E}_i)'\right)(\mathbf{E}_i'\mathbf{E}_i) + \frac{1}{n_i^2}(\mathbf{E}_i'\mathbf{E}_i)(\mathbf{U}_i'\mathbf{U}_i) + \\ &+ \frac{1}{n_i^2}(\mathbf{E}_i'\mathbf{E}_i)\left(\mathbf{U}_i'\mathbf{E}_i + (\mathbf{U}_i'\mathbf{E}_i)'\right) + \frac{1}{n_i^2}(\mathbf{E}_i'\mathbf{E}_i)^2.\end{aligned}$$

and

$$\begin{aligned}\operatorname{tr}(\mathbf{W}_i\mathbf{W}_i) = {}& \frac{1}{n_i^2}\operatorname{tr}\left((\mathbf{U}_i'\mathbf{U}_i)^2\right) + \frac{4}{n_i^2}\operatorname{tr}\left(\mathbf{U}_i'\mathbf{U}_i\mathbf{U}_i'\mathbf{E}_i\right) + \frac{2}{n_i^2}\operatorname{tr}\left((\mathbf{U}_i'\mathbf{U}_i)(\mathbf{E}_i'\mathbf{E}_i)\right) + \\ &+ \frac{4}{n_i^2}\operatorname{tr}\left(\mathbf{E}_i'\mathbf{E}_i\mathbf{U}_i'\mathbf{E}_i\right) + \frac{1}{n_i^2}\operatorname{tr}\left(\left(\mathbf{U}_i'\mathbf{E}_i + (\mathbf{U}_i'\mathbf{E}_i)'\right)^2\right) + \frac{1}{n_i^2}\operatorname{tr}\left((\mathbf{E}_i'\mathbf{E}_i)^2\right)\end{aligned}$$

provide that $\operatorname{tr}(\mathbf{A}) = \operatorname{tr}(\mathbf{A}')$ and $\operatorname{tr}(\mathbf{AB}) = \operatorname{tr}(\mathbf{BA})$ (see [5], page 50). Moreover,

$$\begin{aligned}\mathbf{W}_i\mathbf{W}_j = {}& \frac{1}{n_in_j}(\mathbf{U}_i'\mathbf{U}_i)(\mathbf{U}_j'\mathbf{U}_j) + \frac{1}{n_in_j}(\mathbf{U}_i'\mathbf{U}_i)\left(\mathbf{U}_j'\mathbf{E}_j + (\mathbf{U}_j'\mathbf{E}_j)'\right) + \\ &+ \frac{1}{n_in_j}(\mathbf{U}_i'\mathbf{U}_i)(\mathbf{E}_j'\mathbf{E}_j) + \frac{1}{n_in_j}\left(\mathbf{U}_i'\mathbf{E}_i + (\mathbf{U}_i'\mathbf{E}_i)'\right)(\mathbf{U}_j'\mathbf{U}_j) + \\ &+ \frac{1}{n_in_j}\left(\mathbf{U}_i'\mathbf{E}_i + (\mathbf{U}_i'\mathbf{E}_i)'\right)\left(\mathbf{U}_j'\mathbf{E}_j + (\mathbf{U}_j'\mathbf{E}_j)'\right) + \frac{1}{n_in_j}\left(\mathbf{U}_i'\mathbf{E}_i + (\mathbf{U}_i'\mathbf{E}_i)'\right)(\mathbf{E}_j'\mathbf{E}_j) + \\ &+ \frac{1}{n_in_j}(\mathbf{E}_i'\mathbf{E}_i)(\mathbf{U}_j'\mathbf{U}_j) + \frac{1}{n_in_j}(\mathbf{E}_i'\mathbf{E}_i)\left(\mathbf{U}_j'\mathbf{E}_j + (\mathbf{U}_j'\mathbf{E}_j)'\right) + \frac{1}{n_in_j}(\mathbf{E}_i'\mathbf{E}_i)(\mathbf{E}_j'\mathbf{E}_j).\end{aligned}$$

when $i \neq j$ and so

$$\begin{aligned}\operatorname{tr}(\mathbf{W}_i\mathbf{W}_j) &= \frac{1}{n_in_j}\operatorname{tr}\left(\mathbf{U}_i'\mathbf{U}_i\mathbf{U}_j'\mathbf{U}_j\right) + \frac{2}{n_in_j}\operatorname{tr}\left(\mathbf{U}_i'\mathbf{U}_i\mathbf{U}_j'\mathbf{E}_j\right) + \\ &+ \frac{1}{n_in_j}\operatorname{tr}\left(\mathbf{U}_i'\mathbf{U}_i\right)(\mathbf{E}_j'\mathbf{E}_j) + \frac{2}{n_in_j}\operatorname{tr}\left(\mathbf{U}_i'\mathbf{E}_i\mathbf{U}_j'\mathbf{U}_j\right) + \frac{2}{n_in_j}\operatorname{tr}\left(\mathbf{U}_i'\mathbf{E}_i\mathbf{U}_j'\mathbf{E}_j\right) + \\ &+ \frac{2}{n_in_j}\operatorname{tr}\left(\mathbf{U}_i'\mathbf{E}_i(\mathbf{U}_j'\mathbf{E}_j)'\right) + \frac{2}{n_in_j}\operatorname{tr}\left(\mathbf{U}_i'\mathbf{E}_i\mathbf{E}_j'\mathbf{E}_j\right) + \frac{1}{n_in_j}\operatorname{tr}\left(\mathbf{E}_i'\mathbf{E}_i\mathbf{U}_j'\mathbf{U}_j\right) + \\ &+ \frac{2}{n_in_j}\operatorname{tr}\left(\mathbf{E}_i'\mathbf{E}_i\mathbf{U}_j'\mathbf{E}_j\right) + \frac{1}{n_in_j}\operatorname{tr}\left(\mathbf{E}_i'\mathbf{E}_i\mathbf{E}_j'\mathbf{E}_j\right).\end{aligned}$$

We will apply now the theoretical results of Section 2 to obtain an uniform approximation for the entries of the matrix $\mathbf{S}$ considering $g_i\colon \mathbb{R}^{n_i\times p} \longrightarrow \mathbb{R}$ defined by

$$g_i(\mathbf{X}) = \frac{1}{n_i^2}\operatorname{tr}\left((\mathbf{X}'\mathbf{X})^2\right), \quad i = 1, ..., k.$$

Representing $\mathbf{i}_j$ the jth column of an identity matrix of unspecified dimensions we have

$$\begin{aligned}\frac{\partial g_i(\mathbf{X})}{\partial x_{\ell j}} &= \frac{1}{n_i^2}\operatorname{tr}\left(\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{i}_\ell\mathbf{i}_j' + \mathbf{i}_j\mathbf{i}_\ell'\mathbf{X}) + (\mathbf{X}'\mathbf{i}_\ell\mathbf{i}_j' + \mathbf{i}_j\mathbf{i}_\ell'\mathbf{X})\mathbf{X}'\mathbf{X}\right) \\ &= \frac{4}{n_i^2}\operatorname{tr}\left(\mathbf{X}'\mathbf{X}\mathbf{X}'\mathbf{i}_\ell\mathbf{i}_j'\right)\end{aligned}$$

$\ell = 1, ..., n_i$, $j = 1, ..., p$ (see [5], pages 299 and 300). Therefore,

$$g_i(\mathbf{U}_i) + Dg_i(\mathbf{U}_i)\cdot\mathbf{E}_i = \frac{1}{n_i^2}\operatorname{tr}\left((\mathbf{U}_i'\mathbf{U}_i)^2\right) + \frac{4}{n_i^2}\operatorname{tr}\left(\mathbf{U}_i'\mathbf{U}_i\mathbf{U}_i'\mathbf{E}_i\right)$$

and from Theorem 2.1 we obtain

$$\sup_{x\in\mathbb{R}}\left|F_{g_i(\mathbf{U}_i+\mathbf{E}_i)} - F_{g_i(\mathbf{U}_i)+Dg_i(\mathbf{U}_i)\cdot\mathbf{E}_i}\right| \to 0 \tag{3.1}$$

as $\|\operatorname{vec}(\mathbf{U}_i)\| \to \infty$, where $\operatorname{vec}(\mathbf{U}_i)$ is the vectorization of $\mathbf{U}_i$ (see [5], pages 339 and 340), since the entries of $\mathbf{E}_i$ are continuous r.v.'s. For the case $i \neq j$ consider $g_{i,j}\colon \mathbb{R}^{n_i\times p} \times \mathbb{R}^{n_j\times p} \longrightarrow \mathbb{R}$ defined by

$$g_{i,j}(\mathbf{X}, \mathbf{Y}) = \frac{1}{n_in_j}\operatorname{tr}\left(\mathbf{X}'\mathbf{X}\mathbf{Y}'\mathbf{Y}\right), \quad i, j = 1, ..., k, \quad i \neq j.$$

Thus,

$$\begin{aligned}g_{i,j}(\mathbf{U}_i, \mathbf{U}_j) + Dg_{i,j}(\mathbf{U}_i, \mathbf{U}_j)\cdot(\mathbf{E}_i, \mathbf{E}_j) &= \frac{1}{n_in_j}\operatorname{tr}\left(\mathbf{U}_i'\mathbf{U}_i\mathbf{U}_j'\mathbf{U}_j\right) + \\ &+ \frac{2}{n_in_j}\operatorname{tr}\left(\mathbf{U}_i'\mathbf{U}_i\mathbf{U}_j'\mathbf{E}_j\right) + \frac{2}{n_in_j}\operatorname{tr}\left(\mathbf{U}_j'\mathbf{U}_j\mathbf{U}_i'\mathbf{E}_i\right)\end{aligned}$$

and again, from Theorem 2.1 we have

$$\sup_{x\in\mathbb{R}}\left|F_{g_{i,j}(\mathbf{U}_i+\mathbf{E}_i,\mathbf{U}_j+\mathbf{E}_j)} - F_{g_{i,j}(\mathbf{U}_i,\mathbf{U}_j)+Dg_{i,j}(\mathbf{U}_i,\mathbf{U}_j)\cdot(\mathbf{E}_i,\mathbf{E}_j)}\right| \to 0 \tag{3.2}$$

as $\|\text{vec}(\mathbf{U}_i)\| \to \infty$ for each i. Therefore, considering $\overline{\mathbf{S}} = [\overline{s}_{ij}]_{i,j=1,...,k}$ where

$$\overline{s}_{ii} = g_i(\mathbf{U}_i) + Dg_i(\mathbf{U}_i) \cdot \mathbf{E}_i \quad \text{and} \quad \overline{s}_{ij} = g_{i,j}(\mathbf{U}_i, \mathbf{U}_j) + Dg_{i,j}(\mathbf{U}_i, \mathbf{U}_j) \cdot (\mathbf{E}_i, \mathbf{E}_j),$$

with $i \neq j$ for all $i, j = 1, ..., k$, we can state from (3.1) and (3.2) that the distribution of each entry of $\mathbf{S}$ can be uniformly approximated by the distribution of the same entry of $\overline{\mathbf{S}}$ when $\|\text{vec}(\mathbf{U}_i)\| \to \infty$ for each i.

The above exposure can be summarized in the following result

Theorem 3.1. *If $\mathbf{X}_i$, $i = 1, ..., k$ are n_i-by-p random data tables such that $\mathbf{X}_i = \mathbf{U}_i + \mathbf{E}_i$, where $\mathbf{E}_i$ are independent n_i-by-p random matrices having i.i.d. continuous entries and $\mathbf{U}_i$ are non-random n_i-by-p matrices then, for each $i, j = 1, ..., k$, the distribution of*

$$s_{ij} = \frac{1}{n_i n_j} \text{tr}\left[(\mathbf{U}_i + \mathbf{E}_i)'(\mathbf{U}_i + \mathbf{E}_i)(\mathbf{U}_j + \mathbf{E}_j)'(\mathbf{U}_j + \mathbf{E}_j)\right]$$

is uniformly approximated by the distribution of

$$\overline{s}_{ij} = \frac{1}{n_i n_j} \text{tr}\left(\mathbf{U}_i'\mathbf{U}_i\mathbf{U}_j'\mathbf{U}_j\right) + \frac{2}{n_i n_j} \text{tr}\left(\mathbf{U}_i'\mathbf{U}_i\mathbf{U}_j'\mathbf{E}_j\right) + \frac{2}{n_i n_j} \text{tr}\left(\mathbf{U}_j'\mathbf{U}_j\mathbf{U}_i'\mathbf{E}_i\right),$$

as $\|\text{vec}(\mathbf{U}_i)\| \to \infty$ for each i.

Remark 3.1. Observe that the condition $\|\text{vec}(\mathbf{U}_i)\| \to \infty$ for each i are related with the smallness coefficient of variation and it is a verifiable assumption in some scenarios.

3.1. Estimating the eigenvalue and the eigenvector of E($\mathbf{S}$)

Recovering (1.5) we can write synthetically

$$\overline{\mathbf{S}} \overset{\text{d}}{\approx} \mathbf{S} = \lambda\boldsymbol{\alpha}\boldsymbol{\alpha}' + \boldsymbol{\mathcal{E}} \tag{3.3}$$

when $\|\text{vec}(\mathbf{U}_i)\|$ is large enough for all i (i.e. the distribution of each entry of $\boldsymbol{\mathcal{E}}$ can be computable approximately). From (1.2) and (1.4), one estimates λ by the eigenvalue ρ and one estimates $\boldsymbol{\alpha}$ by the eigenvector $\mathbf{v}$, that is, we will consider the following estimators

$$\widehat{\lambda} = \rho \quad \text{and} \quad \widehat{\boldsymbol{\alpha}} = \mathbf{v}$$

of λ and $\boldsymbol{\alpha}$, respectively. The choice of $\widehat{\lambda}$ and $\widehat{\boldsymbol{\alpha}}$ as estimators of λ and $\boldsymbol{\alpha}$, respectively, arises in a very natural way (the same estimation method of eigenvalues and eigenvectors was already used in Anderson (1963)). On the other hand, the symmetry of $\mathbf{S}$ implies

$$\widehat{\boldsymbol{\beta}} = \mathbf{S}\mathbf{v} = \left(\mathbf{I} \otimes \mathbf{v}'\right) \text{vec}(\mathbf{S})$$

where $\otimes$ denotes the Kronecker product (see [5], page 333) and vec($\mathbf{S}$) is the vectorization of $\mathbf{S}$. Using (3.3) we can compute approximately the distribution of $\widehat{\boldsymbol{\beta}}$. Indeed, we can consider the approximated estimator $\widetilde{\boldsymbol{\beta}} = (\mathbf{I} \otimes \boldsymbol{\alpha}') \operatorname{vec}(\overline{\mathbf{S}})$ instead of $\widehat{\boldsymbol{\beta}}$, since its distribution is always determinable with the aid of the errors distribution.

3.2. Example with i.i.d. normal errors

Let us consider the entries of $\mathbf{E}_i$ i.i.d. normal distributed with zero mean and variance σ^2. Given an non-random q-by-n_i matrix $\mathbf{M}$, it is well-known that the trace $\operatorname{tr}(\mathbf{M}\mathbf{E}_i)$ is distributed normally with zero mean and variance $\sigma^2 \operatorname{tr}(\mathbf{M}\mathbf{M}')$. Therefore, $\overline{s}_{ii}$ is distributed normally with mean $\frac{\operatorname{tr}\left((\mathbf{U}_i'\mathbf{U}_i)^2\right)}{n_i^2}$ and variance $\frac{16\sigma^2 \operatorname{tr}\left((\mathbf{U}_i'\mathbf{U}_i)^3\right)}{n_i^4}$. Moreover, $\overline{s}_{ij}$ is distributed normally with mean $\frac{\operatorname{tr}(\mathbf{U}_i'\mathbf{U}_i\mathbf{U}_j'\mathbf{U}_j)}{n_i n_j}$ and variance $\frac{4\sigma^2 \operatorname{tr}\left((\mathbf{U}_i'\mathbf{U}_i)^2\mathbf{U}_j'\mathbf{U}_j + (\mathbf{U}_j'\mathbf{U}_j)^2\mathbf{U}_i'\mathbf{U}_i\right)}{n_i^2 n_j^2}$ for all $i, j = 1, ..., k$ with $i \neq j$. Using the covariance properties we also get

$$\operatorname{Cov}(\overline{s}_{ii}, \overline{s}_{ii}) = \operatorname{Cov}\left(\tfrac{4}{n_i^2}\operatorname{tr}(\mathbf{U}_i'\mathbf{U}_i\mathbf{U}_i'\mathbf{E}_i), \tfrac{4}{n_i^2}\operatorname{tr}(\mathbf{U}_i'\mathbf{U}_i\mathbf{U}_i'\mathbf{E}_i)\right) = \tfrac{16\sigma^2}{n_i^4}\operatorname{tr}\left((\mathbf{U}_i'\mathbf{U}_i)^3\right)$$

and for $i \neq j$,

$$\operatorname{Cov}(\overline{s}_{ii}, \overline{s}_{jj}) = \operatorname{Cov}\left(\tfrac{4}{n_i^2}\operatorname{tr}(\mathbf{U}_i'\mathbf{U}_i\mathbf{U}_i'\mathbf{E}_i), \tfrac{4}{n_j^2}\operatorname{tr}(\mathbf{U}_j'\mathbf{U}_j\mathbf{U}_j'\mathbf{E}_j)\right) = 0$$

$$\begin{aligned}\operatorname{Cov}(\overline{s}_{ii}, \overline{s}_{ij}) &= \\ &= \operatorname{Cov}\left(\tfrac{4}{n_i^2}\operatorname{tr}\left(\mathbf{U}_i'\mathbf{U}_i\mathbf{U}_i'\mathbf{E}_i\right), \tfrac{2}{n_i n_j}\operatorname{tr}\left(\mathbf{U}_i'\mathbf{U}_i\mathbf{U}_j'\mathbf{E}_j\right) + \tfrac{2}{n_i n_j}\operatorname{tr}\left(\mathbf{U}_j'\mathbf{U}_j\mathbf{U}_i'\mathbf{E}_i\right)\right) \\ &= \tfrac{8\sigma^2}{n_i^3 n_j}\operatorname{tr}\left(\left(\mathbf{U}_i'\mathbf{U}_i\right)^2\mathbf{U}_j'\mathbf{U}_j\right)\end{aligned}$$

$$\begin{aligned}\operatorname{Cov}(\overline{s}_{ij}, \overline{s}_{ij}) = \operatorname{Cov}\Big(&\tfrac{2}{n_i n_j}\operatorname{tr}\left(\mathbf{U}_i'\mathbf{U}_i\mathbf{U}_j'\mathbf{E}_j\right) + \tfrac{2}{n_i n_j}\operatorname{tr}\left(\mathbf{U}_j'\mathbf{U}_j\mathbf{U}_i'\mathbf{E}_i\right), \\ &\tfrac{2}{n_i n_j}\operatorname{tr}\left(\mathbf{U}_i'\mathbf{U}_i\mathbf{U}_j'\mathbf{E}_j\right) + \tfrac{2}{n_i n_j}\operatorname{tr}\left(\mathbf{U}_j'\mathbf{U}_j\mathbf{U}_i'\mathbf{E}_i\right)\Big) \\ = \tfrac{4\sigma^2}{n_i^2 n_j^2}&\operatorname{tr}\left((\mathbf{U}_i'\mathbf{U}_i)^2\mathbf{U}_j'\mathbf{U}_j\right) + \tfrac{4\sigma^2}{n_i^2 n_j^2}\operatorname{tr}\left((\mathbf{U}_j'\mathbf{U}_j)^2\mathbf{U}_i'\mathbf{U}_i\right).\end{aligned}$$

For all different i, j, ℓ, q we still have

$$\begin{aligned}\operatorname{Cov}(\overline{s}_{ii}, \overline{s}_{j\ell}) = \operatorname{Cov}\Big(&\tfrac{4}{n_i^2}\operatorname{tr}(\mathbf{U}_i'\mathbf{U}_i\mathbf{U}_i'\mathbf{E}_i), \\ &\tfrac{2}{n_j n_\ell}\operatorname{tr}(\mathbf{U}_j'\mathbf{U}_j\mathbf{U}_\ell'\mathbf{E}_\ell) + \tfrac{2}{n_j n_\ell}\operatorname{tr}(\mathbf{U}_\ell'\mathbf{U}_\ell\mathbf{U}_j'\mathbf{E}_j)\Big) \\ = 0\end{aligned}$$

$$\begin{aligned}
\operatorname{Cov}(\overline{s}_{ij},\overline{s}_{i\ell}) &= \operatorname{Cov}\Big(\tfrac{2}{n_in_j}\operatorname{tr}(\mathbf{U}_i'\mathbf{U}_i\mathbf{U}_j'\mathbf{E}_j)+\tfrac{2}{n_in_j}\operatorname{tr}(\mathbf{U}_j'\mathbf{U}_j\mathbf{U}_i'\mathbf{E}_i),\\
&\qquad\qquad \tfrac{2}{n_in_\ell}\operatorname{tr}(\mathbf{U}_i'\mathbf{U}_i\mathbf{U}_\ell'\mathbf{E}_\ell)+\tfrac{2}{n_in_\ell}\operatorname{tr}(\mathbf{U}_\ell'\mathbf{U}_\ell\mathbf{U}_i'\mathbf{E}_i)\Big)\\
&= \tfrac{4\sigma^2}{n_i^2n_jn_\ell}\operatorname{tr}\Big(\mathbf{U}_j'\mathbf{U}_j\mathbf{U}_i'\mathbf{U}_i\mathbf{U}_\ell'\mathbf{U}_\ell\Big)\\
\operatorname{Cov}(\overline{s}_{ij},\overline{s}_{q\ell}) &= \operatorname{Cov}\Big(\tfrac{2}{n_in_j}\operatorname{tr}(\mathbf{U}_i'\mathbf{U}_i\mathbf{U}_j'\mathbf{E}_j)+\tfrac{2}{n_in_j}\operatorname{tr}(\mathbf{U}_j'\mathbf{U}_j\mathbf{U}_i'\mathbf{E}_i),\\
&\qquad\qquad \tfrac{2}{n_qn_\ell}\operatorname{tr}(\mathbf{U}_q'\mathbf{U}_q\mathbf{U}_\ell'\mathbf{E}_\ell)+\tfrac{2}{n_qn_\ell}\operatorname{tr}(\mathbf{U}_\ell'\mathbf{U}_\ell\mathbf{U}_q'\mathbf{E}_q)\Big)\\
&= 0
\end{aligned}$$

The next table resumes all covariance computations:

Table 1: Covariance between elements of $\overline{\mathbf{S}}$.

Elements	Covariance	# of elements
$\operatorname{Cov}(\overline{s}_{ii},\overline{s}_{ii})$	$\frac{16\,\sigma^2}{n_i^4}\operatorname{tr}\Big(\big(\mathbf{U}_i'\mathbf{U}_i\big)^3\Big)$	k
$\operatorname{Cov}(\overline{s}_{ii},\overline{s}_{ij})$	$\frac{8\,\sigma^2}{n_i^3n_j}\operatorname{tr}\Big(\big(\mathbf{U}_i'\mathbf{U}_i\big)^2\mathbf{U}_j'\mathbf{U}_j\Big)$	$k\,(k-1)$
$\operatorname{Cov}(\overline{s}_{ii},\overline{s}_{ji})$	$\frac{8\,\sigma^2}{n_i^3n_j}\operatorname{tr}\Big(\big(\mathbf{U}_i'\mathbf{U}_i\big)^2\mathbf{U}_j'\mathbf{U}_j\Big)$	$k\,(k-1)$
$\operatorname{Cov}(\overline{s}_{ij},\overline{s}_{ii})$	$\frac{8\,\sigma^2}{n_i^3n_j}\operatorname{tr}\Big(\big(\mathbf{U}_i'\mathbf{U}_i\big)^2\mathbf{U}_j'\mathbf{U}_j\Big)$	$k\,(k-1)$
$\operatorname{Cov}(\overline{s}_{ji},\overline{s}_{ii})$	$\frac{8\,\sigma^2}{n_i^3n_j}\operatorname{tr}\Big(\big(\mathbf{U}_i'\mathbf{U}_i\big)^2\mathbf{U}_j'\mathbf{U}_j\Big)$	$k\,(k-1)$
$\operatorname{Cov}(\overline{s}_{ii},\overline{s}_{jj})$	0	$k\,(k-1)$
$\operatorname{Cov}(\overline{s}_{ii},\overline{s}_{j\ell})$	0	$k\,(k-1)\,(k-2)$
$\operatorname{Cov}(\overline{s}_{ij},\overline{s}_{\ell\ell})$	0	$k\,(k-1)\,(k-2)$
$\operatorname{Cov}(\overline{s}_{ij},\overline{s}_{ij})$	$\frac{4\,\sigma^2}{n_i^2n_j^2}\operatorname{tr}\Big(\big(\mathbf{U}_i'\mathbf{U}_i\big)^2\mathbf{U}_j'\mathbf{U}_j\Big)+\frac{4\,\sigma^2}{n_i^2n_j^2}\operatorname{tr}\Big(\big(\mathbf{U}_j'\mathbf{U}_j\big)^2\mathbf{U}_i'\mathbf{U}_i\Big)$	$k\,(k-1)$
$\operatorname{Cov}(\overline{s}_{ij},\overline{s}_{ji})$	$\frac{4\,\sigma^2}{n_i^2n_j^2}\operatorname{tr}\Big(\big(\mathbf{U}_j'\mathbf{U}_j\big)^2\mathbf{U}_i'\mathbf{U}_i\Big)+\frac{4\,\sigma^2}{n_i^2n_j^2}\operatorname{tr}\Big(\big(\mathbf{U}_j'\mathbf{U}_j\big)^2\mathbf{U}_i'\mathbf{U}_i\Big)$	$k\,(k-1)$
$\operatorname{Cov}(\overline{s}_{ij},\overline{s}_{i\ell})$	$\frac{4\,\sigma^2}{n_i^2n_jn_\ell}\operatorname{tr}\Big(\mathbf{U}_j'\mathbf{U}_j\mathbf{U}_i'\mathbf{U}_i\mathbf{U}_\ell'\mathbf{U}_\ell\Big)$	$k\,(k-1)\,(k-2)$
$\operatorname{Cov}(\overline{s}_{ij},\overline{s}_{j\ell})$	$\frac{4\,\sigma^2}{n_j^2n_in_\ell}\operatorname{tr}\Big(\mathbf{U}_i'\mathbf{U}_i\mathbf{U}_j'\mathbf{U}_j\mathbf{U}_\ell'\mathbf{U}_\ell\Big)$	$k\,(k-1)\,(k-2)$
$\operatorname{Cov}(\overline{s}_{ij},\overline{s}_{\ell i})$	$\frac{4\,\sigma^2}{n_i^2n_jn_\ell}\operatorname{tr}\Big(\mathbf{U}_i'\mathbf{U}_i\mathbf{U}_j'\mathbf{U}_j\mathbf{U}_\ell'\mathbf{U}_\ell\Big)$	$k\,(k-1)\,(k-2)$
$\operatorname{Cov}(\overline{s}_{ij},\overline{s}_{\ell j})$	$\frac{4\,\sigma^2}{n_j^2n_in_\ell}\operatorname{tr}\Big(\mathbf{U}_i'\mathbf{U}_i\mathbf{U}_j'\mathbf{U}_j\mathbf{U}_\ell'\mathbf{U}_\ell\Big)$	$k\,(k-1)\,(k-2)$
$\operatorname{Cov}(\overline{s}_{ij},\overline{s}_{q\ell})$	0	$k\,(k-1)\,(k-2)\,(k-3)$

Generally, if $\mathbf{\Sigma}_i^+$ is the Moore–Penrose inverse (see [5], page 493) of the covariance matrix $\mathbf{\Sigma}_i$ of $\operatorname{vec}(\mathbf{U}_i+\mathbf{E}_i)$ (the vectorization of $\mathbf{U}_i+\mathbf{E}_i$) then the quadratic form $\operatorname{vec}(\mathbf{U}_i+\mathbf{E}_i)'\mathbf{\Sigma}_i^+\operatorname{vec}(\mathbf{U}_i+\mathbf{E}_i)$ has chi-square distribution with $r_i=\operatorname{rank}(\mathbf{\Sigma}_i)$ degrees of freedom and non-centrality parameter $\delta_i=\mathbf{U}_i'\mathbf{\Sigma}_i^+\mathbf{U}_i$ (see [12], page 182). Hence, we may use δ_i to measure the non-centrality of the sample.

In this case, the covariance matrix of $\mathrm{vec}(\mathbf{E}_i)$ is known, so that the assumption $\|\mathrm{vec}(\mathbf{U}_i)\| \to \infty$ for each i corresponds to consider observations with low variation coefficients which implies a large δ_i.

Since $\mathbf{S} = \lambda\boldsymbol{\alpha}\boldsymbol{\alpha}' + \boldsymbol{\mathcal{E}}$ with $\mathrm{E}(\boldsymbol{\mathcal{E}}) = \mathbf{O}$ we get

$$\mathrm{E}\left(\left(\mathbf{I}\otimes\boldsymbol{\alpha}'\right)\mathrm{vec}(\mathbf{S})\right) = \left(\mathbf{I}\otimes\boldsymbol{\alpha}'\right)\mathrm{vec}(\lambda\boldsymbol{\alpha}\boldsymbol{\alpha}') = \lambda\boldsymbol{\alpha} = \boldsymbol{\beta}$$

that is, $\widehat{\boldsymbol{\beta}}$ is approximately unbiased. The covariance matrix of $\widehat{\boldsymbol{\beta}}$ can also be computed (approximately) through

$$\left(\mathbf{I}\otimes\boldsymbol{\alpha}'\right)\boldsymbol{\Sigma}_{\mathrm{vec}(\overline{\mathbf{S}})}\left(\mathbf{I}\otimes\boldsymbol{\alpha}\right) \tag{3.4}$$

where $\boldsymbol{\Sigma}_{\mathrm{vec}(\overline{\mathbf{S}})}$ is the covariance matrix of $\overline{\mathbf{S}}$ (i.e. the elements of the previous table). Hence, $\widehat{\boldsymbol{\beta}}$ will have (approximately) normal distribution with mean value $\boldsymbol{\beta}$ and covariance matrix $\mathbf{C} = \left(\mathbf{I}\otimes\boldsymbol{\alpha}'\right)\boldsymbol{\Sigma}_{\mathrm{vec}(\overline{\mathbf{S}})}\left(\mathbf{I}\otimes\boldsymbol{\alpha}\right)$.

Remark 3.2. Relation (3.4) lead us even to consider

$$\widehat{\boldsymbol{\Sigma}} = \left(\mathbf{I}\otimes\mathbf{v}'\right)\boldsymbol{\Sigma}_{\mathrm{vec}(\overline{\mathbf{S}})}\left(\mathbf{I}\otimes\mathbf{v}\right)$$

as an estimator of the covariance matrix of $\widehat{\boldsymbol{\beta}}$.

Now, we will construct a non-random k^2-by-k^2 matrix $\mathbf{G}$ ($k > 1$) such that the random vector $\mathbf{y} = \mathbf{G}\mathrm{vec}(\mathbf{S})$ be independent of $\widehat{\boldsymbol{\beta}}$ and

$$\mathrm{E}\big(\mathbf{G}\mathrm{vec}(\overline{\mathbf{S}})\big) = \mathbf{0}, \quad \mathbf{G}\boldsymbol{\Sigma}_{\mathrm{vec}(\overline{\mathbf{S}})}\mathbf{G}' = \begin{bmatrix} \sigma^2\mathbf{I} & \mathbf{O}_{12} \\ \mathbf{O}_{21} & \mathbf{O}_{22} \end{bmatrix}$$

where $\mathbf{I}$ is the identity matrix of some size less than or equal to k^2 and $\mathbf{O}_{12}$, $\mathbf{O}_{21}$, $\mathbf{O}_{22}$ are matrices with zero elements. First, we can obtain an k^2-by-k^2 non-random matrix $\mathbf{B}$ such that

$$\mathbf{B}\,\mathrm{E}\big(\mathrm{vec}(\overline{\mathbf{S}})\big) = \mathbf{0} \quad \text{and} \quad \mathbf{B}\boldsymbol{\Sigma}_{\mathrm{vec}(\overline{\mathbf{S}})}(\mathbf{I}\otimes\boldsymbol{\alpha}) = \mathbf{O}$$

taking, for instance,

$$\mathbf{B} = \mathbf{R}\left(\mathbf{I} - \left[\,\boldsymbol{\Sigma}_{\mathrm{vec}(\overline{\mathbf{S}})}(\mathbf{I}\otimes\boldsymbol{\alpha})\;\; \mathrm{E}\big(\mathrm{vec}(\overline{\mathbf{S}})\big)\,\right]\left[\,\boldsymbol{\Sigma}_{\mathrm{vec}(\overline{\mathbf{S}})}(\mathbf{I}\otimes\boldsymbol{\alpha})\;\; \mathrm{E}\big(\mathrm{vec}(\overline{\mathbf{S}})\big)\,\right]^{+}\right)$$

where $\mathbf{R}$ is an arbitrary (conformatable) matrix (see [1], page 295). Hence, $\mathbf{B}\mathrm{vec}(\overline{\mathbf{S}})$ will have multivariate normal distribution with mean vector $\mathbf{B}\,\mathrm{E}\big(\mathrm{vec}(\overline{\mathbf{S}})\big)$ and covariance matrix $\mathbf{B}\boldsymbol{\Sigma}_{\mathrm{vec}(\overline{\mathbf{S}})}\mathbf{B}'$ (see [14], page 32). Since $\mathbf{B}\boldsymbol{\Sigma}_{\mathrm{vec}(\overline{\mathbf{S}})}\mathbf{B}'$ is an symmetric matrix with rank r such that $\mathbf{B}\boldsymbol{\Sigma}_{\mathrm{vec}(\overline{\mathbf{S}})}\mathbf{B}'$ is either positive definite ($r = k^2$) or positive semidefinite ($r < k^2$) then:

(**i**) If $r = k^2$ then there exists a nonsingular k^2-by-k^2 matrix $\mathbf{H}$ such that $\mathbf{HB}\boldsymbol{\Sigma}_{\mathrm{vec}(\overline{\mathbf{S}})}\mathbf{B}'\mathbf{H}' = \mathbf{I}$.

(**ii**) If $r < k^2$ then there exists a nonsingular r-by-r matrix $\mathbf{H}$ such that

$$\mathbf{HB}\boldsymbol{\Sigma}_{\mathrm{vec}(\overline{\mathbf{S}})}\mathbf{B}'\mathbf{H}' = \begin{bmatrix} \mathbf{I} & \mathbf{O}_{12} \\ \mathbf{O}_{21} & \mathbf{O}_{22} \end{bmatrix}$$

where $\mathbf{O}_{12}$, $\mathbf{O}_{21}$ and $\mathbf{O}_{22}$ are r-by-(k^2-r), (k^2-r)-by-r and (k^2-r)-by-(k^2-r) matrices with zero elements, respectively (see [14], page 27). Therefore, we can take

$$\mathbf{G} = \sigma\mathbf{HB}$$

and the components of $\mathbf{y}$ will be i.i.d. normal distributed with zero mean and variance σ^2.

For testing the assumption $\mathrm{rank}\big(\mathrm{E}(\mathbf{S})\big) = 1$ we can use the following statistical test:

$$\mathcal{F} = \frac{r\sigma^2\widehat{\boldsymbol{\beta}}'\widehat{\boldsymbol{\Sigma}}^{+}\widehat{\boldsymbol{\beta}}}{\nu\,\|\mathbf{y}\|^2}$$

where ν is the rank of $\widehat{\boldsymbol{\Sigma}}$.

Remark 3.3. Let us stand out that $\widehat{\boldsymbol{\Sigma}}^{+} = \dfrac{1}{\sigma^2}\widetilde{\boldsymbol{\Sigma}}^{+}$ where $\widetilde{\boldsymbol{\Sigma}}^{+}$ is independent of σ and so $\mathcal{F}$ do not depend on σ.

The statistical test $\mathcal{F}$ will have (approximately) F-distribution with parameters ν, r and non-centrality parameter $\delta = \boldsymbol{\beta}'\widehat{\boldsymbol{\Sigma}}^{+}\boldsymbol{\beta}$ (see [6], page 609). Since $\mathrm{rank}\big(\mathrm{E}(\mathbf{S})\big) = 1$ if $\delta > 0$, we will use $\mathcal{F}$ to test the null hypothesis $H_0 : \delta = 0$ against $H_1 : \delta > 0$ and the p-value of this statistical test of hypothesis can be computed by

$$p\text{-value} = \Pr(\mathcal{F} > \mathcal{F}_{\mathrm{obs}} | \delta = 0)$$

where $\mathcal{F}_{\mathrm{obs}}$ is the observed value.

After validate the model, we are able to use statistical hypothesis tests on the components of the vector $\boldsymbol{\beta}$. Given a ℓ-by-k non-random matrix $\mathbf{Z}$ and $\boldsymbol{\psi} = \mathbf{Z}\boldsymbol{\beta}$, we can test the hypothesis $H_0 : \boldsymbol{\psi} = \mathbf{z}$, where $\mathbf{z}$ is a non-random vector. Considering the estimator $\widehat{\boldsymbol{\psi}} = \mathbf{Z}\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\psi}$ then $\widehat{\boldsymbol{\psi}}$ will have (approximately) normal distribution with mean value $\boldsymbol{\psi} = \mathbf{Z}\boldsymbol{\beta}$ and covariance matrix $\mathbf{Z}\widehat{\boldsymbol{\Sigma}}\mathbf{Z}'$ (see [14], page 32). Moreover, $\widehat{\boldsymbol{\psi}}$ will be also independent from $\mathbf{y}$ which lead us to use the following statistical test:

$$\mathcal{F} = \frac{r\sigma^2(\widehat{\boldsymbol{\psi}} - \mathbf{z})'\left(\mathbf{Z}\widehat{\boldsymbol{\Sigma}}\mathbf{Z}'\right)^{+}(\widehat{\boldsymbol{\psi}} - \mathbf{z})}{\nu\,\|\mathbf{y}\|^2}$$

where ν is the rank of $\mathbf{Z}\widehat{\mathbf{\Sigma}}\mathbf{Z}'$. Again, $\mathcal{F}$ do not depend on σ (see Remark 4). If H_0 is accepted then $\mathcal{F}$ will have (approximately) F-distribution with parameters ν, r and noncentrality parameter

$$\delta = (\boldsymbol{\psi} - \mathbf{z})' \left(\mathbf{Z}\widehat{\mathbf{\Sigma}}\mathbf{Z}'\right)^{+} (\boldsymbol{\psi} - \mathbf{z}).$$

If $\mathbf{Z}\widehat{\mathbf{\Sigma}}\mathbf{Z}'$ was invertible then $\delta = 0$ is equivalent to the acceptance of H_0. Hence, the p-value of this statistical test of hypothesis is given by

$$p - \text{value} = \Pr(\mathcal{F} > \mathcal{F}_{\text{obs}} | H_0 \text{ is true})$$

where $\mathcal{F}_{\text{obs}}$ is the observed value.

Remark 3.4. Observe that the above statistical test of hypothesis is a generalization of the first one, in the sense that we can use it with $\mathbf{Z} = \mathbf{I}$ and $\mathbf{z} = \mathbf{0}$ to test the assumption $\operatorname{rank}\big(\mathrm{E}(\mathbf{S})\big) = 1$.

Remark 3.5. Furthermore, choosing the matrix $\mathbf{Z}$ and the vector $\mathbf{z}$ appropriately we can perform statistical tests of hypothesis to compare two or more components of $\boldsymbol{\beta} = (\beta_1, ..., \beta_k)$, for instance, $H_0 \colon \beta_i = \beta_j$ against $H_1 \colon \beta_i \neq \beta_j$ (i.e. $\mathbf{z} = \mathbf{0}$ and $\mathbf{Z} = \left[z_{ij}\right]_{\substack{i=1,...,\ell \\ j=1,...,k}}$ defined by $z_{1i} = -z_{1j} = 1$ with all remaining entries being zero). Note also that the statistical test $H_0 \colon \beta_i = \beta_j$ against $H_1 \colon \beta_i \neq \beta_j$ $(i \neq j)$ is equivalent to compare two different components of $\boldsymbol{\alpha}$ provided that $\lambda \neq 0$.

4. CONCLUSIONS

The theoretical results of Section 2 arose to get the solution to the following problem: when there is no limiting distribution for a sequence of r.v.'s X_n can we approximate the limit distribution of $g(X_n)$ for large values of n and some fixed function g? The well-known *delta method* cannot be used to give an answer to this question since there is no limiting distribution for X_n. In Theorem 2.1 we partially answer to the above question considering $X_n = V_n + \mu_n$ and giving sufficient conditions on g to obtain a sequence of random variables which are of the same type of $g(X_n)$ for large values of n. Let us observe also that our result allows us to get "normalizing constants" for $g(X_n)$ when n is large enough.

Therefore, besides the uniform approximation of the distribution function sequence $F_{g(\mathbf{V}_n + \boldsymbol{\mu}_n)}$ by a computable one, this work develops inference results on the components of the vector $\lambda\boldsymbol{\alpha}$, where λ and $\boldsymbol{\alpha}$ are, respectively, the eigenvalue and the eigenvector of the rank one matrix $\mathrm{E}(\mathbf{S})$, with $\mathbf{S}$ the interstructure matrix used in dual STATIS method admitting data tables of the form (1.3). So, our

results appears in the same alignment of Lazraq (see [10]) which considered an inferencial approach for the validation of the compromise matrix obtained by the STATIS method.

In our scenario of data tables, remains an open problem the generalization of the presented inferential results to the case where the rank of E(**S**) is greater than one.

ACKNOWLEDGMENTS

This work was partially supported by the Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) through PEst-OE/MAT/UI0297/2011 (CMA).

REFERENCES

[1] Abadir, K.M. and Magnus, J.R. (2005). *Matrix Algebra*, Cambridge University Press.

[2] Anderson, T.W. (1963). Asymptotic theory for principal component analysis, *The Annals of Mathematical Statistics*, **34**(1), 122–148.

[3] DasGupta, A. (2008). *Asymptotic Theory of Statistics and Probability*, Springer.

[4] Escoufier, Y. (1973). Le traitement des variables vectorielles, *Biometrics*, **29**(4), 751–760.

[5] Harville, D.A (1997). *Matrix Algebra From a Statistician's Perspective*, Springer-Verlag, New York.

[6] Hocking, R.R. (2003). *Methods and Applications of Linear Models: Regression and the Analysis of Variance*, 2nd edn., John Wiley & Sons.

[7] Horn, R.A. and Johnson, C.R. (1985). *Matrix Analysis*, Cambridge University Press.

[8] Lavit, C. (1988). *Analyse Conjointe de Tableaux Quantitatifs*, Masson, Paris.

[9] Lavit, C.; Escoufier, Y.; Sabatier R. and Traissac, P. (1994). The ACT (STATIS) method, *Computational Statistics & Data Analysis*, North-Holland **18**, 97–119.

[10] Lazraq, A.; Hanafi, M.; Cléroux, R.; Allaire, J. and Lepage, Y. (2008). Une approche inférentielle pour la validation du compromis de la méthode STATIS, *Journal de la Société Française de Statistique*, **149**, 97–109.

[11] Lima, E.L. (1989). *Curso de Análise, Volume 2*, 3.ª ed., Projecto Euclides.

[12] Muller, K.E. and Stewart, P.W. (2006). *Linear Model Theory: Univariate, Multivariate, and Mixed Models*, John Wiley & Sons.

[13] Robert, P. and Escoufier, Y. (1976). A unifying tool for linear multivariate statistical methods: the RV-coefficient, *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **25**(3), 257–265.

[14] Tong, Y.L. (1990). *The Multivariate Normal Distribution*, Springer-Verlag.

REVSTAT – Statistical Journal
Volume 12, Number 2, June 2014, 119–134

MADOGRAM AND ASYMPTOTIC INDEPENDENCE AMONG MAXIMA

Authors: Armelle Guillou
– Université de Strasbourg et CNRS, IRMA, UMR 7501,
7 rue René Descartes, 67084 Strasbourg Cedex, France
armelle.guillou@math.unistra.fr

Philippe Naveau
– Laboratoire des Sciences du Climat et de l'Environnement,
Orme des Merisiers / Bat.701 C.E. Saclay, 91191 Gif-sur-Yvette, France
philippe.naveau@lsce.ipsl.fr

Antoine Schorgen
– Université de Strasbourg et CNRS, IRMA, UMR 7501,
7 rue René Descartes, 67084 Strasbourg Cedex, France
schorgen@math.unistra.fr

Received: October 2012 Revised: April 2013 Accepted: April 2013

Abstract:

- A strong statistical research effort has been devoted in multivariate extreme value theory in order to assess the strength of dependence among extremes. This topic is particularly difficult in the case where block maxima are near independence. In this paper, we adapt and study a simple inference tool inspired from geostatistics, the madogram, to the context of asymptotic independence between pairwise block maxima. In particular, we introduce an extremal coefficient and study the theoretical properties of its estimator. Its behaviour is also illustrated on a small simulation study and a real data set.

Key-Words:

- *extremal coefficient; method-of-moment; maximum likelihood.*

AMS Subject Classification:

- 62G05, 62G20, 62G32.

1. INTRODUCTION

One recurrent question in multivariate extreme value theory (MEVT) is how to infer the strength of dependence among maxima. To illustrate this inquiry by an example, monthly maxima of hourly precipitation measured at two french stations from February 1987 to December 2002 are displayed in Figure 1. The two stations belong to the same hydrological basin of Orgeval (`https://gisoracle.cemagref.fr/`) that is located in France, west of Paris.

For each season, a scatterplot between the two stations shows the original 45 (15 years $\times$ 3 months per season) monthly maxima recorded in millimeters. The dependence structure seems to vary according to seasons and it is not clear if the largest summer values are close to independence.

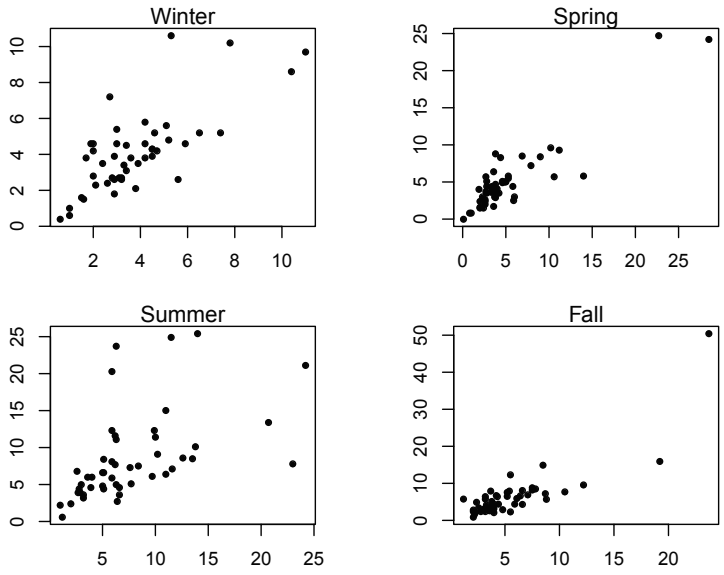


Figure 1: Monthly maxima of hourly precipitation for each season, measured at two stations in the basin of Orgeval (near Paris) during 1987-2002.

This concept of asymptotic independence has been studied by many authors. In this paper, we follow the approach introduced by Ledford and Tawn (1996) and extended by Ramos and Ledford (2009). Before explaining the details of our method, we need to recall a few basic concepts about MEVT and to introduce some notations. Suppose that we have at our disposal n independent and identically distributed pairs (X_i, Y_i) with unit-Fréchet margins ($\mathbb{P}(X_i \leq x) = \exp(-1/x)$ for $x > 0$) and that the component-wise maxima vector $(M_{X,n}, M_{Y,n}) =$

$(\max(X_1, ..., X_n), \max(Y_1, ..., Y_n))$ converges in the following way:

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{M_{X,n}}{n} \leq x, \frac{M_{Y,n}}{n} \leq y\right) = G(x, y), \text{ for } x, y > 0. \tag{1.1}$$

The limiting distribution function G is called the bivariate extreme value distribution and it can be written as $G(x, y) = \exp\{-V(x, y)\}$, with

$$V(x, y) = \int_0^1 \max\left(\frac{\omega}{x}, \frac{1-\omega}{y}\right) dH(\omega),$$

where $H(.)$ is a finite non-negative measure on $[0, 1]$ such that $\int_0^1 \omega dH(w) = \int_0^1 (1-\omega) dH(w) = 1$. This latter condition on H ensures that the margins $G(x, \infty)$ and $G(\infty, y)$ are unit-Fréchet distributed. The function V is called the pairwise extremal dependence function. It is homogeneous of order -1, i.e. $V(tx, ty) = t^{-1}V(x, y)$ for any positive t and G is max-stable, i.e. $G^t(tx, ty) = G(x, y)$. By definition of H, the function V has no explicit form and various non-parametric estimators of V have been studied (e.g. Capéraà *et al.*, 1997). As an example, an approach based on a classical geostatistical tool, the madogram (Matheron, 1987), was proposed by Naveau *et al.* (2009). Its simplest version (Cooley *et al.*, 2006) focused on the estimation of the extremal coefficient $\theta = V(1, 1)$. This coefficient provides a quick summary of the dependence between maxima. It belongs to the interval $[1, 2]$. If θ equals two, the pairwise maxima are independent, and if θ equals one, it is the complete dependence case. Cooley *et al.* (2006) defined the so-called F-madogram

$$\nu = \frac{1}{2}\mathbb{E}\left|F\left(M_{X,n}\right) - F\left(M_{Y,n}\right)\right|, \tag{1.2}$$

where F denotes the distribution function of $M_{X,n}$ and $M_{Y,n}$, in order to express the extremal coefficient as

$$\theta = \frac{1 + 2\nu}{1 - 2\nu}. \tag{1.3}$$

Going back to the maxima displayed in Figure 1, one may wonder if convergence (1.1) provides an appropriate probabilistic framework to study the near independence seen in our summer rainfall data. Convergence (1.1) implies that $\lim n\mathbb{P}\left(\frac{X_i}{n} > x \text{ or } \frac{Y_i}{n} > y\right) = -\log G(x, y)$. Hence

$$\lim_{n\longrightarrow\infty} n\mathbb{P}\left(\frac{X_i}{n} > x \text{ and } \frac{Y_i}{n} > y\right) = \log G(x, y) - \log G(x, \infty) - \log G(\infty, y).$$

If we are in the asymptotically independent case, i.e. $G(x, y)$ can be written as the product $G(x, y) = G(x, \infty)G(\infty, y)$, the right-hand side of the last convergence is nothing else than zero. This result is uninformative about the degree of dependence among our rainfall maxima. A conceptual extension is needed to improve our understanding of the probability of having joint extremes. To fill in this

gap, Ledford and Tawn in a series of papers (e.g. Ledford and Tawn, 1996, 1997, 1998) introduced a new tail model of the distribution which has been simplified by Ramos and Ledford (2009) as follows

$$\mathbb{P}\left(X > x, Y > y\right) = (xy)^{-\frac{1}{2\eta}}\mathcal{L}(x,y), \text{ for some } \eta \in (0,1], \tag{1.4}$$

with $\mathcal{L}$ a bivariate slowly varying function at infinity. The coefficient of tail dependence, η, is a measure of asymptotic independence. It is equal to one in the asymptotic dependence case and less than one in the asymptotic independence one. Condition (1.4) is tailored to analyze large excesses in the asymptotic independent case but it needs a reformulation in order to be used with pairs of maxima, as the ones pictured in Figure 1. This reformulation has been recently proposed by Ramos and Ledford (2011) who studied an extension of (1.1) by proving under the tail model (1.4) that, for $x, y > 0$,

$$\lim_{\varepsilon\to 0}\lim_{n\to\infty}\mathbb{P}\left[\frac{M_{X,n,\varepsilon b_n}}{b_n} \leq x, \frac{M_{Y,n,\varepsilon b_n}}{b_n} \leq y\right] = G_\eta(x,y) = \exp\Big[-V_\eta(x,y)\Big], \tag{1.5}$$

where the normalising constants b_n are defined implicitly as $n\mathbb{P}(X > b_n, Y > b_n) = 1$, $M_{\bullet,n,\varepsilon}$ corresponds to the component-wise maxima such that (X_i, Y_i) occur within the set $R_\varepsilon = \{(x,y) : x > \varepsilon, y > \varepsilon\}$ and

$$V_\eta(x,y) = \eta\int_0^1 \left[\max\left(\frac{\omega}{x}, \frac{1-\omega}{y}\right)\right]^{\frac{1}{\eta}} dH_\eta(\omega), \tag{1.6}$$

with H_η a finite and non-negative measure satisfying the constraint

$$\eta^{-1} = \int_0^{\frac{1}{2}} \omega^{\frac{1}{\eta}} dH_\eta(\omega) + \int_{\frac{1}{2}}^1 (1-\omega)^{\frac{1}{\eta}} dH_\eta(\omega).$$

The new dependence function V_η is homogeneous of order $-\frac{1}{\eta}$:

$$V_\eta(tx, ty) = t^{-\frac{1}{\eta}} V_\eta(x,y),$$

and the distribution $G_\eta(x,y)$ obeys an extended max-stable property:

$$G_\eta^{t^{1/\eta}}(tx, ty) = G_\eta(x,y).$$

In (1.1), a normalisation of n^{-1} is required in order to stabilize the component-wise maxima whereas in (1.5) b_n is of order $O(n^\eta)$.

The main goal of this paper is to adapt the concept of madogram to this framework of asymptotic independence. The asymptotic properties of our estimators are also derived. A small simulation study allows us to compare our inference scheme with the maximum likelihood estimation approach. All these estimators are applied to our rainfall data set.

2. THE F-MADOGRAM IN THE ASYMPTOTIC INDEPENDENCE CASE

Denote (M_X^*, M_Y^*) the bivariate vector that follows the distribution $G_\eta(x, y)$, i.e.

$$\mathbb{P}\left(M_X^* \leq x, M_Y^* \leq y\right) = \exp\{-V_\eta(x, y)\}, \tag{2.1}$$

with $V_\eta(x, y)$ of the form (1.6).

Concerning the marginals, we denote

$$F_X^*(x) := \mathbb{P}\left(M_X^* \leq x\right) = \exp\Big[-\sigma_X x^{-\frac{1}{\eta}}\Big] \text{ and } F_Y^*(y) := \exp\Big[-\sigma_Y y^{-\frac{1}{\eta}}\Big], \tag{2.2}$$

with $\sigma_X = V_\eta(1, \infty)$ and $\sigma_Y = V_\eta(\infty, 1)$. As the scaling coefficients σ_X and σ_Y are not necessarily equal, the Fréchet margins of M_X^* and M_Y^* differs by a multiplicative factor. In the classical MEVT setup defined by (1.1), the extremal coefficient $\theta = V(1, 1)$ was simple to explain. It always varied between one (dependence) and two (independence). Having different marginals in (2.2) makes it difficult to find simple and interpretable summaries like the extremal coefficient. One possible way around this interpretability issue is to go back to the madogram distance because it is trivial to interpret it as a metric and it is marginal free. The F-madogram for the pair (M_X^*, M_Y^*) can be defined as

$$\nu_\eta := \frac{1}{2}\mathbb{E}\left|F_X^*\left(M_X^*\right) - F_Y^*\left(M_Y^*\right)\right|, \tag{2.3}$$

and we can derive from (1.6) and (2.2) the relationship (see Appendix)

$$\theta_\eta = \frac{1 + 2\nu_\eta}{1 - 2\nu_\eta}, \tag{2.4}$$

where $\theta_\eta := V_\eta(\sigma_X^\eta, \sigma_Y^\eta)$ could be viewed as an analog of the classical extremal coefficient, comparing equations (1.3) and (2.4). If ν_η equals zero, then θ_η equals one. As the distance ν_η increases, the coefficient θ_η also increases. If M_X^* and M_Y^* are independent, then $F_X^*\left(M_X^*\right)$ and $F_Y^*\left(M_Y^*\right)$ are independent and uniformly distributed random variables. It follows that $\nu_\eta = 1/6$. From (2.4), we deduce that $\theta_\eta = 2$.

The only difference between equations (1.3) and (2.4) resides in the fact that the pairwise maxima vector belongs now to the largest family G_η instead of the classical G. It is also essential to emphasize that the F-madogram should not be interpreted alone. The coefficient η is paramount to explore the asymptotic independence domain.

3. INFERENCE

3.1. A method-of-moment approach

Our main result is the following theorem that deals with the convergence of the empirical estimator deduced from (2.3).

Theorem 1. *Let* $\left(M^*_{X_i,n}, M^*_{Y_i,n}\right)$ *be a sample of* N *bivariate maxima vectors of block size* n *that converges in distribution, as* $n \to \infty$*, to a bivariate extreme value distribution with an* η*-dependence function defined as in (1.6). Let*

$$\widehat{\nu}_\eta = \frac{1}{2N} \sum_{i=1}^{N} \left| \widehat{F}^*_X(M^*_{X_i,n}) - \widehat{F}^*_Y(M^*_{Y_i,n}) \right|, \tag{3.1}$$

where $\widehat{F}^*_X$*, resp.* $\widehat{F}^*_Y$*, denotes the empirical distribution function of the sample* $M^*_{X_i,n}$*, resp.* $M^*_{Y_i,n}$*. Then, as* $n \to \infty$ *and* $N \to \infty$

$$\sqrt{N}\left(\widehat{\nu}_\eta - \nu_\eta\right) \xrightarrow{d} \int_{[0,1]^2} N_C(u,v) dJ(u,v),$$

where $J(x,y) = \frac{1}{2}|x-y|$ *and* N_C *is a Gaussian process defined by*

$$N_C(u,v) = B_C(u,v) - B_C(u,1)\frac{\partial C}{\partial u}(u,v) - B_C(1,v)\frac{\partial C}{\partial v}(u,v), \tag{3.2}$$

and B_C *is a Brownian bridge on* $[0,1]^2$ *with covariance function*

$$\mathbb{E}\{B_C(u,v) \cdot B_C(u',v')\} = C(u \wedge u', v \wedge v') - C(u,v) \cdot C(u',v'),$$

with $u \wedge u' = \min(u,u')$ *and* C *the copula function with respect to (2.1).*

From (2.4), we introduce the following estimator for the extremal coefficient

$$\widehat{\theta}_\eta = \frac{1 + 2\widehat{\nu}_\eta}{1 - 2\widehat{\nu}_\eta}. \tag{3.3}$$

Applying the delta method, the following corollary follows.

Corollary 2. *Under the assumption of Theorem 1, we have*

$$\sqrt{N}\left(\widehat{\theta}_\eta - \theta_\eta\right) \xrightarrow{d} (1+\theta_\eta)^2 \int_{[0,1]^2} N_C(u,v) dJ(u,v).$$

To infer the value of η, we complement our method-of-moment via a Generalized Probability Weighted Moment (GPWM) approach (Diebolt *et al.*, 2008) based on the following moment equality

$$\mu_\omega = \mathbb{E}(M^*\omega(F_{M^*})) = \int_{-\infty}^{+\infty} x\ \omega(F_{M^*}(x))dF_{M^*}(x),$$

for any variable M^* with a distribution function F_{M^*} and ω a suitable continuous function satisfying

$$\begin{cases} \omega(t) = O((1-t)^b) & \text{for } t \text{ close to 1, } b \geq 0, \\ \omega(t) = O(t^{a'}) & \text{for } t \text{ close to 0, } a' > 0. \end{cases} \tag{3.4}$$

If we take $M^* = \max(M^*_{X,n}, M^*_{Y,n})$, whose distribution function equals $F_{M^*}(x) = \exp\{-V_\eta(1,1)x^{-\frac{1}{\eta}}\}$ (using Equation (2.1) and the homogeneous property of V_η) and if $\omega(t) := \omega_{a,b}(t) = t^a(-\log t)^b, a > a'$ then Diebolt *et al.* (2008) proved that

$$\mu_{a,b} := \mu_\omega = \frac{V_\eta^\eta(1,1)}{(a+1)^{b-\eta+1}}\Gamma(b-\eta+1) \tag{3.5}$$

where $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx$.

A natural estimator for $\mu_{a,b}$ can be obtained by replacing F_{M^*} by its empirical version $\mathbb{F}_n$

$$\widehat{\mu}_{a,b} = \int_0^1 \mathbb{F}_n^{-1}(u)u^a(-\log u)^b du.$$

Using (3.5) with suitable values for (a,b) allows us to obtain an estimator for η in function of $\widehat{\mu}_{a,b}$

$$\widehat{\eta}_{gpwm} = 2\left(1 - \frac{\widehat{\mu}_{1,2}}{\widehat{\mu}_{1,1}}\right). \tag{3.6}$$

The asymptotic normality of $\widehat{\eta}_{gpwm}$ can then be deduced from the asymptotic properties of the GPWM estimator, see our Appendix.

Proposition 3. *Let $\left(M^*_{X_i,n}, M^*_{Y_i,n}\right)$ be a sample of N bivariate maxima vectors of block size n that follows a bivariate extreme value distribution with an η-dependence function defined as in (1.6). Then the GPWM estimator of η defined by $\widehat{\eta}_{gpwm}$ converges in the following way*

$$\sqrt{N}\left(\widehat{\eta}_{gpwm} - \eta\right) \xrightarrow{d} \frac{\eta 2^{3-\eta}}{\Gamma(2-\eta)}\left[I_1 - (1-\eta/2)I_2\right],$$

with $I_1 = \int_0^1 B(t)(-\log t)^{-\eta+1}\,dt$, $I_2 = \int_0^1 B(t)(-\log t)^{-\eta}\,dt$ and B a Brownian bridge.

3.2. The maximum likelihood approach

Besides our aforementioned method-of-moment approach, a Maximum Likelihood (ML) method can also be implemented. Our ML method is based on the normalized sample $\{M_i\} = \left\{\max\left(\frac{M^*_{X_i,n}}{\sigma_X^\eta}, \frac{M^*_{Y_i,n}}{\sigma_Y^\eta}\right)\right\}$, $i = 1, ..., N$ which admits the following log-likelihood

$$\log L(M_1, ..., M_N; \theta_\eta, \eta) = N \log\left(\frac{\theta_\eta}{\eta}\right) - (1/\eta + 1)\sum_{i=1}^{N} \log(M_i) - \theta_\eta \sum_{i=1}^{N} M_i^{-1/\eta}.$$

If $\widehat{\eta}_{mle}$ denotes the ML estimator for η based on the univariate sample $\left\{\max\left(M^*_{X_i,n}, M^*_{Y_i,n}\right)\right\}$ with a distribution function given by (2.1), it allows us to derive

$$\widehat{\theta}_{\eta,mle} = \left[\frac{1}{N}\sum_{i=1}^{N} \min\left(\widehat{\sigma}_{X,mle}(M^*_{X_i,n})^{-1/\widehat{\eta}_{mle}}, \widehat{\sigma}_{Y,mle}(M^*_{Y_i,n})^{-1/\widehat{\eta}_{mle}}\right)\right]^{-1}.$$

The estimates for σ_X and σ_Y in the above equality can be derived from (2.2) as $\widehat{\sigma}_{X,mle} = \left[\frac{1}{N}\sum_{i=1}^{N}(M^*_{X_i,n})^{-1/\widehat{\eta}_{mle}}\right]^{-1}$ and a similar expression for $\widehat{\sigma}_{Y,mle}$. Thus we can define

(3.7)

$$\widehat{\theta}_{\eta,mle} = \left[\frac{1}{N}\sum_{i=1}^{N} \min\left(\frac{(M^*_{X_i,n})^{-1/\widehat{\eta}_{mle}}}{\frac{1}{N}\sum_{j=1}^{N}(M^*_{X_j,n})^{-1/\widehat{\eta}_{mle}}}, \frac{(M^*_{Y_i,n})^{-1/\widehat{\eta}_{mle}}}{\frac{1}{N}\sum_{j=1}^{N}(M^*_{Y_j,n})^{-1/\widehat{\eta}_{mle}}}\right)\right]^{-1}.$$

4. EXAMPLES

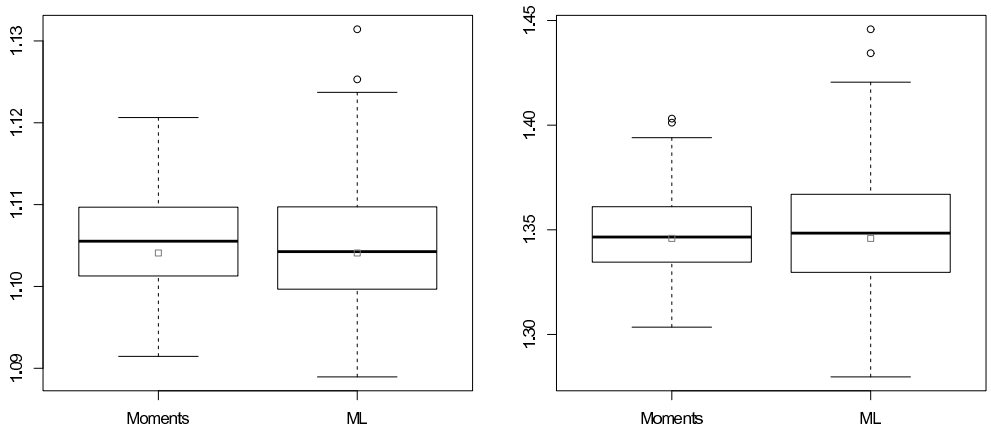
4.1. A small simulation

To compare our estimators with the ML approach, we simulate 300 samples of 500 pairs of maxima from the η-asymmetric logistic dependence model (see Ramos and Ledford, 2011):

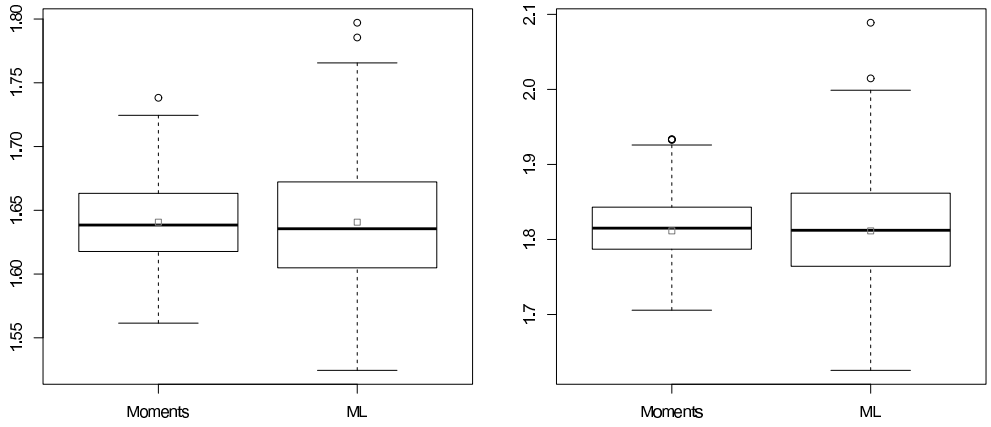
$$V_\eta(x, y) = \frac{1}{2 - 2^{\alpha/\eta}}(x^{-1/\alpha} + y^{-1/\alpha})^{\alpha/\eta}, \text{ for } x, y > 0,$$

with $\alpha \in \{0.1, 0.3, 0.5, 0.6\}$ and $\eta = 0.7$. This specific value of η corresponds to a case of asymptotic independence ($\eta < 1$) and provides $\theta_\eta = 2^{\alpha/\eta}$.

Boxplots of the estimators of θ and η are given in Figures 2 and 3 for different values of α and η. In these figures, the small square represents the true value of the parameter whereas the horizontal line is the median based on the 300 simulations.



(a) Case $\alpha = 0.1, \eta = 0.7$ (b) Case $\alpha = 0.3, \eta = 0.7$

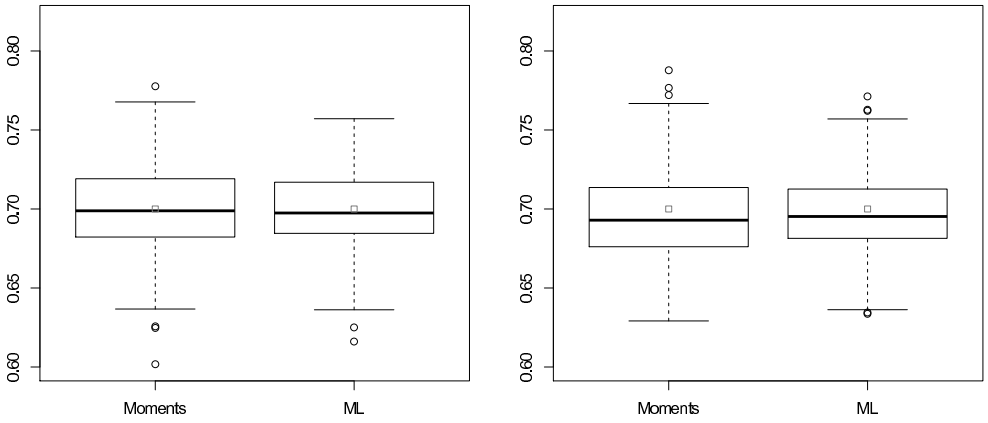


(c) Case $\alpha = 0.5, \eta = 0.7$ (d) Case $\alpha = 0.6, \eta = 0.7$

Figure 2: Simulation: comparing $\widehat{\theta}_\eta$ from (3.3) with $\widehat{\theta}_{\eta,mle}$ from (3.7).

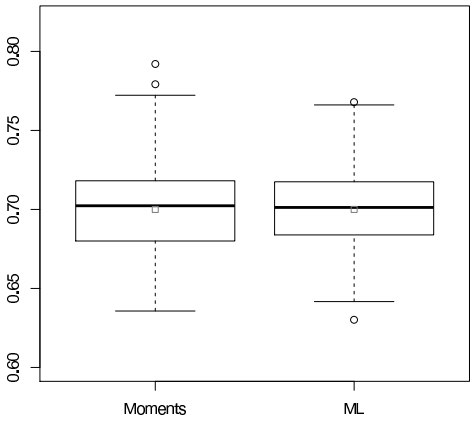
In Figure 2 we can observe that the estimate $\widehat{\theta}_{\eta,mle}$ from (3.7) has a higher variability than $\widehat{\theta}_\eta$ from (3.3). This is particularly true when α is close to η, i.e. $\widehat{\theta}_\eta$ near two.

Concerning the estimation of η, Figure 3 basically tells the opposite story. The ML approach appears slightly better than the method-of-moment. This small simulation study advocates for not restricting one inference approach but rather to combine or at least compare different inference techniques.

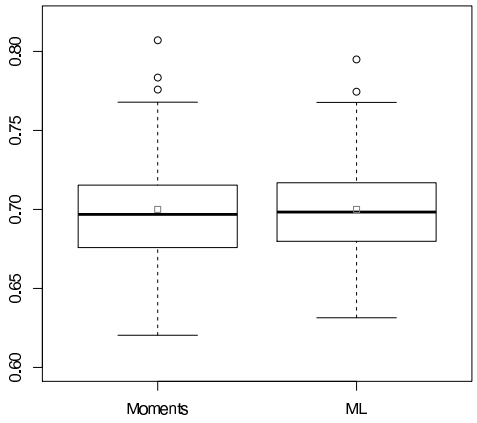


(a) Case $\alpha = 0.1, \eta = 0.7$

(b) Case $\alpha = 0.3, \eta = 0.7$



(c) Case $\alpha = 0.5, \eta = 0.7$



(d) Case $\alpha = 0.6, \eta = 0.7$

Figure 3: Simulation: comparing $\widehat{\eta}_{gpwm}$ from (3.6) with $\widehat{\eta}_{mle}$.

4.2. Orgeval Rainfall data

Table 1 summarizes our inference with respect to the maxima plotted in Figure 1.

If one has to guess from Figure 1 some information about the degree of dependence, precipitation maxima during the Summer season clearly appear to be the less uncorrelated, followed by the Winter ones. The Spring and Fall seasons seem to witness a stronger and similar dependence.

Table 1: Estimates with GPWM and ML approaches for the Orgeval rainfall data.

	$\widehat{\theta}_\eta$	$\widehat{\theta}_{\eta,mle}$	$\widehat{\eta}_{gpwm}$	$\widehat{\eta}_{mle}$
winter	1.44	1.26	0.44	0.71
spring	1.33	1.22	0.50	0.70
summer	1.45	1.47	0.56	0.72
fall	1.36	1.60	0.49	0.51

Concerning the GPWM approach, from Table 1 we can see that the Spring and Fall seasons basically have the same η and the same θ. This parallel confirms Figure 1 where the points are strongly clustered for those two panels. Concerning the Winter and Summer seasons, the corresponding $\widehat{\theta}_\eta$ are much alike, but the $\widehat{\eta}_{gpwm}$ are different. Visually, this does not contradict the Winter and Summer displays, but it is not straightforward to interpret such results.

From Table 1, $\widehat{\eta}_{mle}$ appears to be almost equal to 0.7 for all seasons, but the Fall. It is puzzling that the Spring season belongs to this group because Figure 1 and the GPWM approach clearly discriminates the Spring from the Winter and Summer seasons. On the positive for the MLE approach, having the same η for the Winter, Spring and Summer, we can compare the ML estimates of θ. The ordering among those three $\widehat{\theta}_{\eta,mle}$ is coherent with Figure 1, the Summer has the largest value and the Spring the smallest one. The Fall season is difficult to interpret with the MLE approach, $\widehat{\eta}_{mle}$ being quite different to the values in the other seasons.

Now, if we want to compare the two approaches, GPWM and MLE, looking at Table 1, we can see that $\widehat{\theta}_\eta$ is quite stable, which is not the case for $\widehat{\theta}_{\eta,mle}$. At first sight, as both quantities estimate θ, it is difficult to know what to conclude. However, if we look at Figure 2(b) where the value of θ_η is in the range 1.3–1.4

(corresponding to the values given in Table 1) we can observe that indeed the variability with the maximum likelihood approach is more important than with the moment method. Thus this corroborates the instability of $\widehat{\theta}_{\eta,mle}$ observed in Table 1.

Overall, the time period of 1987–2002 may be too short to clearly compare the dependence among different seasons. Still, this example illustrates that analyzing jointly θ and η can bring relevant information that may not be obtained by simply interpreting θ.

ACKNOWLEDGMENTS

The authors thank the referee for the comments concerning the previous version. Part of this work has been supported by the EU-FP7 ACQWA project (www.acqwa.ch) under Contract Nr 212250, by the PEPER-GIS project, by the ANR MOPERA project, by the ANR McSim project and the MIRACCLE-GICC project. The authors acknowledge the GIS ORACLE for the Orvegal data.

APPENDIX

Proof of (2.4): Applying the equality $|a-b|/2 = \max(a,b) - (a+b)/2$ to ν_η, we get:

$$\frac{1}{2}\mathbb{E}\,|F_X^*(M_X^*) - F_Y^*(M_Y^*)| = \mathbb{E}\max\{F_X^*(M_X^*), F_Y^*(M_Y^*)\} - \frac{1}{2}.$$

Then we calculate

$$\begin{aligned}\mathbb{P}[\max\{F_X^*(M_X^*), F_Y^*(M_Y^*)\} \le u] &= \mathbb{P}[M_X^* \le F_X^{*\leftarrow}(u), M_Y^* \le F_Y^{*\leftarrow}(u)] \\ &= \exp\{-V_\eta(F_X^{*\leftarrow}(u), F_Y^{*\leftarrow}(u))\} \\ &= \exp\{\log(u) V_\eta(\sigma_X^\eta, \sigma_Y^\eta)\} = u^{V_\eta(\sigma_X^\eta, \sigma_Y^\eta)}\end{aligned}$$

because from the margin model (2.2) we have $F_X^{*\leftarrow}(u) = (-\log(u)/\sigma_X)^{-\eta}$ and $F_Y^{*\leftarrow}(u) = (-\log(u)/\sigma_Y)^{-\eta}$. Therefore, $\mathbb{E}\max\{F_X^*(M_X^*), F_Y^*(M_Y^*)\} = \frac{V_\eta(\sigma_X^\eta, \sigma_Y^\eta)}{1+V_\eta(\sigma_X^\eta, \sigma_Y^\eta)}$ from which (2.4) follows. □

Proof of Theorem 1: First, we introduce the 'normalized' empirical distribution functions

$$\widetilde{F}_{n,N,X}(u) := \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{\left[b_n^{-1} M_{X_i,n}^* \le u\right]}, \quad \widetilde{F}_{n,N,Y}(u) := \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{\left[b_n^{-1} M_{Y_i,n}^* \le u\right]},$$

and we rewrite the estimator of the madogram as

$$\widehat{\nu}_\eta = \frac{1}{2N}\sum_{i=1}^{N}\Big|\widetilde{F}_{n,N,X}\Big(b_n^{-1}M^*_{X_i,n}\Big) - \widetilde{F}_{n,N,Y}\Big(b_n^{-1}M^*_{Y_i,n}\Big)\Big|.$$

Before starting the proof, we need to introduce a series of definitions linked to the copula function C. Although very similar, these definitions represent slightly different estimators of the same copula function. One difficulty of the proof is to show how close these versions are:

$$\widetilde{C}_{n,N}(u,v) := \frac{1}{N}\sum_{i=1}^{N}\mathbb{1}_{\left\{\widetilde{F}_{n,N,X}\left(b_n^{-1}M^*_{X_i,n}\right)\le u,\widetilde{F}_{n,N,Y}\left(b_n^{-1}M^*_{Y_i,n}\right)\le v\right\}},$$

$$C_{n,N}(u,v) := \frac{1}{N}\sum_{i=1}^{N}\mathbb{1}_{\left\{b_n^{-1}M^*_{X_i,n}\le\widetilde{F}^{\leftarrow}_{n,N,X}(u),b_n^{-1}M^*_{Y_i,n}\le\widetilde{F}^{\leftarrow}_{n,N,Y}(v)\right\}},$$

$$\widetilde{C}^*_{n,N}(u,v) := \frac{1}{N}\sum_{i=1}^{N}\mathbb{1}_{\left\{U_{X_i,n}\le\widetilde{F}^{*\leftarrow}_{n,N,X}(u),V_{Y_i,n}\le\widetilde{F}^{*\leftarrow}_{n,N,Y}(v)\right\}}$$

where

$$\begin{aligned}
\widetilde{F}_{X,n}(z) &:= \mathbb{P}\Big(b_n^{-1}M^*_{X_i,n}\le z\Big),\quad \widetilde{F}_{Y,n}(z) := \mathbb{P}\Big(b_n^{-1}M^*_{Y_i,n}\le z\Big),\\
U_{X_i,n} &:= \widetilde{F}_{X,n}\Big(b_n^{-1}M^*_{X_i,n}\Big),\quad V_{Y_i,n} := \widetilde{F}_{Y,n}\Big(b_n^{-1}M^*_{Y_i,n}\Big),\\
\widetilde{F}^*_{n,N,X}(u) &:= \frac{1}{N}\sum_{i=1}^{N}\mathbb{1}_{\{U_{X_i,n}\le u\}},\quad \widetilde{F}^*_{n,N,Y}(v) := \frac{1}{N}\sum_{i=1}^{N}\mathbb{1}_{\{V_{Y_i,n}\le v\}}.
\end{aligned}$$

The proof of Theorem 1 is divided into the following five steps.

Step 1. The function $\widetilde{C}_{n,N}(u,v)$ is very similar to $C_{n,N}(u,v)$, i.e. $\sup_{0\le u,v\le 1}|\widetilde{C}_{n,N}(u,v) - C_{n,N}(u,v)| \le 2/N$.

Step 2. We have $C_{n,N}(u,v) = \widetilde{C}^*_{n,N}(u,v)$.

Step 3. Define now the empirical distribution function of $(U_{X_i,n}, V_{Y_i,n})$ as

$$\widetilde{H}^*_{n,N}(u,v) = \frac{1}{N}\sum_{i=1}^{N}\mathbb{1}_{\left\{U_{X_i,n}\le u,V_{Y_i,n}\le v\right\}}$$

and its non-empirical version as

$$\widetilde{H}^*_n(u,v) = \mathbb{P}\left(\widetilde{F}_{X,n}\Big(b_n^{-1}M^*_{X_i,n}\Big)\le u,\widetilde{F}_{Y,n}\Big(b_n^{-1}M^*_{Y_i,n}\Big)\le v\right).$$

We establish that the process $\sqrt{N}(\widetilde{H}^*_{n,N} - \widetilde{H}^*_n)$ tends in distribution to a Brownian bridge B_C. To this end, we prove the convergence of the finite-dimensional distributions and the tightness of the process.

Step 4. The process $\sqrt{N}(\widetilde{C}^*_{n,N} - \widetilde{H}^*_n)$ tends in distribution to a Gaussian process N_C.

Step 5. We conclude the proof of our theorem using the integration by parts. □

This proof is only sketched here as it is a slightly modified version of the one of Proposition 4 in Naveau *et al.* (2009) which is detailed in `http://sama.ipsl.jussieu.fr/Documents/articles/NaveauBiometrika07DetailedProofs.pdf`.

Remark about our Theorem 1. The limiting process is such that

$$\int_{[0,1]^2} N_C(u,v)dJ(u,v) = \frac{1}{2}\int_0^1 N_C(0,v)dv + \frac{1}{2}\int_0^1 N_C(u,0)du - \int_0^1 N_C(u,u)du.$$

This limiting process cannot be precised without specifying the copula function and in special cases where these integrals can be computed. For instance, consider the Product copula, also called the independent copula, defined as $C(u,v) = uv$. In that case

$$N_C(u,v) = B_C(u,v) - vB_C(u,1) - uB_C(1,v),$$

from which direct computations lead to

$$\mathrm{Var}\left(\int_{[0,1]^2} N_C(u,v)dJ(u,v)\right) = \frac{1}{90}.$$

Proof of Proposition 3: As $\eta \in (0,1]$, we have according to Theorem 2.1 in Diebolt *et al.* (2008) that

$$\sqrt{N}\begin{pmatrix} \widehat{\mu}_{1,1} - \mu_{1,1} \\ \widehat{\mu}_{1,2} - \mu_{1,2} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \eta V^\eta_\eta(1,1)\displaystyle\int_0^1 \frac{B(t)}{t}\Big(-\log t\Big)^{-\eta-1} t(-\log t)dt \\ \eta V^\eta_\eta(1,1)\displaystyle\int_0^1 \frac{B(t)}{t}\Big(-\log t\Big)^{-\eta-1} t(-\log t)^2 dt \end{pmatrix},$$

where B denotes a Brownian bridge and $n \to \infty$. It follows

$$\begin{aligned}\sqrt{N}\left(\widehat{\eta}_{gpwm} - \eta\right) &= -\frac{2}{\widehat{\mu}_{1,1}\mu_{1,1}}\sqrt{N}\left(\mu_{1,1}\widehat{\mu}_{1,2} - \mu_{1,2}\widehat{\mu}_{1,1}\right) \\ &= -\frac{2}{\widehat{\mu}_{1,1}\mu_{1,1}}\left[\mu_{1,1}\sqrt{N}(\widehat{\mu}_{1,2} - \mu_{1,2}) - \mu_{1,2}\sqrt{N}(\widehat{\mu}_{1,1} - \mu_{1,1})\right].\end{aligned}$$

An application of Slutzky's theorem leads to

$$\begin{aligned}&\sqrt{N}\left(\widehat{\eta}_{gpwm} - \eta\right) \\ &\qquad\xrightarrow{d} -\frac{2\eta V^\eta_\eta(1,1)}{\mu^2_{1,1}}\int_0^1 \frac{B(t)}{t}\Big(-\log t\Big)^{-\eta-1}\left[\mu_{1,1}\omega_{1,2}(t) - \mu_{1,2}\omega_{1,1}(t)\right]dt,\end{aligned}$$

from which Proposition 3 follows. □

REFERENCES

[1] Capéraà, P.; Fougères, A.-L. and Genest, C. (1997). A nonparametric estimation procedure for bivariate extreme value copulas, *Biometrika*, **84**, 567–577.

[2] Cooley, D.; Naveau, P. and Poncet, P. (2006). Variograms for spatial max-stable random fields. In: Dependence in probability and statistics, *Lecture Notes in Statistics*, **187**, 373–390, Springer, New-York.

[3] Diebolt, J.; Guillou, A.; Naveau, P. and Ribereau, P. (2008). Improving probability-weighted moment methods for the generalized extreme value distribution, *REVSTAT Statistical Journal*, **6**, 35–50.

[4] Ledford, A.W. and Tawn, J.A. (1996). Statistics for near independence in multivariate extreme values, *Biometrika*, **83**, 169–187.

[5] Ledford, A.W. and Tawn, J.A. (1997). Modelling dependence within joint tail regions, *Journal of the Royal Statistical Society Series B Statistical Methodology*, **59**, 475–499.

[6] Ledford, A.W. and Tawn, J.A. (1998). Concomitant tail behaviour for extremes, *Advances in Applied Probability*, **30**, 197–215.

[7] Matheron, G. (1987). Suffit-il, pour une covariance, d'être de type positif? *Sciences de la Terre, Série informatique géologique*, **26**, 51–66.

[8] Naveau, P.; Guillou, A.; Cooley, D. and Diebolt, J. (2009). Modelling pairwise dependence of maxima in space, *Biometrika*, **96**, 1–17.

[9] Ramos, A. and Ledford, A. (2009). A new class of models for bivariate joint tails, *Journal of the Royal Statistical Society Series B Statistical Methodology*, **71**, 219–241.

[10] Ramos, A. and Ledford, A. (2011). Alternative point process framework for modeling multivariate extreme values. *Communications in Statistics — Theory and Methods*, **40**, 2205–2224.

REVSTAT – Statistical Journal
Volume 12, Number 2, June 2014, 135–155

ON THE UPCROSSINGS OF TRIGONOMETRIC POLYNOMIALS WITH RANDOM COEFFICIENTS

Author: K.F. Turkman
– DEIO-CEAUL, FCUL, Bloco C6,
Campo Grande, 1749-016 Lisbon, Portugal
kfturkman@fc.ul.pt

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

Abstract:

- Polynomials are one of the oldest and the most versatile classes of functions which are fundamental in approximating highly complex, deterministic as well as random nonlinear functions and systems. Their use has been acknowledged in every scientific field from physics to ecology. Just to emphasize their great use in many fields, we mention their fundamental role in linear and non-linear time series analysis. In this paper, we give a review of some of the results related to polynomials with random coefficients and highlight the Poisson character of high level upcrossings of certain random coefficient trigonometric polynomials which are used in spectral analysis of time series.

Key-Words:

- *random polynomials; trigonometric polynomials; extreme value theory.*

AMS Subject Classification:

- 60G70.

1. INTRODUCTION

Polynomials form the backbone of mathematics in general and approximation of complex deterministic and random nonlinear functions and dynamic processes in particular. Examples of their use are countless coming from diverse areas such as quantum chaotic dynamics to ecology. Their use repeatedly appears in asymptotic theory of statistics and in particular in time series analysis which will be our primary interest.

In its simplest form, algebraic characteristics of polynomials are very much used in statistics. The celebrated expansions related to central limit theorems such as the Edgeworth expansions, Berry–Esseen type theorems and the delta method all depend on polynomial expansions and form the basis of asymptotic theory of statistics. Polynomials and their algebraic properties are also used in constructing stationary, invertible finite parameter linear representations for time series. Wold decomposition theorem says that under fairly general conditions any stationary time series may be represented as a causal convergent sum

$$X_t = \sum_{j=0}^{\infty} \psi_j \, \epsilon_{t-j} \ ,$$

with uncorrelated finite variance random variables $\{\epsilon_t\}$ and real values $\{\psi_j\}$ such that $\sum_j \psi_j^2 < \infty$. As a class of models, such a representation is not particularly useful due to the infinite number of parameters, and finite parameter versions called the class of stationary and invertible ARMA models are found by using the backshift operator $B^j X_t = X_{t-j}$, then representing the series in the form

$$X_t = \left[\sum_{j=0}^{\infty} \psi_j \, B^j \right] \epsilon_t \ .$$

Under quite general conditions, the polynomial $\psi(B) = \sum_{j=0}^{\infty} \psi_j \, B^j$ can be written as a ratio of two finite order polynomials $\Phi_p(B)$ and $\Theta_q(B)$ of orders p and q respectively, thus permitting us to write $\Phi_p(B) \, X_t = \Theta_q(B) \, \epsilon_t$. The conditions of stationarity and invertibility of the process X_t are then given in terms of the roots of the polynomials $\Phi(B)$ and $\Theta(B)$. In these examples, the well known algebraic results on deterministic polynomials are used. However, in a more general set up, random polynomials are used for very general representations for stationary times series.

Let us start with a collection of standard Gaussian random variables $\{X_s, s \leq t\}$ and consider the space of all measurable functions defined on this collection with the usual inner product

$$\langle f, g \rangle \ = \int_{-\infty}^{\infty} f(x) \, g(x) \, \frac{1}{\sqrt{2\pi}} \, \exp(-x^2/2) \, dx \ . \tag{1.1}$$

This space together with this inner product is a Hilbert space, and (random) Hermite polynomials form a closed and complete orthogonal system.

Hermite polynomials $H_n(x)$ of degree n are defined as

$$\int_{-\infty}^{\infty} H_n(x)\, H_m(x)\, \frac{1}{\sqrt{2\pi}}\, \exp(-x^2/2)\; dx \;=\; n!\, \delta_{n,m}\;, \qquad n, m = 0, 1, 2, ... \tag{1.2}$$

where

$$\delta_{n,m} = \begin{cases} 1\,, & n = m\,; \\ 0\,, & n \neq m\,. \end{cases} \tag{1.3}$$

Hence, every Borel measurable function g with finite variance (with respect to the Gaussian density) such that

$$\int_{-\infty}^{\infty} g^2(x)\, \frac{1}{\sqrt{2\pi}}\, \exp(-x^2/2)\; dx \;<\; \infty\;,$$

can be written as a linear combination (or as a limit) of these Hermite polynomials

$$g(x) \;=\; \lim_{N\to\infty} \sum_{n=0}^{N} \frac{g_n}{n!}\, H_n(x)\;, \tag{1.4}$$

where, the coefficients g_n are given by

$$g_n \;= \int_{-\infty}^{\infty} g(x)\, H_n(x)\, \frac{1}{\sqrt{2\pi}}\, \exp(-x^2/2)\; dx\;.$$

The convergence of (1.4) is in the mean square sense

$$\lim_{N\to\infty} \int_{-\infty}^{\infty} \left(g(x) - \sum_{n=0}^{N} \frac{g_n}{n!}\, H_n(x) \right)^2 \frac{1}{\sqrt{2\pi}}\, \exp(-x^2/2)\; dx \;=\; 0\;.$$

Note that the inner product is a integral with respect to the standard Gaussian density and hence the Hermite polynomials are orthogonal with respect to the standard normal probability distribution. Instead of Hermite polynomials, we can define Hermite functions

$$\psi_n(x) \;=\; \frac{1}{\sqrt{n!\, 2^n\, \sqrt{2\pi}}}\, \exp(-x^2/2)\, H_n(x)\;.$$

Hermite functions are normalized versions of the Hermite polynomials, therefore they form an closed and complete orthonormal basis. The closed linear span of Hermite polynomials is the space of all polynomials, and any element of this space can be represented as sums of products of polynomials given in the form

$$\sum_{p=1}^{\infty} \sum_{i_1=1}^{\infty} \cdots \sum_{i_p=1}^{\infty} a_{i_1 i_2 \cdots i_p} \prod_{v=1}^{p} X_{i_v}\;. \tag{1.5}$$

Here we will not enter into further details, which can be found in Terdik (1999).

The following remarkable result due to Nisio (1964) extends this polynomial representation to any strictly stationary time series.

Let ϵ_t be independent, standard Gaussian random variables. The polynomial representation

$$
\begin{aligned}
Y_t^{(m)} &= \sum_{p=1}^{m} \sum_{i_1=-\infty}^{\infty} \sum_{i_2=-\infty}^{\infty} \cdots \sum_{i_m=-\infty}^{\infty} g_{i_1 i_2 \cdots i_m} \prod_{v=1}^{p} \epsilon_{t-i_v} \\
&= \sum_{i_1=-\infty}^{\infty} g_{i_1} \epsilon_{t-i_1} \\
&\quad + \sum_{i_1=-\infty}^{\infty} \sum_{i_2=-\infty}^{\infty} g_{i_1 i_2} \epsilon_{t-i_1} \epsilon_{t-i_2} \\
&\quad + \sum_{i_1=-\infty}^{\infty} \sum_{i_2=-\infty}^{\infty} \sum_{i_3=-\infty}^{\infty} g_{i_1 i_2 i_3} \epsilon_{t-i_1} \epsilon_{t-i_2} \epsilon_{t-i_3} \\
&\quad + \cdots \\
&\quad + \sum_{i_1=-\infty}^{\infty} \sum_{i_2=-\infty}^{\infty} \cdots \sum_{i_m=-\infty}^{\infty} g_{i_1 i_2 \cdots i_m} \epsilon_{t-i_1} \epsilon_{t-i_2} \cdots \epsilon_{t-i_m}
\end{aligned} \tag{1.6}
$$

is called a Volterra series of order m. We will call

$$
Y_t = \sum_{p=1}^{\infty} \sum_{i_1=-\infty}^{\infty} \sum_{i_2=-\infty}^{\infty} \cdots \sum_{i_p=-\infty}^{\infty} g_{i_1 i_2 \cdots i_p} \prod_{v=1}^{p} \epsilon_{t-i_v} \ , \tag{1.7}
$$

the Volterra series expansion.

Theorem 1.1. *Let X_t be any strictly stationary time series. Then there exists a sequence of Volterra series $Y_t^{(m)}$ such that*

$$
\lim_{m \to \infty} Y_t^{(m)} \stackrel{d}{=} X_t \ ,
$$

in the sense that for any n and for any $\theta_j, |j| \le n$ as $m \to \infty$,

$$
\begin{aligned}
\Big| E \exp\big(i\,\theta_{-n} X_{-n} + \cdots + i\,\theta_n X_n\big) \\
- E \exp\big(i\,\theta_{-n} Y_{-n}^{(m)} + \cdots + i\,\theta_n Y_n^{(m)}\big) \Big| \ \to \ 0 \ .
\end{aligned} \tag{1.8}
$$

If further X_t is Gaussian, then X_t can be represented by

$$
X_t = \sum_{j=-\infty}^{\infty} g_j \, \epsilon_{t-j} \ ,
$$

for some real numbers $\{g_j\}$, such that $\sum_{=-\infty}^{\infty} g_j^2 < \infty$.

Hence random polynomials are basic functions with which we construct very general random processes. There is also a relationship between solutions of random difference equations and random polynomials.

Consider a stochastic difference equation

$$X_t = A_t X_{t-1} + B_t \ , \tag{1.9}$$

where $\{A_t, B_t\}$ is a sequence of random variables. We will call (1.9) a stochastic recurrence equation. It is possible to define (1.9) in a more general form

$$\mathbf{X}_t = \mathbf{A}_t \mathbf{X}_{t-1} + \mathbf{B}_t \ , \tag{1.10}$$

where $\mathbf{X}_t$ and $\mathbf{B}_t$ are random vectors in $\mathcal{R}^d$, $\mathbf{A}_t$ are $d \times d$ random matrices and $\{\mathbf{A}_t, \mathbf{B}_t\}_{n=0}^{\infty}$ is a strictly stationary ergodic process.

Many well known classes of nonlinear time series models such as bilinear, ARCH, GARCH, random coefficient autoregressive models (RCA) as well as threshold models can be represented in this form. Theorems due to Vervaat(1979) and Brant(1986) show that under fairly general conditions on $\{A_t, B_t\}$, stochastic difference equations of the form (1.9) have solutions given by

$$R = \sum_{k=1}^{\infty} \prod_{j=1}^{k-1} A_j B_k \ . \tag{1.11}$$

It is clear that the solution (1.11) are algebraic polynomials of random variables.

Extremal behavior of these polynomial expansions have played important role in understanding the oscillatory behavior of nonlinear processes, and many results on the point processes of upcrossings or exceedances of such polynomial expansions exist. See Turkman and Amaral Turkman(1997) and Scotto and Turkman(2002,2005) and de Haan *et al.*(1989).

Random polynomials of different nature given in the form

$$F_n(\omega, x) = a_0(\omega) F_1(x) + a_1(\omega) F_2(x) + ... + a_n(\omega) F_n(x) \ , \tag{1.12}$$

where $\{a_i(\omega)\}_{i=}^{n}$ are a sequence of random variables defined on a probability space $(\Omega, \mathcal{F}, P)$ and $\{F_i(x)\}_{i=1}^{n}$ are deterministic functions of x, have found significant applications in many fields involving the reliability of complex random physical systems. When $F_i(x) = x^i$, then the solution (1.12) is an algebraic polynomial with random coefficients, taking the form

$$F_n(\omega, x) = \sum_{j=0}^{n} a_j(\omega)\, x^j \ ,$$

whereas when $F_i(x)$ are trigonometric functions then (1.12) is a trigonometric polynomial of order n with random coefficients, often given in the form

$$F_n(\omega, x) = a_0(\omega) + \sum_{j=0}^{n} a_j(\omega) \cos jx + \sum_{j=1}^{n} b_j(\omega) \sin jx \ .$$

Here, contrary to polynomials given in (1.6), the type of polynomials we consider in (1.12) are deterministic in its argument, having random coefficients. We refer the reader to Bharucha-Reid and Sambandham(1986) and Farahmand(1998) for the general treatment of such random polynomials. We also refer the reader to Zygmund(2002) for a full account of developments on trigonometric polynomials.

Polynomials with random coefficients have many interesting characteristics, but the level crossing properties are particularly important and useful. Describing the reliability of a complex physical system subject to random inputs depends on understanding the oscillatory behavior of its sample paths. Level crossing problems of stochastic processes and the related random variable, the number of times the trajectory of a stochastic process crosses an arbitrary level u during the time interval $[0, T]$ has considerable importance and forms the basis of extreme value theory for stochastic processes. We refer the reader to Cramer and Leadbetter(1962), Leadbetter *et al.*(1983) and Albin(1990, 2001) for the general treatment of extreme value theory for stochastic processes and level crossing problems.

Let $X(t)$, $t \geq 0$ be a continuous time, strictly stationary stochastic process with almost surely continuous sample paths $x(t)$. $x(t)$ is said to have an upcrossing of u at the point $t_0 > 0$, if for some $\epsilon > 0$, $x(t) \leq u$ in the interval $(t_0 - \epsilon, t_0)$ and $x(t) > u$ in $(t_0, t_0 + \epsilon)$. Here we assume that the sample paths are not identically equal to u in any subinterval with probability 1. Downcrossings of the level u can similarly be defined with obvious changes. We denote by the random variable $N_u(I)$, the number of upcrossings of the level u by the process $X(t)$ in the time interval I. We will also write $N_u(T) = N_u((0, T])$, and in particular $N_u(1) = N_u((0, 1])$. This random variable plays the fundamental role in studying the level crossing problems of stochastic processes. Much is known on the random variable $N_u(I)$, particularly for Gaussian processes. For example, under general conditions, the mean number of upcrossings of the level u in the unit interval $(0, 1]$ is given by

$$E(N_u(1)) \;=\; \lim_{q \to 0} J_q(u) \;=\; \int_0^\infty z\, p(u, z)\, dz \;, \tag{1.13}$$

where for arbitrarily small $q > 0$, $J_q(u) = \frac{1}{q} P\big(X(0) \leq u < X(q)\big)$ and $g_q(u, z)$ is the joint density of $X(0)$ and $\frac{1}{q}\big(X(q) - X(0)\big)$ such that $p(u, z) = \lim_{q \to 0} g_q(u, z)$. In most cases the limiting density $p(u, z)$ is the joint density of $X(t)$ and its derivative $X'(t)$ calculated at $t = 0$. In this case,

$$E(N_u(1)) \;=\; p(u) \int_0^\infty z\, p(z|u)\, dz \;=\; p(u)\, E\big(\max\{0, X'(0)\} \,|\, X(0) = u\big) \;, \tag{1.14}$$

where $p(u)$ and $p(z|u)$ are respectively the density of $X(t)$ and the conditional density of the derivative $X'(t)$ given $X(t) = u$, again calculated at $t = 0$. Hence, the expected number of upcrossings is given in terms of the density of $X(0)$ and the average positive slope of the sample path at u.

The expected number of upcrossings of a Gaussian process is totally characterized by the behavior of its covariance function at origin. If $r(\tau)$ is the covariance function of the process $X(t)$ such that as $\tau \to 0$,

$$r(\tau) \;=\; 1 - \frac{\lambda_2 \tau^2}{2} + o(\tau^2) \;,$$

then the expected number of upcrossings is given

$$E(N_u(1)) \;=\; \frac{1}{2\pi}\,\lambda_2^{1/2} \exp\!\left(-\frac{u^2}{2}\right). \tag{1.15}$$

Here, $\lambda_2 = r''(0)$ is called the second spectral moment, assumed to be finite and (1.15) is the well known Rice formula. We note that in extreme value theory, a more general class of Gaussian processes with covariance function of the type

$$r(\tau) \;=\; 1 - C|\tau|^\alpha \;=\; O(|\tau|^\alpha) \;, \qquad \tau \to 0 \;,$$

where $0 < \alpha \leq 2$ are considered. This class includes the Gaussian processes with differentiable sample paths, that is, Gaussian process with finite second spectral moments and consequently with finite number of upcrossings. In general, when $\alpha < 2$, the process is nondifferentiable and consequently, has infinitely many upcrossings in any finite interval.

Although the expected number of upcrossings gives quite a lot of information on the oscillatory behavior of the process, more can be learned from the higher moments of upcrossings or indeed from its asymptotic probability distribution. Second upcrossing moment given by

$$E\big(N_u(I)\,(N_u(I)-1)\big) \;=\; \int_0^\infty\!\!\int_0^\infty z_1 z_2\, p(u,u,z_1,z_2)\; dz_1\, dz_2 \;, \tag{1.16}$$

where $p(u,u,z_1,z_2)$ is the joint density of $(X(t_1), X(t_2), X'(t_1), X'(t_2))$ calculated at $X(t_1) = u$, $X(t_2) = u$, plays particularly important role in obtaining limiting results for the extremal behavior of the process. For example, as $u \to \infty$ it can be shown that

$$\begin{aligned} 1 - E\big(N_u(T)\big) + o(1) \;&\leq\; P\Big(\max_{t\in(0,T]} X(t) \leq u\Big) \\ &\leq\; 1 - E\big(N_u(T)\big) + E\big(N_u(T)\,(N_u(T)-1)\big), \end{aligned}$$

from which one can obtain the asymptotic expression for the maximum of a stochastic process over fixed and increasing intervals. See for example, Leadbetter(1978) and Turkman and Walker(1984). It is possible to obtain more complete results on upcrossing events other than their moments. For a given level u, let $\mu(u) = E(N_u(1))$ be the finite mean number of u-crossings per unit time by the process $X(t)$. If we look at the number of u-upcrossings of the process over an interval $[0,T]$ as $T \to \infty$, then almost surely this number would diverge to ∞.

However, if we increase the level u as a function of the the increasing time interval T in such a way that that $T\mu(u) \to \tau$, for some fixed $0 < \tau < \infty$, as $T \to \infty$, then it may be possible to obtain many nice limiting properties. Indeed, Let $N_T(\cdot)$ be the time normalized point process of u-crossings defined by

$$N_T(B) \,=\, N_u(TB) \,=\, \sharp\Big\{u\text{-crossings by } X(t);\;\; t/T \in B\Big\}$$

for any Borel set in $[0,1]$. Thus, N_T has a point at t, if $X(t)$ has an u-upcrossing at tT. Then it is known that under quite general conditions, this point process converges to a homogeneous Poisson process with intensity τ. These results are called complete convergence theorems, since one can obtain many interesting asymptotic results from this basic convergence. For example, the asymptotic distribution of the maximum of the process over increasing intervals as well as the asymptotic distribution of the upper order statistics of the process can be recovered from such basic results. See for example Leadbetter *et al.*(1983) and Resnick(1987, 2007) for convergence of point processes related to exceedances and upcrossings.

The corresponding results for Gaussian processes are well known. Let $X(t)$ be a stationary Gaussian process with covariance function $r(\tau)$ such that

1. as $h \to 0,\; r(h) = 1 - \frac{\lambda_2}{2}h^2 + o(h^2)$;

2. as $h \to \infty,\; r(h)\log h \to 0$.

Let u and T tend to infinity in such a way that $T\mu \sim \tau$, where, $\mu = \frac{1}{2\pi}\,\lambda_2^{1/2}\exp(-u^2/2)$ is the expected number of upcrossings in the unit interval. Then the time normalized point process N_T of u-upcrossings converges in distribution to a Poisson process with intensity τ on the positive real line. For processes other than Gaussian processes, asymptotic results of similar type are very difficult to obtain. We refer the reader to Albin(2001) on asymptotic results on upcrossings by many non-Gaussian processes such as Markov jump processes, α-stable processes and quadratic functionals of Gaussian processes. For specific results on streams of upcrossings by random coefficient polynomials, see Farahmand(1998). See Scotto and Turkman(2005) for similar weak convergence of point processes of u-upcrossings of finite order Volterra series expansions, although such polynomials are quite different in nature than the random coefficient polynomials defined in (1.12).

In section 2, we will look at the point processes of u-upcrossings of certain types of trigonometric polynomials and show Poisson nature of the limiting process.

2. u-UPCROSSINGS OF RANDOM TRIGONOMETRIC POLYNOMIALS

Assume that x_t, $t = 1, 2, ...n$ are n consecutive observations generated by a stationary time series X_t with 0 mean and finite variance. The periodogram of the observations defined by

$$I_{n,X}(\omega) \;=\; \frac{2}{n}\left|\sum_{t=1}^{n} x_t\, e^{i\omega t}\right|^2$$

plays an important role in the inference for spectral distribution function. In particular, crucial tests of hypotheses regarding jumps in the spectral distribution function depend on the statistics

$$M_{n,I} \;=\; \max_{\omega\in[0,\pi]} I_{n,X}(\omega)\ ,$$

and

$$M_{n,K} \;=\; \max_{\omega\in[0,\pi]} K_{n,X}(\omega)\ ,$$

where,

$$K_{n,X}(\omega) \;=\; \frac{I_{n,X}}{2\pi\,\hat{f}(\omega)}\ ,$$

and $\hat{f}(\omega)$ is a suitable estimator of the spectral density function. Hence the asymptotic distribution of $M_{n,I}$ and $M_{n,K}$ have considerable interest. Since,

$$I_{n,X} \;=\; X_n^2(\omega) + Y_n^2(\omega)\ ,$$

where

$$X_n(\omega) \;=\; \sqrt{\frac{n}{2}}\,\sum_{t=1}^{n} x_t \cos\omega t\ ,$$

and

$$Y_n(\omega) \;=\; \sqrt{\frac{n}{2}}\,\sum_{t=1}^{n} x_t \sin\omega t\ ,$$

it is clear that the study of the asymptotic distribution of the maximum periodogram ordinate in $\omega \in [0, \pi]$ can be done by studying similar asymptotic results for $X_n(\omega)$ and $Y_n(\omega)$. Both $X_n(\omega)$ and $Y_n(\omega)$ are trigonometric polynomials with random coefficients. Periodogram is also a trigonometric polynomial since it can be written in the form

$$I_{n,X}(\omega) \;=\; 2\sum_{k=-n}^{n} c_k\, e^{ik\omega}\ ,$$

where $c_k = \frac{1}{n}\sum_{t=1}^{n-|k|} x_t\, x_{t+|k|}$.

The asymptotic distributions for $M_{n,X} = \max_{\omega\in[0,\pi]} X_n(\omega)$ as well as of $M_{n,Y} = \max_{\omega\in[0,\pi]} Y_n(\omega)$ and $M_{n,I}$ are given in Turkman and Walker(1984), under the assumption that x_t are iid, normal random variables. These results are then extended to $M_{n,K}$, when X_t is a stationary time series.

Note that if X_t is a Gaussian time series, then both $X_n(\omega)$ and $Y_n(\omega)$ are continuous parameter Gaussian processes defined over the fixed interval $\omega \in [0,\pi]$. As such, it may be tempting to obtain all desired results on u-upcrossings based on the well known theory for Gaussian processes. However, the second spectral moments of the processes $X_n(\omega)$ and $Y_n(\omega)$ are given by

$$r''(0) \;=\; -\frac{n^2}{3}\big(1+O(1/n)\big)\;,$$

and for finite n, both processes are differentiable having finite number of upcrossings in $\omega \in [0,\pi]$. However, as the sample size n increases, these processes have sample paths that oscillate wildly, having infinitely many upcrossing of any finite level u in any finite subset of $\omega \in [0,\pi]$ with probability one. This is the fundamental reason why periodogram appears as an inconsistent estimator of the spectral density function. Hence, known results on u-upcrossings for Gaussian processes cannot be used in a straightforward fashion. In order to get meaningful asymptotic results for the level crossings of the polynomials $X_n(\omega)$ and $Y_n(\omega)$ as $n \to \infty$, one has to study the u-upcrossings for levels u which increase to infinity in a controlled fashion as $n \to \infty$. We refer the reader to Turkman and Walker(1984, 1991) for details in obtaining the first two moments of the u-crossings by such processes for appropriately chosen level u and the consequent asymptotic distribution of the maxima of these polynomials in the interval $[0,\pi]$. Here, we will derive the asymptotic Poisson character of the u-upcrossings of these trigonometric polynomials, for suitably increasing levels $u = u(n)$, as $n \to \infty$.

2.1. Poisson character of u-upcrossings

Let

$$X_n(\omega) \;=\; \sqrt{\frac{n}{2}}\,\sum_{t=1}^{n} x_t \cos\omega t\;,$$

and

$$Y_n(\omega) \;=\; \sqrt{\frac{n}{2}}\,\sum_{t=1}^{n} x_t \sin\omega t\;,$$

be trigonometric polynomials, where x_t is a realization of iid standard Gaussian random variables. Let $N_X(I) = N_{u_n,X}(I)$ and $N_Y(I) = N_{u_n,Y}(I)$ be respectively the number of upcrossings of a suitable chosen level u_n by the processes $X_n(\omega)$ and $Y_n(\omega)$ in the interval $I \subset [0,\pi]$. In this section we prove the following theorem:

Theorem 2.1. *Let*

$$u_n = \frac{x}{\sqrt{2\log n}} + \sqrt{2\log n} - \frac{\log 12}{2\sqrt{2\log n}} ,$$

and let

$$\tau = \tau(x) = e^{-x} .$$

Then

$$\lim_{n\to\infty} P\big(N_X[0,\pi] = s\big) = \frac{e^{-\tau}\tau^s}{s!} , \tag{2.1}$$

$$\lim_{n\to\infty} P\big(N_Y[0,\pi] = s\big) = \frac{e^{-\tau}\tau^s}{s!} . \tag{2.2}$$

We will give the proof only for (2.1). The proof for (2.2) is similar. For ease in notation, we write

$$N(I) = N_{u_n,X}(I) .$$

The proof of Theorem (2.1) is quite long and we give an outline of the proof.

Let $k > 0$ be a fixed but arbitrarily large integer and divide the interval $[0,\pi]$ into subintervals I_j, $j = 1, 2, ..., k$ such that

$$I_j = \left[\frac{\pi(j-1)}{k}, \frac{\pi j}{k}\right), \qquad j = 1, 2, .., k-1 ,$$

and

$$I_k = \left[\frac{\pi(k-1)}{k}, \pi\right] .$$

For any $\beta \in (0, 1/2)$ arbitrarily small, for every $j = 2, ..., k$ divide every I_j further into two disjoint subintervals

$$I_{j,1} = \left[\frac{\pi(j-1)}{k}, \frac{\pi j}{k}\right) ,$$

$$I_{j,2} = \left[\frac{\pi(j-\beta)}{k}, \frac{\pi j}{k}\right), \qquad 2 \le j \le k ,$$

$$I_{k,2} = \left[\frac{\pi(k-\beta)}{k}, \pi\right], \qquad 2 \le j \le k .$$

Special attention is paid to the interval I_1 and we divide I_1 as

$$I_{1,0} = \left[0, \frac{\pi\beta}{k}\right), \qquad I_{1,1} = \left[\frac{\pi\beta}{k}, \frac{\pi(1-\beta)}{k}\right), \qquad I_{1,2} = \left[\frac{\pi(1-\beta)}{k}, \frac{\pi}{k}\right) .$$

The proof is based on first showing that number of upcrossings over the intervals I_{j_1} are asymptotically independent and that number of upcrossings over the intervals $I_{j,2}$ are asymptotically negligible. Thus the outline of the proof is as follows:

1. For any s, approximate $P\big(N[0,\pi] \geq s\big)$ by $P\big(N\big(\bigcup_j I_{j,1}\big) \geq s\big)$.

2. Approximate $P\big(N\big(\bigcup_j I_{j,1}\big) \geq s\big)$ by $P(A_{n,s})$, where $A_{n,s}$ is the event that $N(I_{j,1}) \geq 1$ for at least s values of $j = 1, ..., k$.

3. Approximate $P(A_{n,s})$ by $P(D_{n,s})$ where $D_{n,s}$ is the event that in exactly s of the intervals $I_{j,1}$ we have $X_n(\omega) \geq u_n$ for some $\omega \in I_{j,1}$, so that $P\big(N[0,\pi] = s\big)$ is approximated by $P(D_{n,s})$.

4. Let

$$p \,=\, p_{k,\beta,\tau} \,=\, \lim_{n\to\infty} P\big(M_n(I_{j,1}) \leq u_n\big) \ ,$$

show that

$$p \,=\, \exp\!\left(-\frac{(1-\beta)\,\tau}{k}\right) \ ,$$

and then approximate $P(D_{n,s})$ by the binomial probability

$$\binom{k}{s} (1-p)^s \, p^{k-s} \ .$$

5. Let $\beta \to 0$, then $k \to \infty$ and use the Poisson approximation to the binomial probability to obtain the desired result.

We now give proofs for these assertions in terms of series of Lemmas.

Lemma 2.1. *For any $s = 0, 1, 2, ...$, as $n \to \infty$,*

$$0 \;\leq\; P\Big(N[0,\pi] \geq s\Big) - P\left(N\left(\bigcup_{j=1}^{k} I_{j,1}\right) \geq s\right) \;\leq\; \beta\tau + o_n(1) \ . \tag{2.3}$$

Proof: The event $\big\{N[0,\pi] \geq s\big\}$ contains the event $\big\{N\big(\bigcup_{j=1}^k I_{j,1}\big) \geq s\big\}$ and the difference is the event

$$\left\{ \bigcup_{j=1}^{k} \big(N(I_{j,2} \geq 1)\big) \,\cup\, \big(N(I_{1,0} \geq 1)\big) \right\} .$$

Hence

$$
\begin{aligned}
0 \;&\le\; P\Big(N[0,\pi] \ge s\Big) - P\bigg(N\Big(\bigcup_{j=1}^{k} I_{j,1}\Big) \ge s\bigg) \\
&=\; P\bigg[\bigcup_{j=1}^{k}\big(N(I_{j,2}\ge 1)\big) \,\cup\, \big(N(I_{1,0}\ge 1)\big)\bigg] \\
&\le\; \sum_{j=1}^{k} P\big(N(I_{j,2}\ge 1)\big) + P\big(N(I_{1,0}\ge 1)\big) \\
&\le\; \sum_{j=1}^{k-1} E\big(N(I_{j,2})\big) + E\big(N(I_{1,0})\big) + E\big(N(I_{k,2})\big)\;.
\end{aligned}
\tag{2.4}
$$

It can be shown that (see Turkman and Walker, 1984) as $n \to \infty$, for every $j = 1, ..., k-1$

$$
E\big(N(I_{j,2})\big) = \frac{\tau\beta}{k}\;.
$$

However, $E(N(I_{1,0}))$ and $E(N(I_{k,2}))$ need special attention in calculations. The reason for this extra complication is that the expected number of upcrossings are calculated as an integral with respect to the joint density of the vector $\big(X_n(\omega), X'(\omega)\big)$ and this vector has a normal density with mean 0 and covariance function given by

$$
\begin{pmatrix} 1 + r_n(2\omega) & r_n'(2\omega) \\ r_n'(2\omega) & \frac{n^2}{3} + r_n''(2\omega) \end{pmatrix}\,,
\tag{2.5}
$$

where $r_n(\omega) = \frac{1}{n}\sum_{j=1}^{n}\cos j\omega$ and $r_n'(\omega), r_n''(\omega)$ are respectively first and second order derivatives of $r_n(\omega)$ respectively. This covariance matrix tends to be singular as ω gets arbitrarily close to 0 or π. Hence $E(N(I_{j,0}))$ needs to be calculated separately over regions

$$
\begin{aligned}
R_{n,1} &= \left\{\omega \in I_{1,0}\colon\; \omega \ge \frac{\log n}{n}\right\}, \\
R_{n,2} &= \left\{\omega \in I_{1,0}\colon\; \frac{1}{nk} \le \omega \le \frac{\log n}{n}\right\}, \\
R_{n,3} &= \left\{\omega \in I_{1,0}\colon\; 0 \le \omega \le \frac{1}{nk}\right\}.
\end{aligned}
$$

It is shown in Turkman and Walker (1984) that

$$
\lim_{n\to\infty} E\big(N(I_{j,0})\big) \;=\; \begin{cases} \dfrac{\tau\beta}{k}\,, & \omega \in R_{n,1}\,, \\[2ex] 0\,, & \omega \in R_{n,2} \cup R_{n,3}\,. \end{cases}
\tag{2.6}
$$

Similar expression can be found for $E(N(I_{k,2}))$ and hence from (2.4) for arbitrarily large k and arbitrarily small β,

$$\sum_{j=1}^{k} E\big(N(I_{j,2})\big) + E\big(N(I_{j,0})\big) \;=\; \tau\beta\Big(1+\frac{1}{k}\Big)\,.$$

This proves the Lemma. □

Define $A_{n,s}$ to be the event that $\{N(I_{j,1}) \geq 1\}$ for at least s values of $j = 1, .., k$. Then

Lemma 2.2. *As $n \to \infty$,*

$$0 \;\leq\; P\big(N[0,\pi] \geq s\big) - P\big(A_{n,s}\big) \;\leq\; \beta\,\tau + \sum_{j=1}^{k} P\big(N(I_{j,1}) \geq 2\big) + o_n(1)\;, \tag{2.7}$$

and

$$\limsup_{n\to\infty} \sum_{j=1}^{k} P\big(N(I_{j,1}) \geq 2\big) \;\leq\; (1-\beta)\tau - k\Big(1 - \exp\Big(-(1-\beta)\frac{\tau}{k}\Big)\Big)\,. \tag{2.8}$$

Proof:

$$\begin{aligned} 0 \;&\leq\; P\Big(N\Big(\bigcup_{j=1}^{k} I_{j,1}\Big) \geq s\Big) - P(A_{n,s}) \\ &\leq\; P\Big(\bigcup_{j=1}^{k} \big(N(I_{j,1}) \geq 2\big)\Big) \qquad (2.9) \\ &\leq\; \sum_{j=1}^{k} P\big(N(I_{j,1}) \geq 2\big)\;. \end{aligned}$$

Now combining this inequality with the inequality (2.4), we get (2.7). To prove (2.8), we proceed as follows: First note that for any $j = 1, ..., k$

$$P\big(N(I_{j,1}) \geq 2\big) \;\leq\; E\big(N(I_{j,1})\big) - P\big(N(I_{j,1}) \geq 1\big)\;. \tag{2.10}$$

also

$$0 \;\leq\; P\big(M_n(I_{j,1}) > u_n\big) - P\big(N(I_{j,1}) \geq 1\big) \;\leq\; P\Big(X_n\Big(\frac{\pi(j-1)}{k}\Big) \geq u_n\Big)\,,$$

which implies that

$$P\big(M_n(I_{j,1}) > u_n\big) - P\Big(X_n\Big(\frac{\pi(j-1)}{k}\Big) \geq u_n\Big) \;\leq\; P\big(N(I_{j,1}) \geq 1\big)\;,$$

so that from (2.10),

$$P\big(N(I_{j,2}) \geq 2\big) \;\leq\; (1-\beta)\frac{\tau}{k} - P\big(M_n(I_{j,1} > u_n)\big) + P\Big(X_n\Big(\frac{\pi(j-1)}{k}\Big) \geq u_n\Big). \tag{2.11}$$

Now, for any $\omega \in I_{j,1}$, $X_n(\omega) \sim N(0, 1 + r_n(2\omega))$, thus as $n \to \infty$,

$$P\Big(X_n\Big(\frac{\pi(j-1)}{k}\Big) \geq u_n\Big) \;=\; o_n(1)\ ,$$

since, from Turkman and Walker(1984) we have

$$\lim_{n\to\infty} P\big(M_n(I_{j,1}) > u_n\big) \;=\; 1 - \exp\Big(-\frac{(1-\beta)\tau}{k}\Big)\ .$$

Now the proof is complete by combining (2.10) with (2.11). □

Denote by $C_{n,s}$, the event that $X_n(\omega) > u_n$ in at least $s = 1, ..., k$ of the intervals $I_{j,1}$ for some $\omega \in I_{j,1}$.

Then

Lemma 2.3. *As $n \to \infty$,*

$$0 \;\leq\; P(C_{n,s}) - P(A_{n,s}) \;=\; o_n(1)\ . \tag{2.12}$$

Proof: The event

$$A = \Big\{N(I_{j,1}) \geq 1\Big\}$$

is contained in the event

$$B = \Big\{X_n(\omega) \geq u_n,\ \text{for some } \omega \in I_{j,1}\Big\}$$

and the difference of these events are given by

$$B - A \;=\; A^c \cap B \;=\; \Big\{X_n\Big(\frac{\pi(j-1)}{k}\Big) > u_n\Big\}\ .$$

Hence it follows from the definitions of the events $C_{n,s}$ and $A_{n,s}$ that as $n \to \infty$,

$$\begin{aligned} 0 \;&\leq\; P(C_{n,s}) - P(A_{n,s}) \\ &\leq\; P\Big(\bigcup_{k=1}^{k}\Big(X_n\Big(\frac{\pi(j-1)}{k}\Big) > u_n\Big)\Big) \\ &\leq\; \sum_{k=1}^{k} P\Big(X_n\Big(\frac{\pi(j-1)}{k}\Big) > u_n\Big) \;=\; o_n(1)\ . \end{aligned} \tag{2.13}$$

Clearly for any $s < k$, $C_{n,s+1} \subset C_{n,s}$. Let $D_{n,s} = C_{n,s} - C_{n,s+1} = C^c_{n,s+1} \cap C_{n,s}$. $D_{n,s}$ is the event that $X_n(\omega) > u_n$ in exactly s of the k intervals and

$$P(D_{n,s}) \;=\; P(C_{n,s}) - P(C_{n,s+1})\ .$$ □

Lemma 2.4.

$$\limsup_{n\to\infty}\Big|P\big(N[0,\pi]=s\big)-P\big(D_{n,s}\big)\Big| \;\le\; \tau - k\bigg(1-\exp\Big(-\frac{(1-\beta)\tau}{k}\Big)\bigg)\,. \tag{2.14}$$

Proof: From (2.12), for any $s<k$, $0\le P(C_{n,s})-P(A_{n,s})=o_n(1)$, therefore, as $n\to\infty$,

$$\Big|P(A_{n,s})-P(A_{n,s+1})-P(D_{n,s})\Big| \;\le\; o_n(1)\;.$$

Hence from (2.7), for any $s<k$,

$$0 \;\le\; P\big(N[0,\pi]\ge s\big)-P\big(D_{n,s}\big) \;\le\; \beta\tau+\sum_{j=1}^{k}P\big(N(I_{j,1})\ge 2\big)+o_n(1)\;, \tag{2.15}$$

and the lemma follows from (2.8).

Let $M_j=\{M_n(I_{j,1})\le u_n\}$ and M_j^c be the compliment of M_j. Let

$$P_{n,j} \;=\; P\big(M_n(I_{j,1})\le u_n\big)\;.$$

We know from Turkman and Walker(1984) that

$$\lim_{n\to\infty}P_{n,j} \;=\; \exp\bigg(-\frac{(1-\beta)\tau}{k}\bigg)=p\,,\qquad \text{say}\,. \qquad \square$$

Lemma 2.5.

$$\limsup_{n\to\infty}\bigg|P(D_{n,s})-\binom{k}{s}(1-p)^s\,p^{k-s}\bigg| \;=\; 0\;. \tag{2.16}$$

Proof:

$$D_{n,s}=\bigcup\big(M_{i_1}^c.\,M_{i_2}^c\ldots M_{i_s}^c.\,M_{i_s+1}\ldots M_{i_k}\big)\;,$$

where the union is taken over all combinations of distinct integers with $i_1<i_2<\ldots<i_k$. Here, we omit the intersection signs, replacing them with ".". We first start by looking at the probability

$$P\big(M_{i_1}^c.\,M_{i_2}^c\ldots M_{i_m}^c.\,M_{i_1}\ldots M_{i_t}\big)\;, \tag{2.17}$$

where m and t are integers such that $m+t\le k$. When $m=0$, (2.17) is equal to

$$P\big(M_{i_1}\ldots M_{i_t}\big) \;=\; P\bigg(\bigcap_{k=1}^{t}\big(M_n(I_{j,1})\le u_n\big)\bigg).$$

It follows from Lemma 2.6 of Turkman and Walker (1984) that for any $t\le k$

$$\limsup_{n\to\infty}\Big|P\big(M_{i_1}\ldots M_{i_t}\big)-p^t\Big| \;=\; 0\;.$$

Now assume that for an $m \geq 1$ and for all $t \leq k-(m-1)$

$$\limsup_{n\to\infty} \left| P\big(M_{i_1}^c . M_{i_2}^c \ldots M_{i_{m-1}}^c . M_{i_1} \ldots M_{i_t}\big) - (1-p)^{m-1} p^t \right| = 0 \ . \tag{2.18}$$

We now show that

$$\limsup_{n\to\infty} \left| P\big(M_{i_1}^c . M_{i_2}^c \ldots M_{i_m}^c . M_{i_1} \ldots M_{i_t}\big) - (1-p)^{m} p^t \right| = 0 \ , \tag{2.19}$$

and the proof will be complete by induction:

$$\begin{aligned} \Big\{ M_{i_1}^c . M_{i_2}^c \ldots M_{i_{m-1}}^c . M_{i_1} \ldots M_{i_t} \Big\} - \Big\{ M_{i_1}^c . M_{i_2}^c \ldots M_{i_m}^c . M_{i_1} \ldots M_{i_t} \Big\} = \\ = \Big\{ M_{i_1}^c . M_{i_2}^c \ldots M_{i_{m-1}}^c . M_{i_1} \ldots M_{i_t} \Big\} \cap \Big\{ M_{i_1}^c . M_{i_2}^c \ldots M_{i_m}^c . M_{i_1} \ldots M_{i_t} \Big\} \\ = \Big\{ M_{i_1}^c . M_{i_2}^c \ldots M_{i_{m-1}}^c . M_{i_m} . M_{i_1} \ldots M_{i_t} \Big\} \ . \end{aligned}$$

Hence,

$$P\big(M_{i_1}^c . M_{i_2}^c \ldots M_{i_m}^c . M_{i_1} \ldots M_{i_t}\big) =$$

$$= P\big(M_{i_1}^c . M_{i_2}^c \ldots M_{i_{m-1}}^c . M_{i_m} . M_{i_1} \ldots M_{i_t}\big) - P\big(M_{i_1}^c . M_{i_2}^c \ldots M_{i_{m-1}}^c . M_{i_m} . M_{i_1} \ldots M_{i_t}\big)$$

From the assumption (2.18) we have

$$\limsup_{n\to\infty} \left| P\big(M_{i_1}^c . M_{i_2}^c \ldots M_{i_{m-1}}^c . M_{i_1} \ldots M_{i_t}\big) - (1-p)^{m-1} p^t \right| = 0 \ ,$$

and

$$\limsup_{n\to\infty} \left| P\big(M_{i_1}^c . M_{i_2}^c \ldots M_{i_{m-1}}^c . M_{i_m} . M_{i_1} \ldots M_{i_t}\big) - (1-p)^{m-1} p^{t+1} \right| = 0 \ ,$$

so that (2.19) follows immediately. Choosing $t = m-k$, we get

$$\limsup_{n\to\infty} \left| P\big(M_{i_1}^c . M_{i_2}^c \ldots M_{i_m}^c . M_{i_{m+1}} . M_{i_1} \ldots M_{i_k}\big) - (1-p)^{m} p^{k-m} \right| = 0$$

and the lemma follows immediately from induction. □

The proof of the theorem 2.1 now follows from lemmas 1–5 by first letting $\beta \to 0$, then $k \to \infty$. First note that

$$\limsup_{n\to\infty} \left| P\big(N[0,\pi] = s\big) - \binom{k}{s} (1-p)^s p^{k-s} \right| \leq \tau - k\left(1 - \exp\Big(-\frac{(1-\beta)\tau}{k}\Big)\right) \ , \tag{2.20}$$

where,

$$p = \exp\left(-\frac{(1-\beta)\tau}{k}\right) \ .$$

Thus,

$$\limsup_{n\to\infty} P\big(N[0,\pi] = s\big) \leq \binom{k}{s} (1-p)^s p^{k-s} + \tau - k\left(1 - \exp\Big(-\frac{(1-\beta)\tau}{k}\Big)\right) \ , \tag{2.21}$$

so that letting $\beta \to 0$,

$$\limsup_{n\to\infty} P\big(N[0,\pi] = s\big) \;\le \tag{2.22}$$
$$\le \binom{k}{s} \big(1-\exp(-\tau/k)\big)^s \big(\exp(-\tau/k)\big)^{k-s} + \tau - k\Big(1-\exp\Big(-\frac{(1-\beta)\tau}{k}\Big)\Big) \;.$$

Now let $k \to \infty$. Then $k(1-\exp(-\tau/k)) \to \tau$, and by Poisson approximation to binomial we get

$$\limsup_{n\to\infty} P\big(N[0,\pi] = s\big) \;\le\; \frac{e^{-\tau}\tau^s}{s!} \;.$$

Similarly we can show that

$$\liminf_{n\to\infty} P\big(N[0,\pi] = s\big) \;\ge\; \frac{e^{-\tau}\tau^s}{s!} \;,$$

and this completes the proof.

It is possible to obtain the following similar asymptotic result for the periodogram, although proofs are slightly more tedious and we omit the proof.

Theorem 2.2. *Let* $u_n = 2(x + \log n + \frac{1}{2}\log\log n - \frac{1}{2}\log\frac{3}{\pi})$. *Then the number of* u_n*-upcrossings* $N_{u_n,I}[0,\pi]$ *of the periodogram in the interval* $[0,\pi]$ *is asymptotically Poisson, in the sense that*

$$\lim_{n\to\infty} P\big(N_{u_n,I}[0,\pi] = s\big) \;=\; \frac{e^{-\tau}\tau^s}{s!} \;, \qquad s = 0,1,\ldots \;,$$

where $\tau = \tau(x) = e^{-x}$.

Asymptotic results given in Theorems 2.1 and 2.2 are very useful and many convergence results for upper order statistics can be recovered from these basic results. For example, if

$$M_n[0,\pi] = \max_{\omega\in[0,\pi]} X_n(\omega) \;,$$

then

$$\Big\{M_n(0,\pi] \le u_n\Big\} \;=\; \Big\{N_X[0,\pi] = 0\Big\} \;,$$

and consequently,

$$\lim_{n\to\infty} P\bigg(M_n \le \frac{x}{\sqrt{2\log n}} + \sqrt{2\log n} - \frac{\log 12}{2\sqrt{2\log n}}\bigg) \;=\; e^{-e^{-x}} \;,$$

which was proved in Turkman and Walker(1984) based on calculating the first two moments of the u-upcrossings.

ACKNOWLEDGMENTS

Research supported by projects Pest-UE/MAT/UI0006/2011 and PTDC/MAT/118335/2010 financed by FCT-Portugal.

REFERENCES

[1] Albin, J.M.P. (2001). On extremes and streams of upcrossings, *Stochastic processes and their applications*, **94**, 271–300.

[2] Albin, J.M.P. (1990). On extremal theory for stationary processes, *Annals of probability*, **18** 92–128.

[3] Bharucha-Reid, A.T. and Sambandham, M. (1986). *Random Polynomials*, Academic Press.

[4] Brant, A. (1986). The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients, *Advances in Applied Probability*, **18**, 211–220.

[5] Cramer, H. and Leadbetter, M.R. (1967). *Stationary and Related Stochastic Processes*, J. Wiley, New York.

[6] de Haan, L.; Resnick, S.I.; Rootzén, H. and de Vries, C.G. (1989). Extremal behavior of solutions to a stochastic difference equation with applications to ARCH processes, *Stochastic Processes and their Applications*, **32**, 213–224.

[7] Farahmand, K. (1998). *Topics on Random Polynomials*, Longman.

[8] Leadbetter, M.R.; Lindgren, G. and Rootzen, H. (1983) *Extremes and Related Properties of Random Sequences and Processes*, Springer-Verlag, New York.

[9] Nisio, M. (1960). On polynomial approximation for strictly stationary processes, *J. Mathematical Society of Japan*, **12**, 207–226.

[10] Resnick, S.I. (2007). *Heavy-tail Phenomena*, Springer, New York.

[11] Resnick, S.I. (2007). *Extreme Values, Regular Variation and Point Processes*, Springer Verlag, New York.

[12] Scotto, M.G. (2007). Extremes for solutions to stochastic difference equations with regularly varying tail, *REVSTAT*, **5**, 229–247.

[13] Scotto, M.G. and Turkman, K.F. (2002). On the extremal behavior of subsampled solutions of stochastic difference equations, *Portugaliae Mathematica*, **59**, 267–283.

[14] Scotto, M.G. and Turkman, K.F. (2005). Extremes of Volterra series expansions with heavy-tailed innovations, *Nonlinear Analysis — Theory Methods Applications*, **63**, 106–122.

[15] TERDIK, G. (1999). *Bilinear Stochastic Models and Related Problems of Nonlinear Time Series Analysis*, Springer-Verlag, New York.

[16] TURKMAN, K.F. and AMARAL TURKMAN, M.A. (1997) Extremes of bilinear time series models, *Journal of Time Series Analysis*, **18**, 305–320.

[17] TURKMAN, K.F. and WALKER, M. (1984). The Asymptotic Distribution of Maxima Of Trigonometric Polynomials with Random Coefficients, *Advances in Applied Probability*, **16**, 819–842.

[18] TURKMAN, K.F. and WALKER, M. (1990). A Stability Result for the Periodogram, *The Annals of Probability*, **18**, 1765–1783.

[19] VERVAAT, W. (1979). On a stochastic difference equation and a representation of non-negative infinitely divisible random variables, *Advances in Applied Probability*, **11**, 750–783.

REVSTAT – Statistical Journal
Volume 12, Number 2, June 2014, 157–168

EXTREMES OF PERTURBED BIVARIATE RAYLEIGH RISKS

Authors: Enkelejd Hashorva
– Department of Actuarial Science, University of Lausanne, Switzerland
enkelejd.hashorva@unil.ch

Saralees Nadarajah
– School of Mathematics, University of Manchester, Manchester M13 9PL, UK
mbbsssn2@manchester.ac.uk

Tibor K. Pogány
– Faculty of Maritime Studies, University of Rijeka, 51000 Rijeka, Croatia
poganj@pfri.hr

Received: January 2013 Revised: April 2013 Accepted: May 2013

Abstract:

- We establish first an asymptotic expansion for the joint survival function of a bivariate Rayleigh distribution, one of the most popular probabilistic models in engineering. Furthermore, we show that the component-wise maxima of a Hüsler–Reiss triangular array scheme of independent perturbed bivariate Rayleigh risks converges to a bivariate Hüsler–Reiss random vector.

Key-Words:

- *asymptotic independence; Gumbel max-domain of attraction; Hüsler–Reiss distribution; Rayleigh distribution; triangular arrays.*

AMS Subject Classification:

- 62E20.

1. INTRODUCTION

Let $(X_1, Y_1), ..., (X_N, Y_N)$ be independent bivariate Gaussian random vectors with $N(0, 1)$ distributed marginals and correlation coefficient $\rho \in (-1, 1)$. We define a bivariate Rayleigh random vector (risk), (U_m, V_m), by

$$U_m = \sum_{i=1}^{N} \Big(X_i - \sum_{i=1}^{N} X_i/N \Big)^2, \text{ and } V_m = \sum_{i=1}^{N} \Big(Y_i - \sum_{i=1}^{N} Y_i/N \Big)^2,$$

where $m := N - 1$. Basic distributional properties of bivariate Rayleigh random vectors are derived in Nadarajah [23]. In view of equation (3) in Nadarajah [23], the joint probability density function (pdf) of (U_m, V_m), $m \geq 1$ is

$$\begin{aligned} h(u,v) \;=\; & \frac{(uv)^{m/2-1}}{\Gamma(m/2)\,(2\widetilde{\varrho})^{m/2}} \exp\Big(-\frac{u+v}{2\widetilde{\varrho}}\Big) \\ & \cdot {}_0F_1\left(;m/2;\rho^2 uv/\,(2\widetilde{\varrho})^2\right), \quad \forall u, v \in (0,\infty), \end{aligned} \tag{1.1}$$

where $\rho \in (-1, 1)$, $\widetilde{\varrho} := 1 - \rho^2$ and

$${}_0F_1\,(;a;x) = \sum_{k=0}^{\infty} \frac{1}{(a)_k} \frac{x^k}{k!}$$

denotes a hypergeometric function, where $(e)_k = e(e+1)\cdots(e+k-1)$ denotes the ascending factorial.

The distribution given by the joint pdf (1.1) is known as the *bivariate Rayleigh distribution*. It has received widespread applications especially in engineering. Some recent applications have included: statistics of wave groups measured in the northern North Sea (Stansell *et al.* [27]); performance analysis of system with selection combining over correlated Rician fading channels in the presence of cochannel interference (Panajotović *et al.* [24]); cochannel interference effect on bit error probability performance of switch and stay combining receiver in correlated Rician fading (Panajotović *et al.* [25]).

The bivariate Rayleigh distribution has also been used to model extreme values, for example, with respect to depth-limited extreme wave heights in a sea state (Méndez and Castanedo [21]), reliability assessment of marine structures (Leira and Myrhaug [17], Leira *et al.* [18]), and asymptotic capacity analysis in point-to-multipoint cognitive radio networks (Ji and Chen [14]). But the asymptotic distribution of the extreme values of (U_m, V_m) has not been known. The principal aim of this short note is to establish the limiting max-stable distribution of (U_m, V_m).

An important max-stable multivariate distribution related to our results is the Hüsler–Reiss distribution due to Hüsler and Reiss [13]. In a bivariate setting,

Hüsler–Reiss distribution has the joint cumulative distribution function (cdf)

$$H_\lambda(x,y) = \exp\left[-\Phi\left(\lambda + \frac{x-y}{2\lambda}\right)\exp(-y) \right.$$
$$\left. -\Phi\left(\lambda + \frac{y-x}{2\lambda}\right)\exp(-x)\right], \quad x,y \in \mathbb{R}, \tag{1.2}$$

where $\Phi(\cdot)$ denotes the standard normal cdf and $\lambda \in (0,\infty)$ is a parameter. For any λ, the marginal cdf's of H_λ are the Gumbel cdf's $\Lambda(x) = \exp\{-\exp(-x)\}$, $x \in \mathbb{R}$.

The parameter λ has a nice representation and comes naturally in the setup of Gaussian triangular arrays. Roughly speaking, if $\rho_n \in (-1,1)$ is the correlation coefficient of a bivariate triangular array, then under the Hüsler–Reiss condition

$$\lim_{n\to\infty}(1-\rho_n)\ln n = \lambda^2 \in (0,\infty),$$

the cdf H_λ appears as the limiting distribution of the normalized maxima.

Hüsler–Reiss distribution has received widespread applications. Hüsler–Reiss distribution arises not only as the limiting max-stable distribution of the componentwise maxima of Gaussian random vectors, but it arises also as the limiting max-stable distribution of the componentwise maxima of random vectors having chi-square, elliptically symmetric and other distributions, see Hashorva [10], Frick and Reiss [8] and Hashorva *et al.* [11].

Some applications of Hüsler–Reiss distribution have included: models for environmental data (Joe [15]); portfolio risk measurement (Bouyé [2]); extremal dependence of multivariate catastrophic losses (Lescourret and Robert [19], Haug *et al.* [12]); inference for bivariate survival data (Ding and Wang [4]); models for spatial extremes (Smith and Stephenson [26]); spatial extreme fields (Bacro *et al.* [1]); models for extremes observed in space and time (Davis *et al.* [3]); multivariate value at risks for operational risk capital computation (Guegan and Hassani [9]); extremal discriminant analysis (Manjunath *et al.* [20]); multiasset derivatives and joint distributions of asset prices (Molchanov and Schmutz [22]). Important recent contributions and insights concerning the Hüsler–Reiss distribution can be found in Kabluchko [16] and Engelke *et al.* [5, 6, 7].

It follows from (1.1) that both U_m and V_m are chi-squared random variables with m degrees of freedom. Let G_m denote the cdf's of U_m and V_m. They belong to the Gumbel max-domain of attraction with scaling function $w(t) = 1/2$, i.e.,

$$\lim_{x\to\infty}\frac{1-G_m\left(x+s/w(x)\right)}{1-G_m(x)} = \exp(-s), \quad s \in \mathbb{R}.$$

Equivalently,

$$\lim_{n\to\infty}\sup_{s\in\mathbb{R}}\left|\left(G_m\left(a_n s + b_n\right)\right)^n - \Lambda(s)\right| = 0$$

with constants

$$a_n = 2, \quad b_n = 2\ln n + (m-2)\ln\ln n - 2\ln\Gamma\left(\frac{m}{2}\right), \quad n > 1.$$

As in Hüsler and Reiss [13] we shall consider a triangular array setup, which is of interest when the components are asymptotically independent. In the Gaussian framework, the asymptotic independence of the components is well known, i.e.,

$$\lim_{u\to\infty} \frac{\mathbb{P}\left(X_1 > u, Y_1 > u\right)}{\mathbb{P}\left(X_1 > u\right)} = 0$$

for any $\rho \in (-1, 1)$. In view of Hashorva *et al.* [11], U_m and V_m are asymptotically independent for any $\rho \in (-1, 1)$. So, we have

$$\lim_{n\to\infty} n\mathbb{P}\left(U_m > x + b_n, V_m > y + b_n\right) = 0. \tag{1.3}$$

Our first result below presented in Theorem 2.1 gives the exact rate of convergence to zero claimed in (1.3). In the light of the aforementioned paper, the component-wise maxima of bivariate triangular arrays of Rayleigh risks is attracted to the Hüsler–Reiss distribution. Indeed, in order to see that let (U'_m, V'_m) be another bivariate random vector defined by the stochastic representation

$$U'_m = \sum_{i=1}^{m} X_i^2, \quad V'_m = \sum_{i=1}^{m} Y_i^2$$

and further

$$U_m + N\left(\overline{X}_N\right)^2 = U'_{m+1}, \quad V_m + N\left(\overline{Y}_N\right)^2 = V'_{m+1}, \quad N = m+1.$$

Moreover, (U_m, V_m) is independent of $(\overline{X}_N, \overline{Y}_N)$, and $(\sqrt{N}\overline{X}_N, \sqrt{N}\overline{Y}_N)$ has the same distribution as (X_1, Y_1), implying the equality in distribution

$$(U_m, V_m) \stackrel{d}{=} \left(U'_m, V'_m\right).$$

Consequently, in view of Hashorva *et al.* [11] the asymptotic behavior of the component-wise maxima of a Hüsler–Reiss triangular array scheme of bivariate Rayleigh risks is known.

In Section 2, we establish the rate of convergence to zero for the joint survival function $\mathbb{P}\left(U_m > x + b, V_m > y + b\right)$ as b tends to infinity. Then we consider a perturbation of Rayleigh risks and derive the limiting distribution of bivariate maxima of triangular arrays of such risks, which turns out to be the bivariate Hüsler–Reiss distribution. All of the proofs are provided in Section 3.

2. MAIN RESULTS

Our first result derives the exact tail asymptotic behavior of the joint survival function of two bivariate Rayleigh risks, which in particular implies (1.3).

Theorem 2.1. With the notation as in Section 1, for any x, y reals and $\rho \in (-1,1)$, we have

$$\mathbb{P}\left(U_m > x+b, V_m > y+b\right) = \frac{\sqrt{2}|\rho|^{(1-m)/2}\widetilde{\varrho}^{3/2}}{\sqrt{\pi}} b^{(m-3)/2} \exp\left(-\frac{b}{1+|\rho|}\right) \cdot \left[1+O\left(b^{-1}\right)\right]$$

as $b \to \infty$.

A direct consequence of Theorem 2.1 is that U_m and V_m are asymptotically independent for any $\rho \in (-1,1)$.

Our second result is concerned with perturbed Rayleigh risks: in order to motivate the definition of such risks, recall that we can write

$$Y_i \stackrel{d}{=} \rho X_i + \sqrt{1-\rho^2} Z_i, \quad 1 \le i \le N$$

with $X_i, Z_i, i \le N$ independent $N(0,1)$ risks. Since in the triangular array framework introduced in Hüsler and Reiss [13] the correlation coefficient $\rho = \rho_n$ tends to one as $n \to \infty$, we see that the base risk is X_i, $i \le N$ and Z_i plays the role of a perturbation. Since as mentioned in Section 1, the case of Rayleigh risks is already dealt with in Hashorva *et al.* [11], we consider the asymptotic distribution of component-wise maxima for triangular arrays of perturbed independent Rayleigh risks. Therefore, we introduce next (X_i, Y_i), $i \ge 1$ with the stochastic representation

$$(X_i, Y_i) \stackrel{d}{=} \left(X, \rho X + \sqrt{1-\rho^2} Z\right), \tag{2.1}$$

where X is a base random variable independent of $Z \sim N(0,1)$. Clearly, if X is also a $N(0,1)$ random variable, then (X_i, Y_i) is a bivariate Rayleigh risk and ρ is the correlation coefficient. Let now $(\mathcal{U}_{m,i}^{(n)}, \mathcal{V}_{m,i}^{(n)})$, $1 \le i \le n$ be independent bivariate random vectors with joint cdf F_{mn} that coincides with the joint cdf of (U_m, V_m) for some fixed $\rho_n \in (-1,1)$, where for the definition of (U_m, V_m) we consider the general bivariate random vectors (X_i, Y_i) given by (2.1) with ρ substituted by $\rho_n \in (-1,1)$. Note that the cdf of V_m depends on n since we use now ρ_n. However, the cdf of U_m does not depend on n.

Under some restrictions on the marginal distributions $F_{mn,i}$, $i = 1,2$ of F_{mn} we have the following result.

Theorem 2.2. Suppose that for some positive constants $a_n > 0$, $b_n \in \mathbb{R}$, $n \geq 1$ we have

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} \Big| \left(F_{mn,i}\left(a_n x + b_n\right)\right)^n - \Lambda(x) \Big| = 0, \quad i = 1, 2.$$

If further the Hüsler–Reiss condition

$$\lim_{n\to\infty} \left(1 - \rho_n\right) \frac{b_n}{a_n} = \lambda^2 \in [0, \infty) \tag{2.2}$$

holds, then

$$\lim_{n\to\infty} \sup_{x,y\in\mathbb{R}} \Big| \left(F_{mn}\left(a_n x + b_n, a_n y + b_n\right)\right)^n - H_\lambda(x, y) \Big| = 0, \tag{2.3}$$

where H_λ is given in (1.2).

Remarks:

a) The convergence in (2.3) can be stated equivalently as the joint weak convergence of $\left(\max_{i\leq n} \mathcal{U}_{mi}^{(n)}, \max_{i\leq n} \mathcal{V}_{mi}^{(n)}\right)$ as $n \to \infty$.

b) In the case $m = 2$ and the base risk is $X = WI$ with W having $N(0, 1)$ distribution and I being a Bernoulli random variable independent of W, we can check that the assumptions of Theorem 2.2 are fulfilled.

3. PROOFS

Proof of Theorem 2.1: Using the well-known fact that

$${}_0F_1\left(;b,z\right) \sim \frac{\Gamma(b)}{2\sqrt{\pi}} z^{(1-2b)/4} \exp\left(2\sqrt{z}\right)$$

as $z \to \infty$, we have

$$\begin{aligned}
& \mathbb{P}\left(U_m > x + b, V_m > y + b\right) \\
= \; & \frac{1}{\Gamma(m/2)\left(2\widetilde{\varrho}\right)^{m/2}} \int_{x+b}^{\infty}\int_{y+b}^{\infty} (uv)^{m/2-1} \exp\Big(-\frac{u+v}{2\widetilde{\varrho}}\Big) \\
& \cdot {}_0F_1\left(;m/2;\rho^2 uv/\left(2\widetilde{\varrho}\right)^2\right) dvdu \\
\sim \; & \frac{|\rho|^{(1-m)/2}}{2\sqrt{\pi}\left(2\widetilde{\varrho}\right)^{1/2}} \int_{x+b}^{\infty}\int_{y+b}^{\infty} (uv)^{(m-3)/4} \exp\Big(\frac{|\rho|\sqrt{uv}}{\widetilde{\varrho}} - \frac{u+v}{2\widetilde{\varrho}}\Big) dvdu \\
=: \; & \frac{|\rho|^{(1-m)/2}}{2\sqrt{\pi}\left(2\widetilde{\varrho}\right)^{1/2}} I\left(b\right). \qquad (3.1)
\end{aligned}$$

Since $\sqrt{uv} \leq (u+v)/2$,

$$\begin{aligned} I(b) &\leq \int_{x+b}^{\infty}\int_{y+b}^{\infty}(uv)^{(m-3)/4}\exp\left(-\frac{u+v}{2(1+|\rho|)}\right)dvdu \\ &= \int_{x+b}^{\infty}u^{(m-3)/4}\exp\left(-\frac{u}{2(1+|\rho|)}\right)du\int_{y+b}^{\infty}v^{(m-3)/4}\exp\left(-\frac{v}{2(1+|\rho|)}\right)dv \\ &= \frac{1}{\left[2\left(1+|\rho|\right)\right]^{(m+1)/2}}\Gamma\Big(\frac{m+1}{4},\frac{2x+b}{2(1+|\rho|)}\Big)\Gamma\Big(\frac{m+1}{4},\frac{2y+b}{2(1+|\rho|)}\Big), \end{aligned}$$

where $\Gamma(s,z)=\int_z^{\infty}t^{s-1}\exp(-t)dt$ denotes the complementary incomplete gamma function. Since

$$\int_z^{\infty}t^{s-1}\exp(-At)dt=\frac{\Gamma(s,Az)}{A^s},\qquad A>0$$

and

$$\Gamma(s,z)=\exp(-z)z^{s-1}\left(1+O\left(z^{-1}\right)\right)$$

as $|z|\to\infty$, we conclude that for $|\rho|<1$

$$\begin{aligned} I(b) &= \left[\exp\left(-\frac{x+b/2}{1+|\rho|}\right)\Big(\frac{x+b/2}{1+|\rho|}\Big)^{\frac{m-3}{4}}\right] \\ &\quad\cdot\left[\exp\left(-\frac{y+b/2}{1+|\rho|}\right)\Big(\frac{y+b/2}{1+|\rho|}\Big)^{\frac{m-3}{4}}\right]\Big(1+O\left(b^{-1}\right)\Big) \\ &= b^{\frac{m-3}{2}\exp\Big(-\frac{b}{1+|\rho|}\Big)\left(1+O\left(b^{-1}\right)\right)} \qquad (3.2) \end{aligned}$$

as $b\to\infty$. The proof follows by combining (3.2) and (3.1). □

Proof of Theorem 2.2: Let (U_m,V_{mn}) be a bivariate random vector with the joint cdf F_{mn}. By the assumptions on the marginal distributions of F_{mn}, the proof follows if we show that

$$\begin{aligned} \lim_{n\to\infty} n\mathbb{P}\left(U_m>a_nx+b_n,V_{mn}>a_ny+b_n\right) &= \exp(-x)+\exp(-y)-\ln H_\lambda(x,y) \\ &=: g(x,y) \end{aligned}$$

holds for any $x,y\in\mathbb{R}$. Let $Z,Z_1,...,Z_n$ be independent $N(0,1)$ random variables and let

$$(X_i,Y_i)\stackrel{d}{=}\left(X_i,\rho_nX_i+\sqrt{1-\rho_n^2}Z_i\right),$$

assuming further that $X_i,Z_i,i\leq n$ are mutually independent and $X_i\stackrel{d}{=}X,i\geq 1$. Hence, we obtain

$$\begin{aligned} V_{mn}^2 &\stackrel{d}{=} \sum_{i=1}^{N}\left(Y_i-\sum_{i=1}^{N}Y_i/N\right)^2 \\ &\stackrel{d}{=} \sum_{i=1}^{N}\left(\rho_n\left(X_i-\overline{X}_N\right)-\sqrt{1-\rho_n^2}\left(Z_i-\overline{Z}_N\right)\right)^2 \\ &\stackrel{d}{=} \rho_n^2U_m-2\rho_n\sqrt{1-\rho_n^2}T_m+\left(1-\rho_n^2\right)V_m^*, \end{aligned}$$

where

$$T_m = \sum_{i=1}^{N} \left(X_i - \overline{X}_N\right) Z_i, \quad V_m^* = \sum_{i=1}^{N} \left(\overline{Z}_N - Z_i\right)^2.$$

By the independence of $X_i, Z_i, i \leq n, \overline{X}_N$ and the fact that $Z, Z_1, ..., Z_N$ are independent $N(0,1)$ random variables, we may further write

$$T_m = \sum_{i=1}^{N} \left(X_i - \overline{X}_N\right) Z_i \stackrel{d}{=} Z_1 \sqrt{\sum_{i=1}^{N} \left(X_i - \overline{X}_N\right)^2} \stackrel{d}{=} Z\sqrt{U_m}. \tag{3.3}$$

Hence, as in Hashorva *et al.* [11], we have for any $\varepsilon > 0$ and any $x, y \in \mathbb{R}$

$$\begin{aligned}
&\mathbb{P}\left(U_m > a_n x + b_n, V_{mn} > a_n y + b_n\right) \\
&= \mathbb{P}\left(U_m > a_n x + b_n,\, \rho_n^2 U_m - 2\rho_n\sqrt{1-\rho_n^2} T_m + \left(1-\rho_n^2\right) V_m^* > a_n y + b_n\right) \\
&\leq \mathbb{P}\Big(U_m > a_n x + b_n,\, \rho_n^2 U_m - 2\rho_n\sqrt{1-\rho_n^2} T_m + \left(1-\rho_n^2\right) V_m^* > a_n y + b_n, \\
&\quad \left(1-\rho_n^2\right) V_m^* \leq \varepsilon\Big) + \mathbb{P}\left(U_m > a_n x + b_n,\, \left(1-\rho_n^2\right) V_m^* > \varepsilon\right) \\
&\leq \mathbb{P}\left(U_m > a_n x + b_n,\, \rho_n^2 U_m - 2\rho_n\sqrt{1-\rho_n^2} Z\sqrt{U_m} > a_n y - \varepsilon + b_n\right) \\
&\quad + \mathbb{P}\left(U_m > a_n x + b_n,\, \left(1-\rho_n^2\right) V_m^* > \varepsilon\right).
\end{aligned}$$

By the assumptions, we have

$$\lim_{n\to\infty} n\mathbb{P}\left(U_m > a_n x + b_n\right) = \exp(-x), \quad \forall x \in \mathbb{R}.$$

Consequently, for some ε sufficiently small

$$\begin{aligned}
&\lim_{n\to\infty} n\mathbb{P}\left(U_m > a_n x + b_n, \left(1-\rho_n^2\right) V_m^* > \varepsilon\right) \\
&= \lim_{n\to\infty} n\mathbb{P}\left(U_m > a_n x + b_n\right) \mathbb{P}\left(\left(1-\rho_n^2\right) V_m^* > \varepsilon\right) = 0.
\end{aligned}$$

By the fact that V_m^* is non-negative, we have further

$$\begin{aligned}
&\mathbb{P}\left(U_m > a_n x + b_n, V_{mn} > a_n y + b_n\right) \\
&\geq \mathbb{P}\left(U_m > a_n x + b_n, \rho_n^2 U_m - 2\rho_n\sqrt{1-\rho_n^2} Z\sqrt{U_m} > a_n y + b_n\right).
\end{aligned}$$

We have with H the cdf of U_m (which does not depend on n)

$$n\left[1 - H\left(b_n\right)\right] \to 1, \quad \overline{H}_n(x) := \frac{1 - H\left(a_n x + b_n\right)}{1 - H\left(b_n\right)} \to \exp(-x), \quad \forall x \in \mathbb{R}$$

as $n \to \infty$. Furthermore, by condition (2.2) and the fact that $Z \stackrel{d}{=} -Z$

$$\begin{aligned}
l_n(x,y) &:= \mathbb{P}\left(\rho_n\left(a_n x + b_n\right) - 2\rho_n\sqrt{1-\rho_n^2}\sqrt{a_n x + b_n} Z > a_n y + b_n\right) \\
&\to \mathbb{P}\left(4\lambda Z > 2y - 2x + 2\lambda^2\right), \quad n \to \infty
\end{aligned}$$

holds locally uniformly for $x \in \mathbb{R}$. Using a conditional argument as in Hashorva *et al.* [11] and utilizing further (3.3), we obtain

$$\begin{aligned}
& g(x,y) \\
&= \lim_{n\to\infty} n \int_{a_n x+b_n}^{\infty} \mathbb{P}\left(\rho_n^2 U_m - 2\rho_n\sqrt{1-\rho_n^2}Z\sqrt{U_m} > a_n y + b_n \Big| U_m = s\right) dH(s) \\
&= \lim_{n\to\infty} \frac{1}{1-H\left(b_n\right)} \int_x^{\infty} \mathbb{P}\Big(\rho_n^2\left(a_n t + b_n\right) - 2\rho_n\sqrt{1-\rho_n^2}Z\sqrt{U_m} > a_n y + b_n \Big| \\
& \quad U_m = a_n t + b_n\Big)\, dH\left(a_n t + b_n\right) \\
&= -\lim_{n\to\infty} \int_x^{\infty} \mathbb{P}\left(\rho_n^2\left(a_n t + b_n\right) - 2\rho_n\sqrt{1-\rho_n^2}\sqrt{a_n t + b_n}Z > a_n y + b_n\right) dH_n(t) \\
&= -\lim_{n\to\infty} \int_x^{\infty} l_n(t,y) dH_n(t) \\
&= \int_x^{\infty} \mathbb{P}\left(Z > (y-t)/(2\lambda) + \lambda/2\right) \exp(-t) dt.
\end{aligned}$$

Utilizing the explicit expression of $g(x,y)$ derived in Hüsler and Reiss [13] establishes the proof. □

ACKNOWLEDGMENTS

E. Hashorva acknowledges support by the Swiss National Science Foundation Grants 200021-134785 and 200021-1401633/1. The authors would like to thank the Editor and the referee for careful reading and for their comments which improved the paper.

REFERENCES

[1] Bacro, J.-N.; Bel, L. and Lantuéjoul, C. (2010). Testing the independence of maxima: From bivariate vectors to spatial extreme fields: Asymptotic independence of extremes, *Extremes*, **13**, 155–175.

[2] Bouyé, E. (2002). Multivariate extremes at work for portfolio risk measurement, *Finance*, **23**, 125–144.

[3] Davis, R.A.; Klüppelberg, C. and Steinkohl, C. (2011). Max-stable processes for modelling extremes observed in space and time, Preprint.

[4] Ding, A.A. and Wang, W. (2007). Inference for bivariate survival data by copula models adjusted for the boundary effect, *Communications in Statistics — Theory and Methods*, **36**, 2927–2936.

[5] Engelke, S.; Kabluchko, Z. and Schlather, M. (2012a). Maxima of independent, non-identically distributed Gaussian vectors, Preprint.

[6] Engelke, S.; Malinowski, A.; Kabluchko, Z. and Schlather, M. (2012b). Estimation of Huesler–Reiss distributions and Brown–Resnick processes, Preprint.

[7] Engelke, S.; Malinowski, A.; Oesting, M. and Schlather, M. (2012c). Representations of max-stable processes based on single extreme events, Preprint.

[8] Frick, M. and Reiss, R.-D. (2010). Limiting distributions of maxima under triangular schemes, *Journal of Multivariate Analysis*, **101**, 2346–2357.

[9] Guegan, D. and Hassani, B. (2011). Multivariate VaRs for operational risk capital computation: A vine structure approach, CES Working Paper 2011.17.

[10] Hashorva, E. (2005). Elliptical triangular arrays in the max-domain of attraction of Hüsler–Reiss distribution, *Statistics and Probability Letters*, **72**, 125–135.

[11] Hashorva, E.; Kabluchko, Z. and Wübker, A. (2012). Extremes of independent chi-square random vectors, *Extremes*, **15**, 35–42.

[12] Haug, S.; Klüppelberg, C. and Peng, L. (2011). Statistical models and methods for dependence in insurance data, *Journal of the Korean Statistical Society*, **40**, 125–139.

[13] Hüsler, J. and Reiss, R.-D. (1989). Maxima of normal random vectors: Between independence and complete dependence. *Statistics and Probability Letters*, **7**, 283–286.

[14] Ji, J. and Chen, W. (2012). Asymptotic capacity analysis in point-to-multipoint cognitive radio networks. In: *Proceedings of the IEEE International Conference on Communications*, pp. 1–5.

[15] Joe, H. (1994). Multivariate extreme-value distributions with applications to environmental data, *Canadian Journal of Statistics*, **22**, 47–64.

[16] Kabluchko, Z. (2011). Extremes of independent Gaussian processes, *Extremes*, **14**, 285–310.

[17] Leira, B.J. and Myrhaug, D. (2011). Multivariate Gumbel distributions for reliability assessment. In: *Advances in Safety, Reliability and Risk Management*, pp. 1061–1068.

[18] Leira, B.J.; Myrhaug, D. and Holm, H. (2011). A bivariate extreme value distribution for reliability assessment of marine structures. In: *Proceedings of the 30th International Conference on Ocean, Offshore and Arctic Engineering*, pp. 703–709.

[19] Lescourret, L. and Robert, C.Y. (2006). Extreme dependence of multivariate catastrophic losses, *Scandinavian Actuarial Journal*, 203–225.

[20] Manjunath, B.G.; Frick, M. and Reiss, R.-D. (2012). Some notes on extremal discriminant analysis, *Journal Multivariate Analysis*, **103**, 107–115.

[21] Méndez, F.J. and Castanedo, S. (2007). A probability distribution for depth-limited extreme wave heights in a sea state, *Coastal Engineering*, **54**, 878-Ű882.

[22] Molchanov, I. and Schmutz, M. (2011). Multiasset derivatives and joint distributions of asset prices. In: *Musiela Festschrift*, Springer Verlag, Berlin.

[23] Nadarajah, S. (2010). Simple expressions for a bivariate chisquare distribution, *Statistics*, **44**, 189–201.

[24] Panajotović, A.S.; Stefanović, M.C. and Drača, D.L. (2009). Performance analysis of system with selection combining over correlated Rician fading channels in the presence of cochannel interference, *AEU — International Journal of Electronics and Communications*, **63**, 1061–1066.

[25] Panajotović, A.S.; Stefanović, M.C. and Drača, D.L. (2010). Cochannel interference effect on BEP performance of SSC receiver in correlated Rician fading, *Journal of the Franklin Institute — Engineering and Applied Mathematics*, **347**, 1242–1252.

[26] Smith, E.L. and Stephenson, A.G. (2009). An extended Gaussian max-stable process model for spatial extremes, *Journal of Statistical Planning and Inference*, **139**, 1266–1275.

[27] Stansell, P.; Wolfram, J. and Linfoot, B. (2002). Statistics of wave groups measured in the northern North Sea: Comparisons between time series and spectral predictions, *Applied Ocean Research*, **24**, 91–106.

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, $P_{\boldsymbol{\Omega}}$ with $\boldsymbol{\Omega} = (\boldsymbol{\theta}, \boldsymbol{\tau})$, 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 $\mathbf{X} \sim P_{\boldsymbol{\Omega}}$. 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 $\hat{\boldsymbol{\theta}}^{nr} = \bar{\mathbf{x}}$, $\mathbb{RESIIF}(\mathbf{x}; \hat{\boldsymbol{\theta}}^{nr}, \hat{\boldsymbol{\Omega}}^r)$ 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

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

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

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

b) Compute weights, w_i, according to

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

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 $\mathbb{SIIF}$ 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,

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

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, $w_i = w(x_i; P_n)$, $1 \leq i \leq n$, 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 $\hat{\mathbf{\Omega}}^r \xrightarrow{a.s.} \mathbf{\Omega}$ 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, $\sqrt{n}\left(\mu(P_{w_n,n}) - \mu(P_w)\right) \xrightarrow{d} N\left(0, \sigma^2(P_w)\right)$ (see Proposition 6.1b) in the Appendix for a related result concerning the Influence Function Bootstrap distribution). Thus, for large n, we have that

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

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 influ-

ence 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		$n = 20$		$n = 50$	
				Length		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

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

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,

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

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 $\mathbf{X}_i = (1, \mathbf{Z}_i^{\text{T}})^{\text{T}}$ with $\mathbf{Z}_i \sim N_{10}(0, \mathbf{I})$ and binary responses Y_i such that $Y_i|\mathbf{X}_i = \mathbf{x} \sim Bi(1, H(\mathbf{x}^{\text{T}}\boldsymbol{\beta}))$. In the first case, $\boldsymbol{\beta} = (0, 0, ..., 0)^{\text{T}}$, while in the second one, we choose $\boldsymbol{\beta} = (1, ..., 1)^{\text{T}}/3\sqrt{11}$.

We calculate the classical maximum likelihood (ML) and the robust weighted estimators introduced in Croux and Haesbroeck (2003) and denoted $\widehat{\boldsymbol{\beta}}_{\text{WBY}}$. 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 $\widehat{\boldsymbol{\beta}}_{\text{WBY}}$. 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 $\hat{\mu}_j = (\sum_{i=1}^{n} w_i x_{ij}) (\sum_{i=1}^{n} w_i)^{-1}$, $j = 1, 2$.

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 $\mu_w(P_n)$,

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

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

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

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

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

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

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

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 $\mathbf{X}_i \sim N_2(0, \mathbf{\Sigma})$ for $1 \leq i \leq n - [\varepsilon n]$ while $\mathbf{X}_i \sim \delta_{\mathbf{x}}$ when $n - [\varepsilon n] + 1 \leq i \leq n$. 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

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

  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

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

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

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 $Pw_1 = \mathbb{E}_P w_1(\mathbf{X}, \boldsymbol{\theta}) = 1$ and that $W_1(x) = \sup_{\boldsymbol{\theta} \in \Theta} w_1(\mathbf{x}, \boldsymbol{\theta})$ is such that $PW_1^2 < \infty$.

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. $|w_1(\mathbf{x}, \boldsymbol{\theta}_1) - w_1(\mathbf{x}, \boldsymbol{\theta}_2)| \leq \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\| F(\mathbf{x})$, 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

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

with $\rho_P^2(f) = P(f - Pf)^2$. Note that, $\rho_P^2(f_{\boldsymbol{\theta}_1,\mathbf{t}} - f_{\boldsymbol{\theta}_2,\mathbf{t}}) \leq \mathbb{E}_P(w_1(\mathbf{X},\boldsymbol{\theta}_1) - w_1(\mathbf{X},\boldsymbol{\theta}_2))^2 = B(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)$ 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

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

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 $\sqrt{n}\big(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\big)$.

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

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

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 $\sqrt{n}\,[M(\boldsymbol{\theta}_0) - M_n(\boldsymbol{\theta}_0)]$, so, using that $M(\boldsymbol{\theta}_0) \neq 0$, we have

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

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

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

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

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

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

Similarly, we get that

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

Thus,

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

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

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

which imply that

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

so,

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

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

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

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

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

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

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

Hence,

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

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

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

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

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

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.

REVSTAT – Statistical Journal

Background

Statistical Institute of Portugal (INE, I.P.), well aware of how vital a statistical culture is in understanding most phenomena in the present-day world, and of its responsibility in disseminating statistical knowledge, started the publication of the scientific statistical journal *Revista de Estatística*, in Portuguese, publishing three times a year papers containing original research results, and application studies, namely in the economic, social and demographic fields.

In 1998 it was decided to publish papers also in English. This step has been taken to achieve a larger diffusion, and to encourage foreign contributors to submit their work.

At the time, the Editorial Board was mainly composed by Portuguese university professors, being now composed by national and international university professors, and this has been the first step aimed at changing the character of *Revista de Estatística* from a national to an international scientific journal.

In 2001, the *Revista de Estatística* published three volumes special issue containing extended abstracts of the invited contributed papers presented at the 23rd European Meeting of Statisticians.

The name of the Journal has been changed to REVSTAT – STATISTICAL JOURNAL, published in English, with a prestigious international editorial board, hoping to become one more place where scientists may feel proud of publishing their research results.

— The editorial policy will focus on publishing research articles at the highest level in the domains of Probability and Statistics with emphasis on the originality and importance of the research.

— All research articles will be refereed by at least two persons, one from the Editorial Board and another, external.

— The only working language allowed will be English.

— Three volumes are scheduled for publication, one in April, one in June and the other in November.

— On average, four articles will be published per issue.

Aims and Scope

The aim of REVSTAT is to publish articles of high scientific content, in English, developing innovative statistical scientific methods and introducing original research, grounded in substantive problems.

REVSTAT covers all branches of Probability and Statistics. Surveys of important areas of research in the field are also welcome.

Abstract / indexed in

REVSTAT is expected to be abstracted / indexed at least in *Current Index to Statistics*, *Statistical Theory and Method Abstracts* and *Zentralblatt für Mathematik*.

Instructions to Authors, special-issue editors and publishers

Papers may be submitted in two different ways:

— By sending a paper copy to the Executive Editor and one copy to one of the two Editors or Associate Editors whose opinion the author(s) would like to be taken into account, together with a postscript or a PDF file of the paper to the e-mail: revstat@fc.ul.pt.

— By sending a paper copy to the Executive Editor, together with a postscript or a PDF file of the paper to the e-mail: revstat@fc.ul.pt.

Submission of a paper means that it contains original work that has not been nor is about to be published elsewhere in any form.

Submitted manuscripts (text, tables and figures) should be typed only in black, on one side, in double spacing, with a left margin of at least 3 cm and not have more than 30 pages.

The first page should include the name, affiliation and address of the author(s) and a short abstract with the maximum of 100 words, followed by the key words up to the limit of 6, and the AMS 2000 subject classification.

Authors are obliged to write the final version of accepted papers using LaTeX, in the REVSTAT style.

This style (REVSTAT.sty), and examples file (REVSTAT.tex), which may be download to *PC Windows System* (Zip format), *Mackintosh*, *Linux* and *Solaris Systems* (StuffIt format), and *Mackintosh System* (BinHex Format), are available in the REVSTAT link of the National Statistical Institute's Website: http://www.ine.pt/revstat/inicio.html

Additional information for the authors may be obtained in the above link.

Accepted papers

Authors of accepted papers are requested to provide the LaTeX files and also a postscript (PS) or an acrobat (PDF) file of the paper to the Secretary of REVSTAT: liliana.martins@ine.pt.

Such e-mail message should include the author(s)'s name, mentioning that it has been accepted by REVSTAT.

The authors should also mention if encapsulated postscript figure files were included, and submit electronics figures separately in .tiff, .gif, .eps or .ps format. Figures must be a minimum of 300 dpi.

Also send always the final paper version to:

Maria José Carrilho
Executive Editor, REVSTAT – STATISTICAL JOURNAL
Instituto Nacional de Estatística, I.P.
Av. António José de Almeida
1000-043 LISBOA
PORTUGAL

Copyright