

Estimating the Components of a Mixture of Extremal Distributions under Strong Dependence

Carolina Crisci , Gonzalo Perera , Lia Sampognaro 

Departamento Modelización Estadística de Datos e Inteligencia Artificial (MEDIA), CURE, Rocha, Universidad de la República, Rocha, Uruguay

Email: gperera@cure.edu.uy

How to cite this paper: Crisci, C., Perera, G. and Sampognaro, L. (2023) Estimating the Components of a Mixture of Extremal Distributions under Strong Dependence. *Advances in Pure Mathematics*, 13, 425-441. <https://doi.org/10.4236/apm.2023.137027>

Received: June 1, 2023

Accepted: July 9, 2023

Published: July 12, 2023

Copyright © 2023 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

In this paper, we provide a method based on quantiles to estimate the parameters of a finite mixture of Fréchet distributions, for a large sample of strongly dependent data. This is a situation that appears when dealing with environmental data and there was a real need of such method. We validate our approach by means of estimation and goodness-of-fit testing over simulated data, showing an accurate performance.

Keywords

Mixture of Extremal Distributions, Strongly Dependent Data

1. Introduction

In many applications of Statistics, the finite mixture model had been widely used to describe the distribution of data. A finite mixture model is a distribution that may be written as a finite, convex linear combination of distributions belonging to parametric classes. For instance, a mixture of k normal distributions, each one with its mean and variance, is a basic example, where the parameters involved are $k - 1$ non-negative weights (because their sum is one), and the $2k$ parameters corresponding to each mean and variance, making a total of $3k - 1$ parameters. In both theoretical developments and specific applications, the use of finite mixture models and the development of techniques of estimation of the unknown parameters have been deeply studied, with developments such as the expectation-maximization algorithm (EM) and its variants [1] [2] [3] [4] [5].

It should be noticed that the parametric classes of distributions involved in the mixture may be different. For instance, one may consider a mixture of a Normal

distribution with an exponential one.

In a general, abstract framework, one of the first questions to answer when considering a finite mixture model is if it is identifiable, that is, if there is a unique combination of all the involved parameters to express a given distribution. It is obvious that if the finite mixture model is not identifiable, estimations will be affected seriously by the fact that there are different sets of parameters leading to the same distribution.

More recently, both in theoretical and applied developments, the finite mixture of extremal distributions has increased its consideration [6] [7] [8] [9]. In Propositions 2.3.3 of [6], it is shown that finite mixtures of extremal distributions are identifiable, leading to the estimation of weights and parameters of the extremal components based on a random, *iid*, sample.

In a recent paper [10], another reason to pay attention to finite mixtures of extremal distributions is provided, because it is shown in its ***Theorem 1*** that, trying to mimic Fisher-Tippet-Gnedenko theory, when studying the asymptotic distribution of the maximum of a large sample, if data are non-stationary and strongly dependent, under very mild assumptions, the limit distribution is a finite mixture of extremal distributions, instead of an extremal one. This means that, when trying to fit a sample consisting on the list of the maximum values of blocks of a large number of continuous measures to a Generalized Extremal Value distribution (GEV), if the result of testing or diagnostic analysis is rejection, it may be related to a non-detected strong dependence and non-stationary structure on data. In addition, in many real data sets, in particular in environmental studies, non-stationarity and strong dependence should be expected. Consider the case mentioned in [11], when each data of our sample is the maximum wind speed registered by an online anemometer in a 10 minutes period, that may be well-fitted to a mixture of extremal distributions. If we dispose of several years of data, since one year has 52.560 periods of 10 minutes, and wind speed is affected by global phenomena that induce dependence trough years, one finds a significative correlation between data with lags of the order of 10^5 (or more), and non-stationarity is often evident.

Therefore, we need to develop a method for the estimation of the components of a finite mixture of extremal distributions, for large samples of strong dependent, and non-stationary data. Such a method will be a substantial improvement for the statistical analysis of large samples of complex environmental data.

This is the focus of the paper. More precisely, we will first recall the strong dependent and non-stationary models presented in [10] and propose an estimation method for the components of the mixture of $k = 2, 3$ extremal distributions. We will focus on the mixture of Fréchet distributions, for the sake of simplicity, and because they correspond to the most heavy-tailed data. Further, we will prove the consistency of our estimators and expose their performance using data simulated following models presented in [10], and checking the quality of the fitting of the estimated model to data, using the test for these types of models provided in [11].

Therefore, the method introduced here is a new and effective tool for the statistical analysis of strong-dependent data, as is required in several environmental applications.

2. Preliminary Results

At first, we will now recall the main result of [10] in a compressed manner. We assume that classical Fisher-Tippet-Gnedenko theory, in particular concepts like maximal domain of attraction (MDA, in what follows), are well-known for the reader. For a reference in the topic as well as some examples of its wide domain of application to real data, see [12] [13] [14] [15].

Our data will be X_1, \dots, X_n with $X_i = f(\xi_i, Y_i)$, where $Y_i \in \{1, \dots, k\} \forall i$ and we will assume the following hypotheses:

(H1) $\frac{1}{n} \sum_{j=1}^n 1_{\{Y_i=j\}} \xrightarrow[n]{a.s.} b_j$ where b_j is a positive random variable. More precisely, if $I(t) = \sigma\{Y_i : i \geq t\}$, and $I(\infty) = \bigcap_{t>1} I(t)$, then, since for any j , b_j is $I(\infty)$ -measurable, if $I(\infty)$ is trivial (what means weak dependence on the process Y), b_1, \dots, b_k are deterministic, but, if $I(\infty)$ is not trivial (what means strong dependence on the process Y), for some j , b_j may be non-deterministic.

(H2) For any j , b_j assume a finite numbers of values.

(H3) The three following conditions are fulfilled.

1) $(\xi_i)_{i \in \mathbb{N}}$ is iid

2) Y_1, Y_n, \dots satisfy (H1) and (H2)

3) The processes $(\xi_i)_{i \in \mathbb{N}}$ and $(Y_i)_{i \in \mathbb{N}}$ are independent.

(H4) For any $j = 1, \dots, k$ the process $X_i^j = f(\xi_i, j)$ belongs to the MDA of the GEV G^j , where G^1 is the most heavy-tailed of them, and corresponds to a Fréchet distribution of order α (we will denote Φ_α the standard Fréchet distribution of order α).

We are now in conditions to present the main result of [10].

Theorem 1 of [10].

Under (H3) and (H4) there exists a random variable Z such that

$$\frac{\max(X_1, \dots, X_n)}{n^{1/\alpha}} \xrightarrow[n \rightarrow \infty]{w} Z$$

In addition:

1) If $I(\infty)$ is trivial, then the distribution of Z is $F_z(x) = \Phi_\alpha\left(\frac{x}{b_1^{1/\alpha}}\right)$.

2) If $I(\infty)$ is not trivial and b_1 assumes the values v_1, \dots, v_r with probabilities p_1, \dots, p_r , then the distribution of Z is

$$F_z(x) = \sum_{i=1}^r p_i \Phi_\alpha\left(\frac{x}{v_i^{1/\alpha}}\right) \quad (\text{Mixture of Fréchet distributions}).$$

Remark 1

Part *b* of **Theorem 1** means that finite mixtures of Fréchet distributions of the same order, but with different scale parameters, appear when one tries to approximate the distribution of the maximum of a large sample of strongly-dependent

data. As mentioned in the introduction, this is a situation that appears when dealing in practice with environmental data. Therefore, from now on, we will try to provide statistical procedures to estimate the order α , the weights p_1, \dots, p_r and the scale parameters v_1, \dots, v_r assuming that such a mixture applies to our data and validate (or not) its fitting by means of the test provided in [11]. Finally, for the sake of simplicity, and taking into account that estimations will be tested, in the case of the order α , we will just use an exploratory estimator. Even if the results exposed in this paper are auspicious, it is clear that for a deeper approach, the estimation of the order α must be refined.

We will provide now some classical statistical procedures enabling to prove consistency of estimators.

First, remember that for Z_1, \dots, Z_n, \dots independent, centered and bounded we have that

$$P\left(\left|\frac{1}{n}\sum_{i=1}^n Z_i\right| > \varepsilon\right) \leq \frac{C}{n^2 \varepsilon^2}$$

Let us also remember that this implies complete convergence of $\frac{1}{n}\sum_{i=1}^n Z_i$ to zero for n tending to infinity, i.e.,

$$\sum_{n=1}^{\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^n Z_i\right| > \varepsilon\right) < \infty$$

what in turn implies almost sure convergence, i.e.,

$$\frac{1}{n}\sum_{i=1}^n Z_i \xrightarrow[n \rightarrow \infty]{a.s.} 0$$

Then we have the following consistency result.

Theorem 1: If ξ, Y satisfy (H3) of [10], and φ is a bounded function, then

$$\frac{1}{n}\sum_{i=1}^n \varphi(\xi_i, Y_i) \xrightarrow[n]{a.s.} \sum_{j=1}^k m(j)b_j,$$

where

$$m(j) = E\{\varphi(\xi_0, j)\}, j = 1, \dots, k$$

Proof

First, consider

$$Z_i^* = \varphi(\xi_i, Y_i) - \sum_{j=1}^k m(j)1_{\{Y_i=j\}}$$

It is clear that $E(Z_i^*) = 0 \quad \forall i$, and that $Z_1^*, \dots, Z_n^*, \dots$ are bounded. Then, calling $Y = (Y_i)_{i \in \mathbb{N}}$, and $y = (y_i)_{i \in \mathbb{N}}$ a fixed element of $S = \{1, \dots, k\}^\infty$, we have, for any $\varepsilon > 0$,

$$P\left(\left|\frac{1}{n}\sum_{i=1}^n Z_i^*\right| > \varepsilon\right) = \int_S P\left(\left|\frac{1}{n}\sum_{i=1}^n Z_i^*\right| > \varepsilon / Y = y\right) dP^Y(y)$$

But

$$P\left(\left|\frac{1}{n}\sum_{i=1}^n Z_i^*\right| > \varepsilon / Y = y\right) = P\left(\left|\frac{1}{n}\sum_{i=1}^n \hat{Z}_i\right| > \varepsilon / Y = y\right)$$

where $\hat{Z} = \varphi(\xi_i, y_i) - m(y_i)$, that are clearly independent, centered and bounded variables, and therefore

$$\begin{aligned} \sum_{n=1}^{\infty} P\left(\left|\frac{1}{n} \sum_{i=1}^n \hat{Z}_i\right| > \varepsilon\right) &\leq \sum_{n=1}^{\infty} \int_S P\left(\left|\frac{1}{n} \sum_{i=1}^n \hat{Z}_i\right| > \varepsilon / Y = y\right) dP^Y(y) \\ &\leq \sum_{n=1}^{\infty} \frac{Ck^2}{\varepsilon^2 n^2} < \infty, \end{aligned}$$

what implies that $\frac{1}{n} \sum_{i=1}^n \hat{Z}_i \xrightarrow[n \rightarrow \infty]{a.s.} 0$, what in turn implies that

$$\frac{1}{n} \sum_{i=1}^n Z_i^* \xrightarrow[n \rightarrow \infty]{a.s.} 0.$$

Therefore

$$\frac{1}{n} \sum_{i=1}^n \varphi(\xi_i, Y_0) - \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^k m(j) 1_{\{Y_i=j\}} \xrightarrow[n \rightarrow \infty]{a.s.} 0$$

But

$$\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^k m(j) 1_{\{Y_i=j\}} = \sum_{j=1}^k m(j) \frac{1}{n} \sum_{i=1}^n 1_{\{Y_i=j\}} \xrightarrow[n \rightarrow \infty]{a.s.} \sum_{j=1}^k m(j) b_j$$

and hence, we conclude that

$$\frac{1}{n} \sum_{i=1}^n \varphi(\xi_i, Y_i) \xrightarrow[n \rightarrow \infty]{a.s.} \sum_{j=1}^k m(j) b_j$$

Remark 2

As a clear consequence of **Theorem 1** the empirical distribution of a large sample satisfying **(H3)**, and where data are equally distributed, converges to the theoretical distribution at any given point. That is, the empirical distribution is a consistent estimator of the theoretical one at any given point. Calling F to the theoretical distribution and F_n to the empirical one, when F is continuous, since F_n is monotonous, by well-known elementary arguments, consistency is uniform, that is

$$\sup_{t \in \mathbb{R}} |F_n(t) - F(t)| \xrightarrow[n \rightarrow \infty]{a.s.} 0$$

This result is consistent with (slightly more general, in fact) **Theorem 1** of [11].

3. Mixture of Two Components - Simulation of Data

We will consider now the case of a mixture of $k = 2$ extremal distributions. The procedure to simulate our data follows very closely the one proposed in [10], but we will explain it here, for a better reading and comprehension.

Example I:

Let U be a random variable such that $P(U = 1) = p$, $P(U = 2) = 1 - p$. Let $\sigma_1, \dots, \sigma_n, \dots$ an iid sequence of random variables on 1,2 independent of U such that $P(\sigma_i(1) = 1) = \delta$, $P(\sigma_i(1) = 2) = 1 - \delta$, $P(\sigma_i(2) = 1) = \eta$, $P(\sigma_i(2) = 2) = 1 - \eta$, with $\sigma_i(1)$, $\sigma_i(2)$ independent among them for any i , $0 < \delta < 1$, $0 < \eta < 1$, $\delta \neq \eta$. Set $Y_i = \sigma_i(U)$.

Thus, $\frac{1}{n} \sum_{i=1}^n 1_{\{Y_i=1\}} / U = 1$ has the same distribution as

$\frac{1}{n} \sum_{i=1}^n 1_{\{\sigma_i(1)=1\}} \xrightarrow[n \rightarrow \infty]{a.s.} P(\sigma_i(1)=1) = \delta$ (by the Strong Law of Large Numbers).

On the other hand $\frac{1}{n} \sum_{i=1}^n 1_{\{Y_i(1)=1\}} / U = 2$ has the same distribution as

$\frac{1}{n} \sum_{i=1}^n 1_{\{\sigma_i(2)=1\}} \xrightarrow[n \rightarrow \infty]{a.s.} P(\sigma_i(2)=1) = \eta$. Therefore, we have that

$$b_1 = \begin{cases} \delta & \text{if } U = 1 \\ \eta & \text{if } U = 2 \end{cases}$$

Hence, b_1 is not-deterministic and $I(\infty)$ is not trivial. Similar treatment applies to b_2 .

Example II:

Now, if $Y_i = \sigma_i(U)$, we have that Y_1, \dots, Y_n, \dots fulfills **(H1)**, **(H2)** of section 0.2, with b_1, b_2 random variables such that

$$b_2 = \begin{cases} 1 - \delta & \text{if } U = 1 \\ 1 - \eta & \text{if } U = 2 \end{cases}$$

Thus, if we assume $0 < \alpha_1 < \alpha_2$ and consider two independent sequences $V_1^{(1)}, \dots, V_n^{(1)}, \dots, iid \sim F^{(1)}$, $V_1^{(2)}, \dots, V_n^{(2)}, \dots, iid \sim F^{(2)}$, with $F^{(i)} \in MDA(\Phi_{\alpha_i})$, $i = 1, 2$ and we set:

- 1) If $\sigma_i(U) = 1, X_i = V_i^{(1)}$
- 2) If $\sigma_i(U) = 2, X_i = V_i^{(2)}$

Then, X_1, \dots, X_n, \dots fulfills **(H3)**, **(H4)** of section 0.2 and therefore, **Theorem**

1 of [10] applies and, $\frac{\max(X_1, \dots, X_n)}{n^{1/\alpha_1}} \xrightarrow[n \rightarrow \infty]{w} MF$, with

$MF(x) = p\Phi_{\alpha_1}\left(\frac{x}{\delta^{1/\alpha_1}}\right) + (1-p)\Phi_{\alpha_1}\left(\frac{x}{\eta^{1/\alpha_1}}\right) \quad \forall x > 0$, a mixture of Fréchet distributions of order α_1 . We use this algorithm to simulate our data for evaluation of estimation methods in the case of $k = 2$.

4. A Method for Estimation of Parameters

As explained in **Remark 1** we will just provide a very rough estimation procedure for the order α .

4.1. An Exploratory Estimation for α

In our model:

$$MF(x) = p\Phi_{\alpha}\left(\frac{x}{v_1}\right) + (1-p)\Phi_{\alpha}\left(\frac{x}{v_2}\right)$$

where $0 < p < 1$, $\alpha > 0$, $v_1 > 0$, $v_2 > 0$, we may assume, without loss of generality that $v_2 > v_1$. Since $\Phi_{\alpha}(x) = e^{-x^{\alpha}}$, we have then

$$MF(x) = pe^{-\frac{v_1^{\alpha}}{x^{\alpha}}} + (1-p)e^{-\frac{v_2^{\alpha}}{x^{\alpha}}}$$

For x large enough,

$$\frac{-v_1^{\alpha}}{x^{\alpha}}$$

and

$$\frac{-v_2^\alpha}{x^\alpha}$$

are close to zero, and since $e^u \approx 1 + u$ for u close to zero, we then have that, for x large enough

$$MF(x) \approx p \left(1 - \frac{v_1^\alpha}{x^\alpha}\right) + (1-p) \left(1 - \frac{v_2^\alpha}{x^\alpha}\right) = 1 - \frac{pv_1^\alpha + (1-p)v_2^\alpha}{x^\alpha}$$

and, therefore,

$$\frac{\log(1 - MF(x))}{\log(x)} = \frac{\log(pv_1^\alpha + (1-p)v_2^\alpha)}{\log(x)} - \alpha$$

which tends to $-\alpha$ as x goes to infinity.

Then, since by **Theorem 1** the empirical distribution F_n is an uniformly consistent estimator of MF , α will be estimated by the values of: $\frac{-\log(1 - F_n(x))}{\log(x)}$

for x large enough.

As we will see later on, we simulate a mixture of two Fréchet distributions of order 1, and **Figure 1** shows that the estimation procedure is consistent.

4.2. Estimation of p, v_1, v_2

From now on, we shall assume α known, and we will focus on the estimation of p ($0 < p < 1$), and v_1, v_2 , ($0 < v_1 < v_2$).

Let us consider three particular values: $1, 2^{1/\alpha}, 4^{1/\alpha}$. It is clear that $1 < 2^{1/\alpha} < 4^{1/\alpha}$, and that $(2^{1/\alpha})^2 = 4^{1/\alpha}$. We have:

$$\begin{aligned} MF(1) &= pe^{-v_1^\alpha} + (1-p)e^{-v_2^\alpha} \\ MF(2^{1/\alpha}) &= pe^{-\frac{v_1^\alpha}{2}} + (1-p)e^{-\frac{v_2^\alpha}{2}} \\ MF(4^{1/\alpha}) &= pe^{-\frac{v_1^\alpha}{4}} + (1-p)e^{-\frac{v_2^\alpha}{4}} \end{aligned} \tag{1}$$

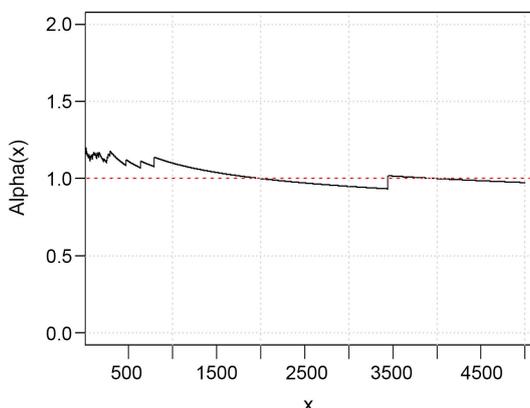


Figure 1. Estimation of α .

Calling:

$$u = e^{-\frac{v_1^\alpha}{4}}, v = e^{-\frac{v_2^\alpha}{4}} \tag{2}$$

and since $0 < b_1 < b_2$, we have that $0 < v < u < 1$ and we get:

$$\begin{aligned} (-4\log(u))^{1/\alpha} &= v_1 \\ (-4\log(v))^{1/\alpha} &= v_2 \end{aligned} \tag{3}$$

and, thus, the estimation of u, v leads to the estimation of v_1, v_2 . Further observe that:

$$e^{-\frac{v_1^\alpha}{2}} = u^2, e^{-\frac{v_2^\alpha}{2}} = v^2, e^{-v_1^\alpha} = u^4, e^{-v_2^\alpha} = v^4,$$

and therefore, (1) may be rewritten as:

$$\begin{aligned} MF(1) &= pu^4 + (1-p)v^4 \\ MF(2^{1/\alpha}) &= pu^2 + (1-p)v^2 \\ MF(4^{1/\alpha}) &= pu + (1-p)v \end{aligned} \tag{4}$$

As usual in Statistics, and taking into account that **Theorem 1** shows the uniform consistency of F_n as an estimator of MF for our model, if we replace in (4) MF by F_n and we manage to solve the equations in p, u, v , this will lead to a consistent estimation of p, u, v . For the sake of simplicity, we will denote p, u, v , their estimated values (instead of p_n, u_n, v_n). Therefore, we will solve (4):

$$\begin{aligned} F_n(1) &= pu^4 + (1-p)v^4 \\ F_n(2^{1/\alpha}) &= pu^2 + (1-p)v^2 \\ F_n(4^{1/\alpha}) &= pu + (1-p)v \end{aligned} \tag{5}$$

Taking the first two Equation of (5), it is clear that they can be rewritten in matrix terms as:

$$\begin{pmatrix} F_n(1) \\ F_n(2^{1/\alpha}) \end{pmatrix} = \begin{pmatrix} u^4 & v^4 \\ u^2 & v^2 \end{pmatrix} \begin{pmatrix} p \\ 1-p \end{pmatrix} \tag{6}$$

Calling $\mathbb{A} = \begin{pmatrix} u^4 & v^4 \\ u^2 & v^2 \end{pmatrix}$ we have that $\det(\mathbb{A}) = u^4v^2 - u^2v^4 = u^2v^2(u^2 - v^2) > 0$, since $0 < v < u$, what means that \mathbb{A} is invertible with inverse matrix

$$\mathbb{A}^{-1} = \begin{pmatrix} v^2 & -v^4 \\ -u^2 & u^4 \end{pmatrix} \frac{1}{u^2v^2(u^2 - v^2)}$$

and therefore, we have

$$\begin{pmatrix} p \\ 1-p \end{pmatrix} = \frac{\begin{pmatrix} v^2 & -v^4 \\ -u^2 & u^4 \end{pmatrix}}{u^2v^2(u^2 - v^2)} \begin{pmatrix} F_n(1) \\ F_n(2^{1/\alpha}) \end{pmatrix} \tag{7}$$

Remark 3

It should be noticed, as be used later on, that, more in general, if $k \geq 3$, $0 < u_1 < u_2 < \dots < u_k$ and we consider the $k \times k$ matrix:

$$\mathbb{A} = \begin{pmatrix} u_1^{2^k} & u_2^{2^k} & \dots & u_k^{2^k} \\ \vdots & \vdots & & \vdots \\ u_1^4 & u_2^4 & \dots & u_k^4 \\ u_1^2 & u_2^2 & \dots & u_k^2 \end{pmatrix}$$

then \mathbb{A} is invertible.

Thus

$$\begin{pmatrix} p \\ 1-p \end{pmatrix} = \mathbb{A}^{-1} \begin{pmatrix} F_n(1) \\ F_n(2^{1/\alpha}) \end{pmatrix},$$

calling $\mathbb{A}_1^{-1}, \mathbb{A}_2^{-1}$ to the first and second rows of \mathbb{A}^{-1} , we get the non-linear system:

$$\begin{cases} p = \mathbb{A}_1^{-1} \begin{pmatrix} F_n(1) \\ F_n(2^{1/\alpha}) \end{pmatrix} \\ 1-p = \mathbb{A}_2^{-1} \begin{pmatrix} F_n(1) \\ F_n(2^{1/\alpha}) \end{pmatrix} \end{cases} \tag{8}$$

with $0 < p < 1$, $0 < v < u$ as variables. Adding to this system the only equation of (4) that we have not used yet, $F_n(4^{1/\alpha}) = pu + (1-p)v$, that can be rewritten

$$v = \frac{F_n(4^{1/\alpha}) - pu}{1-p} \tag{9}$$

and imposing the restriction

$$F_n(4^{1/\alpha}) < u \tag{10}$$

we replace (9), and (10) in (7), obtaining

$$\begin{aligned} p &= C_1 \begin{pmatrix} F_n(1) \\ F_n(2) \end{pmatrix} \\ 1-p &= C_2 \begin{pmatrix} F_n(1) \\ F_n(2) \end{pmatrix} \end{aligned} \tag{11}$$

with $C_1 = \mathbb{A}_1^{-1}$, $C_2 = \mathbb{A}_2^{-1}$ depending only on p, u , because v is replaced by (8), and p, u , restricted to the constraints

$$0 < p < 1, F_n(4^{1/\alpha}) < u \tag{12}$$

we arrive to the non-linear equation

$$0 = \left(p - C_1 \begin{pmatrix} F_n(1) \\ F_n(2^{1/\alpha}) \end{pmatrix} \right)^2 + \left(1-p - C_2 \begin{pmatrix} F_n(1) \\ F_n(2^{1/\alpha}) \end{pmatrix} \right)^2 \tag{13}$$

under the constraints (12). (13) is solved by the Newton-Raphson method or any

other non-linear equation-solving method. Then, using (13), from the estimators (p, u, v) we get the estimators (p, v_1, v_2) .

Remark 4

As mentioned before, the estimation procedure leads to consistent estimators of the parameters. Then, one may ask by their asymptotic distribution to perform confidence intervals, etc. Even if this is not included in the main goals of this work (because, as pointed out in the introduction, we will validate estimations by suitable testing), we shall explain briefly how this asymptotic distribution is obtained. The solutions of the non-linear Equation (13), using the Implicit Function Theorem may be expressed in the following way

$$(p, u, v) = h\left(F_n(1), F_n(2^{1/\alpha}), F_n(4^{1/\alpha})\right) \quad (14)$$

with h a differentiable function.

Since in the preliminaries of **Theorem 2** of [11] the asymptotic distribution of the empirical process is derived, a standard application of the Delta Method ([16]), leads to the asymptotic distribution of the estimators (p, u, v) . Of course, the same applies to (p, v_1, v_2) . Its estimation will be treated later on, but this remark also applies in that context.

5. Testing the Estimated Model

As a concrete example of the method as well as a validation procedure, we will now simulate a large sample with strong-dependence, where the common distribution of all the data is a mixture of two Fréchet distributions. We will test if data fits to a single Fréchet distribution and rejection is expected. Further, we will use our method to estimate the parameters of a mixture of two Fréchet distributions, and in this case it is expected that the goodness of fit test does not reject the estimated model.

We will then choose as the true model a mixture of Fréchet distributions with $p = 0.3$, $v_1 = 0.3$, $v_2 = 0.7$, that is

$$MF(x) = 0.3\Phi_1\left(\frac{x}{0.3}\right) + 0.7\Phi_1\left(\frac{x}{0.7}\right)$$

We computed 4000 maximums, each one coming from samples of size 500 of the simulation procedure described in section 0.3, with parameters $p = 0.3$, $\delta = 0.3$ and $\eta = 0.7$. As mentioned before in section 0.3, by **Theorem 1** of [10], these maximum should follow a distribution very close to our choice of MF .

Remark 5

It should be noticed that indeed, we are not simulating data following the distribution MF , but following a distribution that is very close to MF , according to Theorem 1 of [10]. There are two reasons for that choice. At first, obviously, this choice makes harder the work for the estimation procedure because the real model is not exactly a mixture of two extremal distributions. At second, it is of particular interest this kind of data, because as pointed out in the Introduction and [10], they appear in many applications.

With our simulated sample of 4000 maximum values, we first proposed for fitting (*i.e.*, as H_0 in our test) a simple Fréchet model with $\alpha = 1$ (F1). In this example and all the further ones, we have used the adaptation of Kolmogorov-Smirnov test (KS) for this type of models provided by [11]. In this context, for F1, the KS statistic was 0.1928443, what means that p -value $\ll 0.001$, implying a clear rejection.

Figure 2 shows the difference between the empirical distribution of our sample, and the theoretical distribution of the proposed model (F1). Clearly the distribution of the proposed model (red curve) is below the empirical distribution (black curve), reflecting the much more heavy-tailed nature of the proposed model with respect to the true model.

Therefore we turn our attention to the estimation of a mixture of two components. The exploratory estimation of α corresponds to the **Figure 1** leading to $\alpha = 1$. Then, following the procedure of the previous section, we get the following results: $p = 0.25$, $v_1 = 0.35$, $v_2 = 0.7200471$. We perform the KS -test proposing as H_0 the mixture of two Fréchet of order 1 with the estimated parameters, leading to a KS statistic equal to 0.01380997, which implies p -value > 0.20 (**Figure 3**).

In conclusion, the simulated model fits the estimated two components mixture and does not fit an extremal distribution.

6. Mixture of Three Components - Simulation of Data

We will now turn our attention to the case of a mixture of $k = 3$ extremal distributions.

Again, the basis of the models that we will present here is provided in [10], but we have to explain them for the sake of clarity.

Example III:

Let U be a random variable such that $P(U = 1) = p$, $P(U = 2) = q$, $P(U = 3) = 1 - p - q$, $p > 0$, $q > 0$, $p + q < 1$. Let $\sigma_1, \dots, \sigma_n, \dots$ an *iid* sequence of random variables on $1, 2, 3$.

$P(\sigma_i(1) = 1) = \delta$, $P(\sigma_i(1) = 2) = \lambda$, $P(\sigma_i(1) = 3) = 1 - \delta - \lambda$, with $\delta > 0$, $\lambda > 0$, $\delta + \lambda < 1$.

$P(\sigma_i(2) = 1) = \eta$, $P(\sigma_i(2) = 2) = \rho$, $P(\sigma_i(2) = 3) = 1 - \eta - \rho$, with $\eta > 0$, $\rho > 0$, $\eta + \rho < 1$.

$P(\sigma_i(3) = 1) = \tau$, $P(\sigma_i(3) = 2) = \nu$, $P(\sigma_i(3) = 3) = 1 - \tau - \nu$, with $\tau > 0$, $\nu > 0$, $\tau + \nu < 1$.

Set $Y_i = \sigma_i(U)$. Thus,

$$\frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{Y_i=1\}} / U = 1 \xrightarrow[n \rightarrow \infty]{a.s.} \delta$$

$$\frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{Y_i=2\}} / U = 2 \xrightarrow[n \rightarrow \infty]{a.s.} \eta$$

$$\frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{Y_i=3\}} / U = 3 \xrightarrow[n \rightarrow \infty]{a.s.} \tau$$

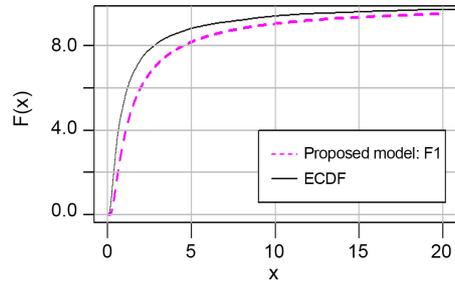


Figure 2. The difference between the empirical distribution (*ECDF*), and the theoretical distribution of the proposed model *F1*.

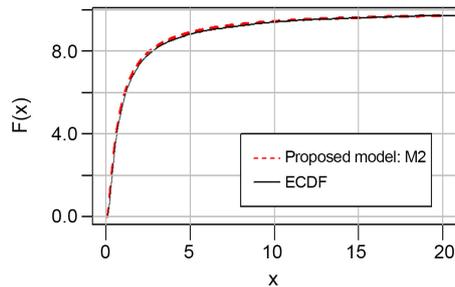


Figure 3. The difference between the empirical distribution (*ECDF*), and the theoretical distribution of the proposed model *M2*.

Therefore if we assume that $\delta \neq \eta \neq \tau$, $\delta \neq \tau$, $\lambda \neq \rho \neq \nu$, $\lambda \neq \nu$, $\delta + \lambda \neq \eta + \rho \neq \tau + \nu$, $\delta + \lambda \neq \tau + \nu$, we have that

$$b_1 = \begin{cases} \delta & \text{if } U = 1 \\ \eta & \text{if } U = 2 \\ \tau & \text{if } U = 3 \end{cases}$$

$$b_2 = \begin{cases} \lambda & \text{if } U = 1 \\ \rho & \text{if } U = 2 \\ \nu & \text{if } U = 3 \end{cases}$$

$$b_3 = \begin{cases} 1 - \delta - \lambda & \text{if } U = 1 \\ 1 - \eta - \rho & \text{if } U = 2 \\ 1 - \tau - \nu & \text{if } U = 3 \end{cases}$$

Example IV:

Now, we define $Y_i = \sigma_i(U)$, for any $i \in \mathbb{N}$. We then have that Y_1, \dots, Y_n, \dots fulfills (H1), (H2) of section 0.2, with b_1, b_2, b_3 random variables as in **Example III**.

Thus, if we assume $0 < \alpha_1 < \alpha_2 < \alpha_3$, and consider three independent sequences $V_1^{(1)}, \dots, V_n^{(1)}, \dots, iid \sim F^{(1)}$, $V_1^{(2)}, \dots, V_n^{(2)}, \dots, iid \sim F^{(2)}$, $V_1^{(3)}, \dots, V_n^{(3)}, \dots, iid \sim F^{(3)}$, $F^{(i)} \in MDA(\Phi_{\alpha_i})$, and for any i we set:

- 1) If $\sigma_i(U) = 1, X_i = V_i^{(1)}$
- 2) If $\sigma_i(U) = 2, X_i = V_i^{(2)}$

3) If $\sigma_i(U) = 3, X_i = V_i^{(3)}$

Then, X_1, \dots, X_n, \dots fulfills **(H3)**, **(H4)** of section 0.2 and therefore, **Theorem**

1 of [10] applies and, $\frac{\max(X_1, \dots, X_n)}{n^{1/\alpha_1}} \xrightarrow[n \rightarrow \infty]{w} MF$, with

$$MF(x) = p\Phi_{\alpha_1}\left(\frac{x}{\delta^{1/\alpha_1}}\right) + q\Phi_{\alpha_1}\left(\frac{x}{\eta^{1/\alpha_1}}\right) + (1-p-q)\Phi_{\alpha_1}\left(\frac{x}{\tau^{1/\alpha_1}}\right)$$

7. A Method for Estimation of Parameters, Case $k = 3$

As pointed out in **Remark 1**, for the estimation of α we just use an exploratory method. Therefore, we will concentrate our attention in the estimation of weights and scale parameters.

Estimation of p, q, v_1, v_2, v_3

Let us consider now

$$\begin{aligned} F_n(16^{1/\alpha}) &= pu + qv + (1-p-q)w \\ F_n(8^{1/\alpha}) &= pu^2 + qv^2 + (1-p-q)w^2 \\ F_n(4^{1/\alpha}) &= pu^4 + qv^4 + (1-p-q)w^4 \\ F_n(2^{1/\alpha}) &= pu^8 + qv^8 + (1-p-q)w^8 \\ F_n(1^{1/\alpha}) &= pu^{16} + qv^{16} + (1-p-q)w^{16} \end{aligned} \tag{15}$$

with

$$\begin{aligned} u &= \exp\left(\frac{-v_1^\alpha}{16}\right) \\ v &= \exp\left(\frac{-v_1^\alpha}{16}\right) \\ w &= \exp\left(\frac{-v_1^\alpha}{16}\right) \end{aligned} \tag{16}$$

Following the ideas of section 0.4.2 we write down

$$\begin{pmatrix} F_n(1^{1/\alpha}) \\ F_n(2^{1/\alpha}) \\ F_n(4^{1/\alpha}) \end{pmatrix} = \begin{pmatrix} u^{16} & v^{16} & w^{16} \\ u^8 & v^8 & w^8 \\ u^4 & v^4 & w^4 \end{pmatrix} \begin{pmatrix} p \\ q \\ 1-p-q \end{pmatrix} \tag{17}$$

Setting

$$A = \begin{pmatrix} u^{16} & v^{16} & w^{16} \\ u^8 & v^8 & w^8 \\ u^4 & v^4 & w^4 \end{pmatrix}$$

and using **Remark 2** we get

$$\begin{pmatrix} p \\ q \\ 1-p-q \end{pmatrix} = A^{-1} \begin{pmatrix} F_n(1^{1/\alpha}) \\ F_n(2^{1/\alpha}) \\ F_n(4^{1/\alpha}) \end{pmatrix}$$

From the equations

$$F_n(16^{1/\alpha}) = pu + qv + (1 - p - q)w$$

$$F_n(8^{1/\alpha}) = pu^2 + qv^2 + (1 - p - q)w^2$$

we may express u, v as a function of p, q, w . Calling C_1, C_2, C_3 to the first, second, and third (respectively) row of \mathbb{A}^{-1} with u, v replaced as a function of p, q, w , we have then the non-linear equation on p, q, w :

$$0 = \left(p - C_1 \begin{pmatrix} F_n(1) \\ F_n(2^{1/\alpha}) \\ F_n(4^{1/\alpha}) \end{pmatrix} \right)^2 + \left(q - C_2 \begin{pmatrix} F_n(1) \\ F_n(2^{1/\alpha}) \\ F_n(4^{1/\alpha}) \end{pmatrix} \right)^2 + \left(1 - p - q - C_3 \begin{pmatrix} F_n(1) \\ F_n(2^{1/\alpha}) \\ F_n(4^{1/\alpha}) \end{pmatrix} \right)^2 \quad (18)$$

Solving this equation we get the estimates of p, q, w and therefore of u, v . As in the case of two components we will denote this estimations omitting its dependence of the sample size n . From the estimations of p, q, u, v, w , we finally get the estimations of p, q, v_1, v_2, v_3 .

8. Testing the Estimated Model

Now as another concrete example of the method as well as a validation procedure, we will simulate a large sample with strong-dependence, where the common distribution of all the data is a mixture of three Fréchet distributions. In this case we will first estimate the parameters of a mixture of two Fréchet distributions. The estimated model will be tested, and rejection is expected. Further, we will use again our method but to estimate the parameters of a mixture of three Fréchet distributions, and in this case it is expected that the goodness of fit test does not reject the estimated model.

Therefore, here we consider as the true model a mixture of three Fréchet distributions of order 1, with parameters $p = 0.3, q = 0.3, v_1 = 0.55, v_2 = 0.9, v_3 = 0.2$, that is:

$$MF(x) = 0.3\Phi_1\left(\frac{x}{0.55}\right) + 0.3\Phi_1\left(\frac{x}{0.9}\right) + 0.4\Phi_1\left(\frac{x}{0.2}\right)$$

We computed 4000 maximums, each one coming from samples of size 1000 of the simulation procedure described in section 0.6, with parameters $p = 0.3, q = 0.3, \delta = 0.55, \eta = 0.9, \tau = 0.2$. As mentioned before in section 0.6, by Theorem 1 of [10], these maximum should follow a distribution very close to our choice of MF .

We first proposed for fitting (*i.e.*, as H_0 in our test) a mixture of two Fréchet distributions with $\alpha = 1$ ($M2$). In this context, for $M2$, the estimated parameters were: $p = 0.625, v_1 = 0.28, v_2 = 0.76$, and the corresponding KS statistic was 0.0326095, what means that p -value $\ll 0.001$, implying a clear rejection.

Then, we proposed for fitting (as H_0) a mixture of three Fréchet distributions with $\alpha = 1$ ($M3$). In this context, for $M3$, the estimated parameters were: $p = 0.285, q = 0.334, v_1 = 0.51, v_2 = 0.86, v_3 = 0.29$ and the KS statistic was 0.01936157 with a p -value > 0.10 , non-rejecting H_0 .

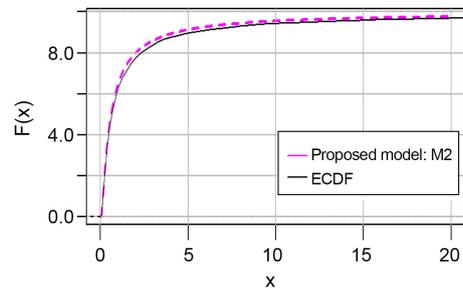


Figure 4. The difference between the empirical distribution (*ECDF*), and the theoretical distribution of the proposed model *M2*.

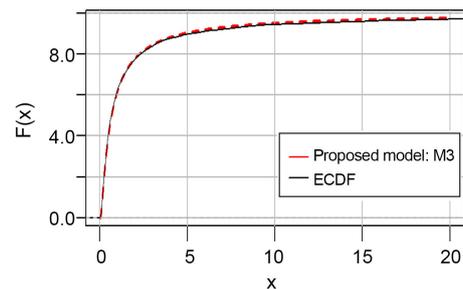


Figure 5. The difference between the empirical distribution (*ECDF*), and the theoretical distribution of the proposed model *M3*.

In **Figure 4**, we can appreciate a moderate deviation of the proposed *M2* model with respect to the empirical distribution, but this discrepancy is systematic, in the sense that most of the time the proposed model is above the empirical distribution, what means that real data have heavier tails, what is coherent with a very small p -value.

In **Figure 5**, the proposed *M3* model and the empirical distribution are almost equal, what is coherent with the no rejection decision of the test.

9. Discussion & Conclusions

Finite mixtures of extremal distributions appear in practice when dealing with environmental data (as well as in other fields) with a strong dependence structure. Therefore, one needs to be able to estimate the parameters of such a mixture under strong dependence, and test whether data fits to the estimated mixture.

In this paper, we successfully accomplish this task for the case of a mixture of two or three extremal distributions of the Fréchet type. The results obtained in simulated data show that this new estimation procedure developed here has an efficient performance.

Therefore, this work completes a line of research that includes [10] and [11], what obviously make new questions and subjects of interest arise.

10. Further Work

As pointed out, the estimation of the order α should be improved in a similar

way as the classical methods for the *iid* case [17] [18]. In addition, the asymptotic distribution of the weights and scale parameters, and their corresponding confidence regions can be more precisely exposed following the ideas mentioned in *Remark 4*.

In the case that methods based on moments (instead of quantiles) are applicable [19], an alternative method must be developed and its performance compared to the estimation procedure of this paper should be studied.

Another direction of work is the study of mixtures of different types of extremal distributions, or mixtures of extremal distributions and non-extremal ones, or more general finite mixture models under strong dependence, as it has been done in the *iid* case [4] [5] [6] [8] [9].

In a forthcoming paper, we deal with the problem of the estimation of the components in larger dimensions mixtures (k large) by using other techniques (*i.e.*, Machine Learning) for faster estimations of k .

Acknowledgements

This work was partial supported by *Proyecto CSIC-VUSP “Análisis de eventos climáticos extremos y su incidencia sobre la producción hortifrutícola en Salto” (Uruguay)*. Authors also thank to an anonymous reviewer for his valuable suggestions.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Wolfe, J.H. (1970) Pattern Clustering by Multivariate Mixture Analysis. *Multivariate Behavioral Research*, **5**, 329-350. https://doi.org/10.1207/s15327906mbr0503_6
- [2] Everitt, B.S. (1981) A Monte Carlo Investigation of the Likelihood Ratio Test for the Number of Components in a Mixture of Normal Distributions. *Multivariate Behavioral Research*, **16**, 171-180. https://doi.org/10.1207/s15327906mbr1602_3
- [3] Hathaway, R.J. (1986) Another Interpretation of the EM Algorithm for Mixture Distributions. *Statistics Probability Letters*, **4**, 53-56. [https://doi.org/10.1016/0167-7152\(86\)90016-7](https://doi.org/10.1016/0167-7152(86)90016-7)
- [4] Titterton, D.M., Smith, A.M.F. and Makov, U.E. (1985) *Statistical Analysis of Finite Mixture Distributions*. Wiley, Chichester.
- [5] McLachlan, G.J. and Peel, D. (1985) *Finite Mixture Models*. Wiley, New York.
- [6] Otiniano, C.E.G., Gonalves, C.R. and Dorea C.C.Y. (2017) Mixture of Extreme-Value Distributions: Identifiability and Estimation. *Communications in Statistics—Theory and Methods*, **46**, 6528-6542. <https://doi.org/10.1080/03610926.2015.1129423>
- [7] Tendijck, S., Eastoe, E., Tawn, J., Randell, D. and Jonathan, P. (2021) Modeling the Extremes of Bivariate Mixture Distributions with Application to Oceanographic Data. *Journal of the American Statistical Association*, **118**, 1373-1384.
- [8] Kollu, R., Rayapudi, S.R., Narasimham, S., *et al.* (2012) Mixture Probability Distri-

- bution Functions to Model Wind Speed Distributions. *International Journal of Energy and Environmental Engineering*, **3**, Article No. 27. <https://doi.org/10.1186/2251-6832-3-27>
- [9] Fahmi, K.J. and Al Abbasi, J.N. (1987) Mixture Distributions—An Alternative Approach for Estimating Maximum Magnitude Earthquake Occurrence. *Geophysical Journal International*, **89**, 741-747. <https://doi.org/10.1111/j.1365-246X.1987.tb05190.x>
- [10] Crisci, C. and Perera, G. (2022) Asymptotic Extremal Distribution for Non-Stationary, Strongly-Dependent Data. *Advances in Pure Mathematics*, **12**, 479-489. <https://doi.org/10.4236/apm.2022.128036>
- [11] Crisci, C., Perera, G. and Sampognaro, L. (2023) Goodness-of-Fit Test for Non-Stationary and Strongly Dependent Samples. *Advances in Pure Mathematics*, **13**, 226-236. <https://doi.org/10.4236/apm.2023.135016>
- [12] Embrechts, P., Kluppelberg, C. and Mikosch, T. (1997) Modelling Extremal Events for Insurance and Finance. Springer, New York. <https://doi.org/10.1007/978-3-642-33483-2>
- [13] Katz, R.W., Brush, G.S. and Parlange, M.B. (2005) Statistics of Extremes: Modeling Ecological Disturbances. *Ecology*, **86**, 1124-1134. <https://doi.org/10.1890/04-0606>
- [14] Reiss, R.D. and Thomas, M. (2007) Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields. Springer, Birkhäuser.
- [15] Batt, R.D., Carpenter, S.R. and Ives, A.R. (2005) Extreme Events in Lake Ecosystem Time Series. *Limnology and Oceanography Letters*, **2**, 63-69. <https://doi.org/10.1002/lol2.10037>
- [16] Vaart, A.W. (1998) Asymptotic Statistics (Cambridge Series in Statistical and Probabilistic Mathematics). Cambridge University Press, Cambridge. <https://doi.org/10.1017/CBO9780511802256>
- [17] Ramos, P.L., Louzada, F., Ramos, E. and Dey, S. (2005) The Fréchet Distribution: Estimation and Application—An Overview. *Journal of Statistics and Management Systems*, **23**, 549-578. <https://doi.org/10.1080/09720510.2019.1645400>
- [18] Huang, C., Lin, J.G. and Ren, Y.Y. (2005) Testing for the Shape Parameter of Generalized Extreme Value Distribution Based on the Lq-Likelihood Ratio Statistic. *Metrika*, **76**, 641-671. <https://doi.org/10.1007/s00184-012-0409-5>
- [19] Luong, A. (2020) Generalized Method of Moments and Generalized Estimating Functions Based on Probability Generating Function for Count Models. *Open Journal of Statistics*, **10**, 516-539. <https://doi.org/10.4236/ojs.2020.103031>