Chapter 1

On the variance of the Madogram with extreme value copula

1 Introduction

Management of environmental resources often requires the analysis of multivariate extreme values. In the classical theory, one is often interested in the behavior of the mean or average of a random variable X defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. This average will then be described through the expected value $\mathbb{E}[X]$ of the distribution. The central limit theorem yields, under some assumptions on the moments, the asymptotic behavior of the sample mean $\bar{X}$. This result can be used to provide a confidence interval for $\mathbb{E}[X]$ for a level $\alpha \in [0,1]$. But in case of extreme events, it can be just as important to estimate tails probabilities. Furthermore, what if the second moment $\mathbb{E}[X^2]$ or even the mean is not finite? Then the central limit theorem does not apply and the classical theory, carried by the normal distribution, is no longer relevant [Beirlant et al., 2004].
Some extreme events, such as heavy precipitation or wind speed has spatial characteristics and geostatisticians are striving to better understand the physical processes in hand. In geostatistics, we often consider $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space, S a set of locations and $(E, \mathcal{E})$ a measurable state space. We define on this probability space a stochastic process $X = \{X_s, s \in S\}$ with values on $(E, \mathcal{E})$. It is classical to define the following second-order statistic as the variogram (see [Gaetan and Guyon, 2008] Chapter 1.3 for definition and basic properties) :

$$2\gamma(h) = \mathbb{E}[|X(s+h) - X(s)|^2],$$

where $\{X(s), s \in S\}$ represents a spatial and stationary process with a well-defined covariance function. The function $\gamma(\cdot)$ is called the semi-variogram of X. With respect to extremes, this definition is not well adapted because a second order statistic is difficult to interpret inside the framework of extreme value theory or may not even be defined. To ensure that we always work with finite moments quantities, the following type of first-order variogram is introduced :

$$\nu(h) = \frac{1}{2}\mathbb{E}\left[|F(X(s+h)) - F(X(s))|\right], \tag{1.1}$$

where $F(u) = \mathbb{P}(X(s) \leq u)$ is named as the FMadogram. Let us define the pairwise extremal dependence function (Section 4.3 of [Coles et al., 1999]) such as :

$$V_h(x,y) = \int_0^1 max\left(\frac{w}{x}, \frac{1-w}{y}\right) 2dH_h(w), \tag{1.2}$$

where x, y are two reals and $H_h(\cdot)$ is a distribution function on $[0,1]$ such that $\int_{[0,1]} wdH_h(w) = 0.5$. It has been shown ([Cooley et al., 2006]) that $\nu(h)$ fully characterizes the extremal coefficient $V_h(1,1)$ since, using $2^{-1}|a-b| = max(a,b) - 2^{-1}(a+b)$, we have the following relationship,

$$V_h(1,1) = \frac{1+2\nu(h)}{1-2\nu(h)}. \tag{1.3}$$

Then, the estimation and the study of the madogram gives us an estimator of the extremal coefficient's estimator. We point out that this identity whose equation (1.3) is raised also permits a multivariate extension in higher dimension of the FMadogram. This approach was tackled by [Marcon et al., 2017] using a multivariate madogram in order to estimate the Pickands dependence function. This method extends [Capéraà et al., 1997] which proposes a non-parametric estimator to estimate the Pickands dependence function for bivariate extreme value copulas (see Equation (1.5) in Section 2). Let's go back to the estimation of our extremal coefficient, his main drawback is that it only focuses on the values $V_h(x,x)$ but does not provide any information about $V_h(x,y)$, for $x \neq y$. To overpass this drawback, [Naveau et al., 2009] introduce the λ-FMadogram defined as,

$$\nu(h,\lambda) = \frac{1}{2}\mathbb{E}\left[\left|F^{\lambda}(X(s+h)) - F^{1-\lambda}(X(s))\right|\right], \tag{1.4}$$

for every $\lambda \in [0,1]$. It is shown in the same paper that the λ-Fmadogram fully characterizes the dependence function $V_h(x,y)$ with the following relationship,

$$V_h(\lambda, 1-\lambda) = \frac{c(\lambda) + \nu(h,\lambda)}{1 - c(\lambda) - \nu(h,\lambda)}.$$

Furthermore, this statistic kept our attention because it can be seen as a dissimilarity measure among bivariate maxima to be used in a clustering algorithm [Bernard et al., 2013]. This first chapter aims to study the variance of the λ-madogram with the fewest possible assumptions. In our knowledge, only [Guillou et al., 2014] has computed the variance of the sole madogram in Equation (1.1) by assuming independence and found $1/90$.

This chapter is organized as follows, in Section 3, we introduce our estimator and we discuss its properties. An explicit formula for the asymptotic variance is also given. In

Section 4, we investigate the finite-sample performance of the estimator by means of Monte Carlo simulations. All proofs are postponed to the final section.

2 Notations and preliminaries

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and (X, Y) be a bivariate random vector with values in $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$. This random vector has a joint distribution function H and marginal distribution function F and G. A function $C : [0,1]^2 \rightarrow [0,1]$ is called a bivariate copula if it is the restriction to $[0,1]^2$ of a bivariate distribution function whose marginals are given by the uniform distribution on the interval $[0,1]$. Since the work of [Sklar, 1959], it is well known that every distribution function H can be decomposed as $H(x,y) = C(F(x), G(y))$, for all $(x,y) \in \mathbb{R}^2$. This function C characterizes the dependence between X and Y and is called an extreme value copula if and if it admits a representation of the form [Gudendorf and Segers, 2009]

$$C(u,v) = (uv)^{A(log(v)/log(uv))}, \tag{1.5}$$

for all $u, v \in [0,1]$ and where $A(\cdot)$ is the Pickands dependence function, *i.e.*, $A : [0,1] \longrightarrow [1/2, 1]$ is convex and satisfies $t \vee (1-t) \leq A(t) \leq 1$, $\forall t \in [0,1]$. The upper and lower bound of A has special meanings, the upper bound $A(t) = 1$ corresponds to independence, whereas the lower bound $A(t) = t \vee 1 - t$ corresponds to the perfect dependence (comonotonicity). Notice that, on sections, the extreme value copula is of the form

$$C(u^t, u^{1-t}) = u^{A(t)}. \tag{1.6}$$

Let $(X_t)_{t=1,\dots,T}$ be an *i.i.d.* sample of a bivariate random vector whose underlying copula is denoted by C and whose margins by F, G. For $x, y \in \mathbb{R}$, let $x \wedge y = min(x,y)$ and $x \vee y = max(x,y)$. Let $(b_{t,j})_{t \geq 1, j \in \{1,2\}}$ and $(a_{t,j})_{t \geq 1, j \in \{1,2\}}$ be respectively a sequence of numbers and a sequence of positive numbers. We say that the sequence $(a_{t,1}^{-1}(\bigvee_{t=1}^T X_t - b_{t,1}), a_{t,2}^{-1}(\bigvee_{t=1}^T Y_t - b_{t,2}))$ belongs to the domain of attraction of H, if for all real values x, y (at which the limit is continuous)

$$\mathbb{P}\left(\frac{\bigvee_{t=1}^T X_t - b_{t,1}}{a_{t,1}} \leq x, \frac{\bigvee_{t=1}^T Y_t - b_{t,2}}{a_{t,2}} \leq y \right) \underset{T \rightarrow \infty}{\rightarrow} H(x,y).$$

If this relationship hold, we would say that H is a multivariate extreme value distribution. We will call by FMadogram the following quantity

$$\nu = \frac{1}{2}\mathbb{E}\left[|F(X) - G(Y)| \right]. \tag{1.7}$$

We will write the generalized inverse function of F (respectively G) as $F^{\leftarrow}(u) = inf\{v \in \mathbb{R} | F(v) \geq u\}$ (respectively $G^{\leftarrow}(u) = inf\{v \in \mathbb{R} | G(v) \geq u\}$) where $0 < u, v < 1$. Given $\mathcal{X} \subset \mathbb{R}^2$, let $l^{\infty}(\mathcal{X})$ denote the spaces of bounded real-valued function on $\mathcal{X}$. For $f : \mathcal{X} \rightarrow \mathbb{R}$, let $||f||_{\infty} = sup_{x \in \mathcal{X}} |f(x)|$. The arrows $\xrightarrow{a.s.}$, $\xrightarrow{d}$ and $\rightsquigarrow$ denote almost sure convergence, convergence in distribution of random vectors and weak convergence in the sense of J.Hoffman-Jørgensen (see Part 1 in the monograph by [van der Vaart and Wellner, 1996]).

3 Weak convergence of the λ-FMadogram

For the rest of this report we will assume that the copula C is of extreme value type as defined in Equation (1.5). Following [Fermanian et al., 2004], to guarantee the weak convergence of our empirical copula process, we introduce the following assumptions.

Assumption A. *(i) The bivariate distribution function H has continuous margins F, G.*

(ii) The derivative of the Pickands dependence function $A'(t)$ exists and is continuous on $(0,1)$.

The Assumption A *(i)* guarantee the uniqueness of the representation $H(x,y) = C(F(x), G(y))$ on the range of (F,G). Under the Assumption A *(ii)*, the first-order partial derivatives of C with respect to u and v exists and are continuous on the set $\{(u,v) \in [0,1]^2 : 0 < u, v < 1\}$. Indeed, we have

$$\frac{\partial C(u,v)}{\partial u} = \begin{cases} \frac{C(u,v)}{u} \left(A(\frac{log(v)}{log(uv)}) - A'(\frac{log(v)}{log(uv)}) \frac{log(v)}{log(uv)} \right), & \text{if } 0 < u, v < 1, \\ 0, & \text{if } v = 0, \quad 0 < u < 1, \end{cases}$$

$$\frac{\partial C(u,v)}{\partial v} = \begin{cases} \frac{C(u,v)}{v} \left(A(\frac{log(v)}{log(uv)}) + A'(\frac{log(v)}{log(uv)}) \frac{log(u)}{log(uv)} \right), & \text{if } 0 < u, v < 1, \\ 0, & \text{if } u = 0, \quad 0 < v < 1, \end{cases}$$

The properties of A imply $0 \leq A(t) - tA'(t) \leq 1$ and $0 \leq A(t) + (1-t)A'(t) \leq 1$ where $t = log(v)/log(uv)$ (see Lemma 1 in Section 5.1). Therefore if $v \searrow 0$, then $\partial C(u,v)/\partial u \rightarrow 0$ as required. The FMadogram (see Equation (1.7)) is an estimator commonly used with extrema due to its relation with the pairwise extremal dependence coefficient defined in Equation (1.2) (for example [Cooley et al., 2006] or [Guillou et al., 2014]). In this study, we aim to analyze the asymptotic variance structure of the λ-FMadogram defined in [Naveau et al., 2009].

Definition 1. *Let (X,Y) be a bivariate random vector with margins F and G. A λ-FMadogram is the quantity defined by :*

$$\nu(\lambda) = \frac{1}{2}\mathbb{E}[|F^{\lambda}(X) - G^{1-\lambda}(Y)|]. \tag{1.8}$$

Having a sample $(X_1, Y_1), \dots, (X_T, Y_T)$ of T bivariate vector with unknown margins F and G, we construct the empirical distribution function :

$$\hat{H}_T(x, y) = \frac{1}{T} \sum_{t=1}^{T} \mathbb{1}_{\{X_t \leq x, Y_t \leq y\}},$$

and let $\hat{F}_T(x)$ and $\hat{G}_T(y)$ be its associated marginal distributions, that is,

$$\hat{F}_T(x) = \hat{H}_T(x, +\infty), \quad \text{and,} \quad \hat{G}_T(y) = \hat{H}_T(+\infty, y), \quad -\infty < x, y < +\infty.$$

Based on these identical and independent copies $(X_1, Y_1), \dots, (X_T, Y_T)$, it is natural to define the following estimator of the λ-Madogram:

$$\hat{\nu}_T(\lambda) = \frac{1}{2T} \sum_{t=1}^{T} \left| \hat{F}_T^{\lambda}(X_t) - \hat{G}_T^{1-\lambda}(Y_t) \right|. \tag{1.9}$$

Concerning the boundaries of the estimators, the λ-FMadogram defined by (1.8) satisfies $\nu(0) = \nu(1) = 0.25$. Hence we can force our estimator to satisfy $\hat{\nu}_T(0) = \hat{\nu}_T(1) = 0.25$. This leads to the following definition :

$$\begin{aligned} \hat{\nu}_T^*(\lambda) = \frac{1}{2T} \sum_{t=1}^{T} \left| \hat{F}_T^{\lambda}(X_t) - \hat{G}_T^{1-\lambda}(Y_t) \right| - \frac{\lambda}{2T} \sum_{t=1}^{T} \{1 - \hat{F}_T^{\lambda}(X_t)\} \\ - \frac{1-\lambda}{2T} \sum_{t=1}^{T} \{1 - \hat{G}_T^{1-\lambda}(Y_t)\} + \frac{1}{2} \frac{1 - \lambda + \lambda^2}{(2-\lambda)(1+\lambda)}. \end{aligned} \tag{1.10}$$

In the following proposition, we establish some properties of the λ-FMadogram.

Proposition 1. *Let $(X_1, Y_1), \dots, (X_T, Y_T)$ a sample of $\mathbb{R}^2$-valued independent random vector. We have, for each $\lambda \in [0, 1]$,*

(i) $0 \leq \nu(\lambda) \leq \frac{1}{1+\lambda(1-\lambda)} - \frac{1}{2}\left(\frac{1}{1+\lambda} + \frac{1}{1+1-\lambda}\right)$,

(ii) $\hat{\nu}_T(\lambda) \xrightarrow{a.s.} \nu(\lambda)$,

(iii) $\nu(0) = \nu(1) = 0.25$, *and if* $\lambda \in (0, 1)$,

$$\nu(\lambda) = \frac{A(\lambda)}{A(\lambda) + \lambda(1-\lambda)} - \frac{1}{2}\left(\frac{1}{1+\lambda} + \frac{1}{1+1-\lambda}\right). \tag{1.11}$$

The proof is given in Section 5.2. Note that the upper bound in *(i)* is exactly the value of the λ-FMadogram when X and Y are independent, *i.e.* when $A(t) = 1$. Define the empirical copula function $\hat{C}_T(u,v)$ by,

$$\hat{C}_T(u,v) = \hat{H}_T(\hat{F}_T^{\leftarrow}(u), \hat{G}_T^{\leftarrow}(v)), \quad 0 \leq u, v \leq 1,$$

and the (ordinary) empirical copula process,

$$\mathbb{C}_T(u,v) = \sqrt{n}(\hat{C}_T - C)(u,v), \quad 0 \leq u, v \leq 1.$$

The weak convergence of $\mathbb{C}_T$ has already been proved by [Fermanian et al., 2004] using previous results on the Hadamard differentiability of the map $\phi : D([0,1]^2) \rightarrow l^\infty([0,1]^2)$ which transforms the cumulative distribution function H into its copula function C (see lemma 3.9.28 from [van der Vaart and Wellner, 1996]). Some auxiliary results concerning the convergence of $\mathbb{C}_T$ to a Gaussian process and the convergence of the λ-FMadogram to a centered Gaussian law are recalled in Appendix. The limiting Gaussian process of $\mathbb{C}_T$ can be written as

$$N_C(u,v) = B_C(u,v) - \frac{\partial C}{\partial u}(u,v)B_C(u,1) - \frac{\partial C}{\partial u}(u,v)B_C(1,v), \tag{1.12}$$

where B_C is a Brownian bridge in $[0,1]^2$ with covariance function

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

Furthermore, the asymptotic variance of the λ-FMadogram can be written as:

$$\int_{[0,1]^2} f(x,y)dJ(x,y) = \frac{1}{2}\int_{[0,1]} f(0, y^{1/(1-\lambda)})dy + \frac{1}{2}\int_{[0,1]} f(x^{1/\lambda}, 0)dx - \int_{[0,1]} f(x^{1/\lambda}, x^{1/(1-\lambda)})dx, \tag{1.13}$$

where $J(x,y) = 2^{-1}|x^\lambda - y^{1-\lambda}|$. Some details explaining Equation (1.13) are given in Lemma 3 in section 5. Now, one may consider the weak convergence of the stochastic process $(\sqrt{T}(\hat{\nu}_T(\lambda) - \nu(\lambda)))_{\lambda \in [0,1]}$ in $l^\infty([0,1])$. To establish such a result, we use empirical process arguments as formulated in [van der Vaart and Wellner, 1996]. This allows us to show the following theorem.

Theorem 1. *Let $\lambda \in [0,1]$. Under Assumption A, then in $l^\infty([0,1])$ (je spécifie l'espace dans lequel vivent les trajectoires du processus), as $T \rightarrow \infty$,*

$$\sqrt{T}(\hat{\nu}_T(\lambda) - \nu(\lambda)) \rightsquigarrow \left(-\int_{[0,1]} N_C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}})du \right)_{\lambda \in [0,1]}.$$

The covariance structure of the limiting process is given for $\lambda, \mu \in [0,1]$

$$Cov\left(\left(\int_{[0,1]} N_C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}}) du \right), \left(\int_{[0,1]} N_C(u^{\frac{1}{\mu}}, u^{\frac{1}{1-\mu}}) du \right) \right) = \int_{[0,1]^2} \mathbb{E}\left[N_C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}}) N_C(v^{\frac{1}{\mu}}, v^{\frac{1}{1-\mu}}) \right] duv. \tag{1.14}$$

Proof of Theorem 1 may be found in Section 5.4. Furthermore, we are able to prove that the asymptotic behavior of $(\sqrt{T}\hat{\nu}_T^*(\lambda) - \nu(\lambda))_{\lambda \in [0,1]}$ is the same as $\sqrt{T}(\hat{\nu}_T(\lambda) - \nu(\lambda))_{\lambda \in [0,1]}$, we refer the reader to Section 5.5 for details. We now want to give a closed expression for the variance of our process, *i.e.* when $\lambda = \mu$ in Equation (1.14). As N_C is a tight Gaussian process, we know that the random variable, for a fixed $\lambda \in [0,1]$,

$$-\int_{[0,1]} N_C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}}) du,$$

is a centered Gaussian random variable for every $\lambda \in [0,1]$ as it is a linear transformation of a tight Gaussian process. Furthermore, its variance is given by :

$$\begin{aligned} Var\left(\int_{[0,1]} N_C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}}) du \right) = Var\Bigg(&\int_{[0,1]} B_C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}}) du - \int_{[0,1]} B_C(u^{\frac{1}{\lambda}}, 1) \frac{\partial C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}})}{\partial u} du \\ &- \int_{[0,1]} B_C(1, v^{\frac{1}{1-\lambda}}) \frac{\partial C(v^{\frac{1}{\lambda}}, v^{\frac{1}{1-\lambda}})}{\partial v} \Bigg). \end{aligned}$$

Using the property exhibited in Equation (1.6). We may find a similar pattern for partial derivatives,

$$\begin{aligned} \frac{\partial C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}})}{\partial u} &= \frac{C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}})}{u^{\frac{1}{\lambda}}} \left(A(\lambda) - A'(\lambda)\lambda \right), \\ \frac{\partial C(v^{\frac{1}{\lambda}}, v^{\frac{1}{1-\lambda}})}{\partial v} &= \frac{C(v^{\frac{1}{\lambda}}, v^{\frac{1}{1-\lambda}})}{v^{\frac{1}{1-\lambda}}} \left(A(\lambda) + A'(\lambda)(1-\lambda) \right). \end{aligned}$$

Let $\mathcal{A}$ be the space of Pickands dependence functions. We will denote by $\kappa(\lambda, A)$ and $\zeta(\lambda, A)$ two functional such as :

$$\kappa \colon [0,1] \times \mathcal{A} \to [0,1], \quad (\lambda, A) \mapsto A(\lambda) - A'(\lambda)\lambda,$$

$$\zeta \colon [0,1] \times \mathcal{A} \to [0,1], \quad (\lambda, A) \mapsto A(\lambda) + A'(\lambda)(1-\lambda).$$

Furthermore, the integral $\int_{[0,1]^2} C(u,v) duv$ does not admit, in general, a closed form. But we are able to express it with respect to a simple integral of the Pickands dependence function. We note, for notational convenience the following functional

$$f \colon [0,1] \times \mathcal{A} \rightarrow [0,1], \quad (\lambda, A) \mapsto \left(\frac{\lambda(1-\lambda)}{A(\lambda) + \lambda(1-\lambda)} \right)^2.$$

For a fixed $\lambda \in [0,1]$, using properties of the extreme value copula permit us to give an explicit formulas of the asymptotic variance of the normalized error of the λ-Madogram $\sqrt{T}\,(\hat{\nu}_T(\lambda) - \nu(\lambda))$. Let $A_1(\lambda) = A(\lambda)/\lambda$, $A_2(\lambda) = A(\lambda)/(1-\lambda)$. Before presenting the theorem, let us introduce some notation for convenience.

$$\gamma_1^2 = \frac{A(\lambda)}{A(\lambda) + 2\lambda(1-\lambda)}, \quad \gamma_2^2 = \frac{\kappa(\lambda, A)^2(1-\lambda)}{2A(\lambda) - (1-\lambda) + 2\lambda(1-\lambda)}, \quad \gamma_3^2 = \frac{\zeta(\lambda, A)^2 \lambda}{2A(\lambda) - \lambda + 2\lambda(1-\lambda)},$$

$$\gamma_{12} = \kappa(\lambda, A)\Bigg(f(\lambda, A) \left(\frac{(1-\lambda)^2 - A(\lambda)}{2A(\lambda) - (1-\lambda) + 2\lambda(1-\lambda)} \right)$$
$$+ \lambda(1-\lambda) \int_{[0,\lambda]} [A(s) + (1-s)(A_2(\lambda) - (1-\lambda) - 1) - s\lambda + 1]^{-2} \, ds \Bigg),$$

$$\gamma_{13} = \zeta(\lambda, A)\Bigg(f(\lambda, A) \left(\frac{\lambda^2 - A(\lambda)}{2A(\lambda) - \lambda + 2\lambda(1-\lambda)} \right)$$
$$+ \lambda(1-\lambda) \int_{[\lambda,1]} [A(s) + s(A_1(\lambda) - 1 - \lambda) - (1-s)(1-\lambda) + 1]^{-2} \, ds \Bigg),$$

$$\gamma_{23} = \kappa(\lambda, A)\zeta(\lambda, A)\Bigg(- f(\lambda, A)$$
$$+ \lambda(1-\lambda) \int_{[0,1]} [A(s) + (1-s)(A_2(\lambda) - (1-\lambda) - 1) + s(A_1(\lambda) - \lambda - 1) + 1]^{-2} \, ds \Bigg).$$

Theorem 2. *For $\lambda \in [0,1]$, the asymptotic variance of $\sqrt{T}\,(\hat{\nu}_T(\lambda) - \nu(\lambda))$ has the following closed form,*

$$Var\left(\int_{[0,1]} N_C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}}) du \right) = f(\lambda, A) \left(\gamma_1^2 + \gamma_2^2 + \gamma_3^2 \right) - 2\gamma_{12} - 2\gamma_{13} + 2\gamma_{23} \tag{1.15}$$

Proof is given in Section 5.6. From Theorem 2, we are able to infer the closed form of the λ-Madogram's variance in the case of an independent Copula, *i.e.* when $C(u,v) = uv$. Indeed, we just have to take $A(t) = 1$ for every $t \in [0,1]$. We thus obtain that $\kappa(\lambda, A) = \zeta(\lambda, A) = 1$ for every $\lambda \in [0,1]$. This result is summarised in the following statement:

Corollary 1. *Under the framework of theorem 2 and if $C(u,v) = uv$, then the asymptotic variance of $\sqrt{T}(\hat{\nu}_T(\lambda) - \nu(\lambda))$ has the following form, for $\lambda \in (0,1)$*

$$Var\left(\int_{[0,1]} N_C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}}) du \right) = \left(\frac{\lambda(1-\lambda)}{1 + \lambda(1-\lambda)} \right)^2 \Bigg(\frac{1}{1 + 2\lambda(1-\lambda)}$$
$$- \frac{1-\lambda}{2 - (1-\lambda) + 2\lambda(1-\lambda)} - \frac{\lambda}{2 - \lambda + 2\lambda(1-\lambda)} \Bigg)$$

4 Simulation

4.1 Comparison with several models for the bivariate case

We present several models that would be used in the simulation section in order to assess our findings remains in finite-sample settings.

1. The asymmetric logistic model [Tawn, 1988] defined by the following dependence function :

$$A(t) = (1-\psi_1)t + (1-\psi_2)(1-t) + [(\psi_1 t)^\theta + (\psi_2(1-t))^\theta]^{\frac{1}{\theta}},$$

   with parameters $\theta \in [1, \infty[$, ψ_1, $\psi_2 \in [0,1]$. The special case $\psi_1 = \psi_2 = 1$ gives us the symmetric model of Gumbel. In the symmetric model, as we retrieve the independent case when $\theta = 1$, the dependence between the two variables is stronger as θ goes to infinity.

2. The asymmetric negative logistic model [Joe, 1990], namely,

$$A(t) = 1 - [(\psi_1(1-t))^{-\theta} + (\psi_2 t)^{-\theta}]^{-\frac{1}{\theta}},$$

   with parameters $\theta \in (0, \infty)$, $\psi_1, \psi_2 \in (0,1]$. The special case $\psi_1 = \psi_2 = 1$ returns the symmetric negative logistic of [Oliveira and Galambos, 1977].

3. The asymmetric mixel model [Tawn, 1988] :

$$A(t) = 1 - (\theta + \kappa)t + \theta t^2 + \kappa t^3,$$

   with parameters θ and κ satisfying $\theta \geq 0$, $\theta + 3\kappa \geq 0$, $\theta + \kappa \leq 1$, $\theta + 2\kappa \leq 1$. The special case $\kappa = 0$ and $\theta \in [0,1]$ yields the symmetric mixed model.

4. The model of Hüsler and Reiss [Hüsler and Reiss, 1989],

$$A(t) = (1-t)\Phi\left(\theta + \frac{1}{2\theta} log\left(\frac{1-t}{t}\right)\right) + t\Phi\left(\theta + \frac{1}{2\theta} log\left(\frac{t}{1-t}\right)\right),$$

   where $\theta \in (0, \infty)$ and Φ is the standard normal distribution function. As θ goes to 0^+, the dependence between the two variables is stronger. When θ goes to infinity, we are near independence.

5. The t-EV model [Demarta and McNeil, 2005], in which

$$A(w) = w t_{\chi+1}(z_w) + (1-w)t_{\chi+1}(z_{1-w}),$$
$$\text{with } z_w = (1+\chi)^{1/2}[w/(1-w)^{\frac{1}{\chi}} - \theta](1-\theta^2)^{-1/2},$$

and parameters $\chi > 0$, and $\theta \in (-1, 1)$, where $t_{\chi+1}$ is the distribution function of a Student-t random variable with $\chi + 1$ degrees of freedom.

A vast Monte Carlo study was used to illustrate Theorem 2 of Section 3 in finite-sample settings. Specifically, for each $\lambda \in [0, 1]$, 500 random samples of size $n = 256$ were generated from the Gumbel copula with $\theta \in \{1.0, 1.5, 2.0, 2.5, 3.0\}$. For each sample, the λ-FMadogram estimators were computed where the margins are unknown. For each estimator, the empirical version of the normalized estimation error's variance, namely

$$\mathcal{E}_T(\lambda) := Var\left(\sqrt{T}\left(\hat{\nu}_T^*(\lambda) - \nu(\lambda)\right)\right), \tag{1.16}$$

was computed by taking the variance over 500 samples. For each estimator, we represent its theoretical asymptotic variance using the integral exhibits in Theorem 2.

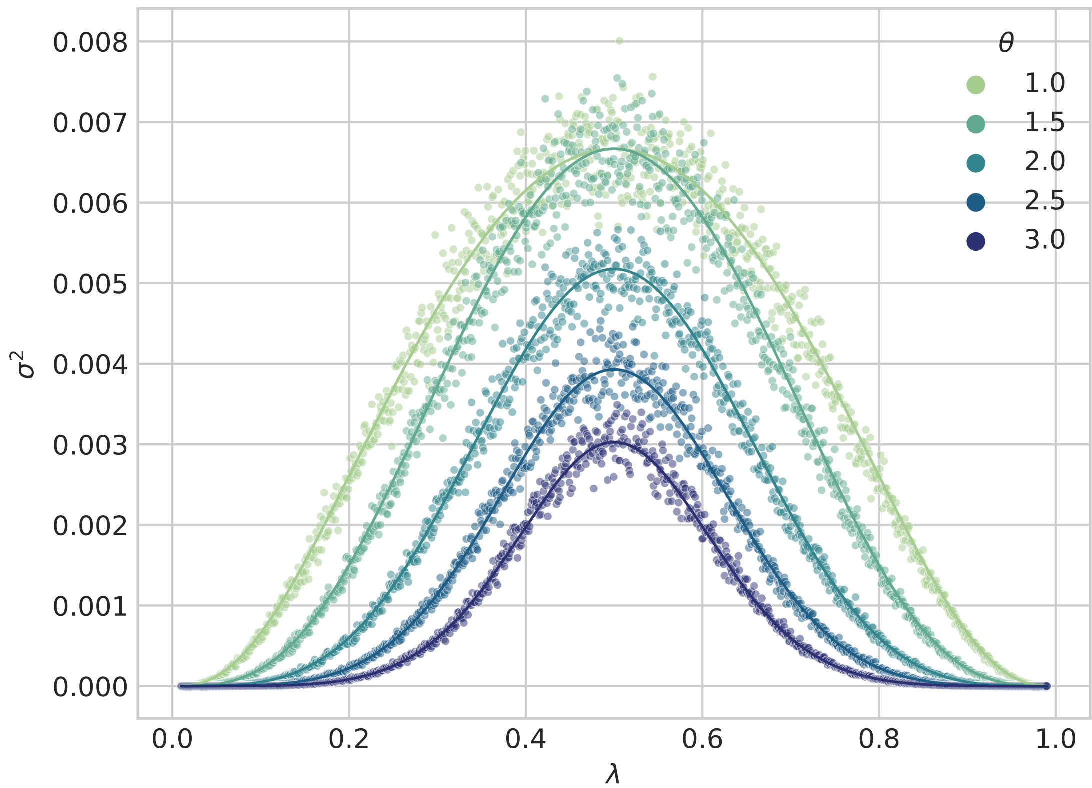


Figure 1: Variance of the normalized estimation error based on 500 samples of size $T = 256$ from the Gumbel copula with $\theta = \{1.0, 1.5, 2.0, 2.5, 3.0\}$ chosen in such a way that $\lambda \in \{i/1000 : i = 0.01, \ldots, 0.99\}$.

Similar results were obtained for many other extreme-value dependence models (see figure 2), we can note the following :

1. When A is symmetric, one would expect the asymptotic variance of the estimator to

reach its maximum at $\lambda = 1/2$. Such is not always the case, however, as illustrated by the t-EV model.

2. In the asymmetric negative logistic model, the asymptotic of the λ-FMadogram is close to zero for all $\lambda \in [0, 0.3]$. This is due to the fact that $A(\lambda) \approx 1 - t$ for this model.

Remarks 1. and 2. are also observed in [Genest and Segers, 2009]. We propose in Figure 3 the theoretical asymptotic variance depending of θ and λ for six model of extreme-value copula. The parameters are chosen accordingly to [Genest and Segers, 2009].

4.2 Non-monotonicity of the variance with respect to the dependence parameter

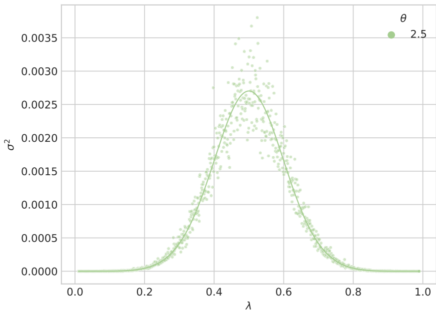
Looking at Figure 1, one may make the third remark :

3. Interestingly, as our variables (X, Y) are becoming more positively dependent (in Figure 1, as θ increase), then the asymptotic variance is, for every $\lambda \in [0, 1]$, lower or equal than the asymptotic variance in the independent case. As shown in the proof section 5.8 where we exhibit a counterexample, it is not always the case.

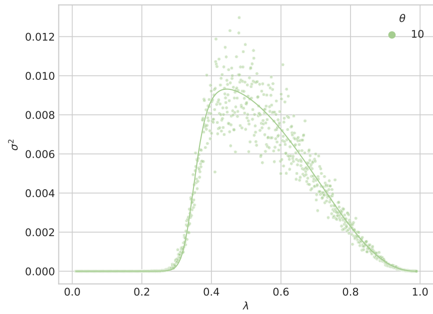
The Figure 4a shows the same model with different values for θ and λ. The moving average is computed out of 10 empirical variances for each θ. Each theoretical asymptotic variance depending of λ is fitted by its empirical counterpart represented by the moving average. As the dependency parameter θ increases, we can find some λ for which the asymptotic variance is greater than the asymptotic variance in the case of independance. That figure supports our counterexample that draws the same conclusion. Also, Figure 4b depict the asymptotic variance for a fixed $\lambda = 0.5$ for the symmetric mixed model with a varying $\theta \in [0, 1]$. When $\theta = 0$, we are turning back to the independent copula $C(u, v) = uv$ and it's asymptotic variance is given by $1/150$ for this value of λ. When the vector (X, Y) are positively dependent, *i.e.* when θ increase, the asymptotic variance for this given λ increase also, but after a certain threshold which depends of the chosen model, the variance starts to decrease.

4.3 Estimation on Max-Stable processes

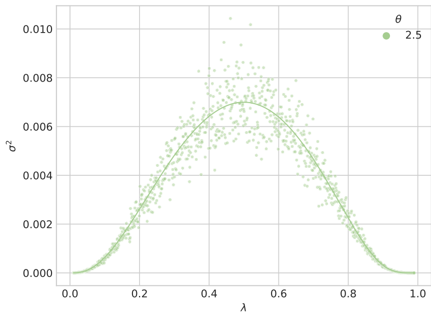
To determine the quality of the λ-FMadogram for estimating the pairwise dependence of maxima in space, [Naveau et al., 2009] compute on a particular class of simulated max-stable random fields. They focus on the Smith's max-stable process [Smith, 2005]. We recall the bivariate distribution for the max-stable process model proposed by Smith is equal to :



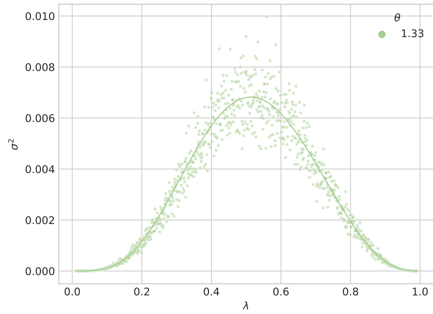
(a) **Asym. Neg. Logistic** ($\theta = 2.5$, $\psi_1 = 1.0$, $\psi_2 = 1.0$)



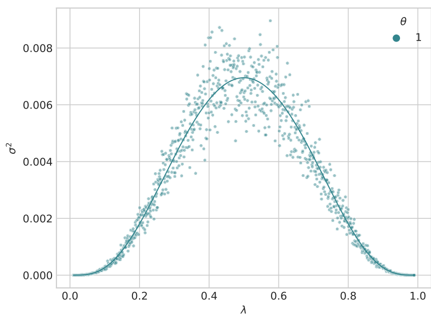
(b) **Asym. Neg. Logistic** ($\theta = 10$, $\psi_1 = 0.5$, $\psi_2 = 1.0$)



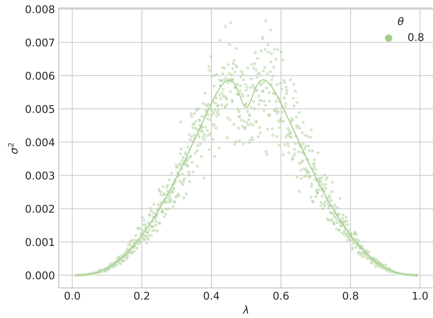
(c) **Asymmetric Logistic** ($\theta = \frac{5}{2}$, $\psi_1 = 0.1$, $\psi_2 = 1.0$)



(d) **Asymmetric Mixed** ($\theta = \frac{4}{3}$, $\kappa = -\frac{1}{3}$)

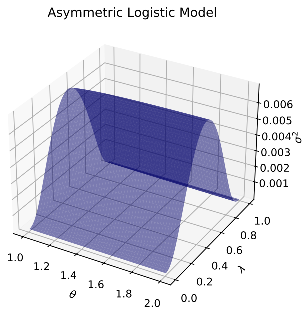


(e) **Hüsler-Reiss** ($\theta = 1$)

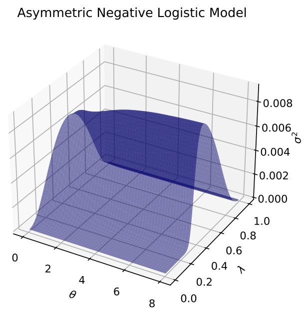


(f) **t-EV** ($\theta = 0.8$, $\chi = 0.2$)

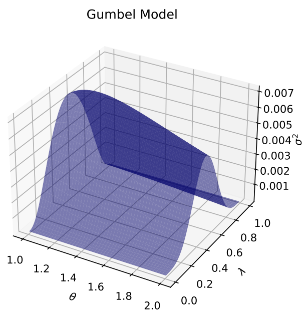
Figure 2: Graph, as a function of λ, of the asymptotic variances of the normalized estimation errors based on 500 samples of size $T = 256$ of the λ-FMadogram for six extreme-value copula models.



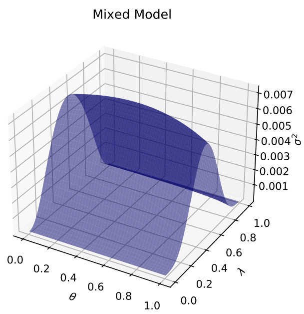
(a) **Asym. Logistic** ($\psi_1 = 0.1$, $\psi_2 = 1.0$)



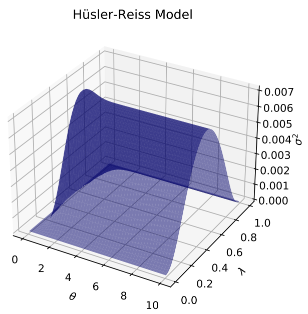
(b) **Neg. Logistic** ($\psi_1 = 0.5$, $\psi_2 = 1.0$)



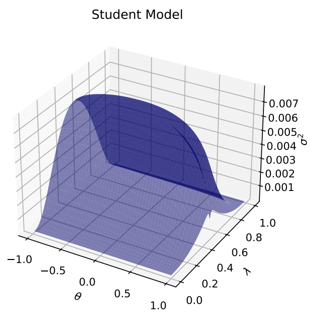
(c) **Gumbel** ($\psi_1 = 0.1$, $\psi_2 = 1.0$)



(d) **Symmetric Mixed** ($\kappa = -\frac{1}{3}$)

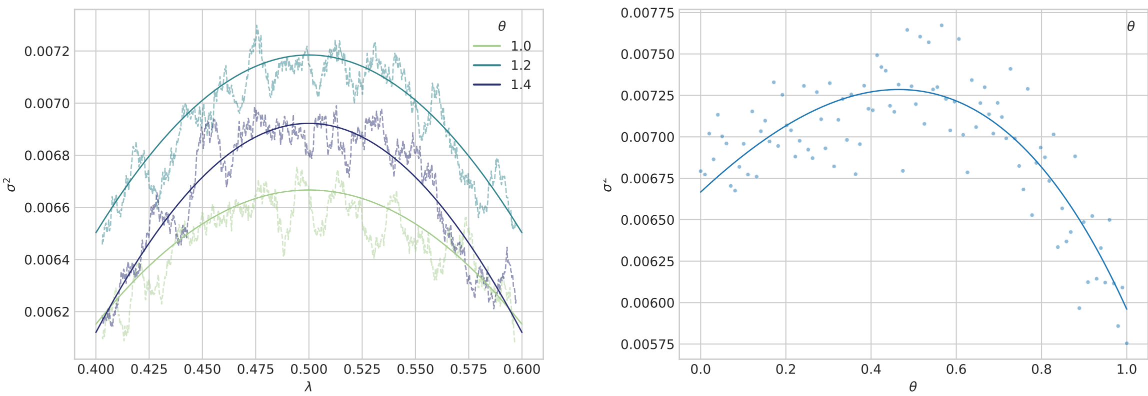


(e) **Hüsler-Reiss**



(f) **t-EV** ($\chi = 0.2$)

Figure 3: Graph, as a function of λ and the θ, of the asymptotic variance of the λ-FMadogram for six extreme-value copula models.



(a) (b)

Figure 4: Panel (a) depicts the variance ($\times 256$) of the estimators $\hat{\nu}(\lambda)$ based on 500 samples of size $T = 256$ from the Gumbel copula with $\theta = \{1.0, 1.2, 1.4\}$ chosen in such a way that $\lambda \in \{i/1000 : i = 0.4, \ldots, 0.6\}$. The dotted lines are moving averages made out of the 1000 empirical estimators of the variance. Panel (b) shows the variance ($\times 512$) of the estimators $\hat{\nu}(\lambda)$ based on 2000 sample of size $T = 512$ from the symmetric mixed model with $\lambda = 0.5$ chosen in such a way that $\theta \in \{i/100 : i = 0.0, \ldots 1.0\}$. The solid line is the asymptotic variance computed numerically using Theorem 2.

$$\mathbb{P}\left(X(s) \leq u, X(s+h) \leq v\right) = exp\left[-\frac{1}{u}\Phi(\frac{a}{2} + \frac{1}{a}log(\frac{v}{u})) - \frac{1}{v}\Phi(\frac{a}{2} + \frac{1}{a}log(\frac{u}{v}))\right], \quad (1.17)$$

where Φ denotes the standard normal distribution function, with $a^2 = (h^\top \Sigma^{-1} h)$ and Σ is a covariance matrix. In case of isotropic field, we set $\Sigma = \sigma^2 I_2$. We recall that, for this kind of process, the pairwise extremal dependence function $V_h(\cdot, \cdot)$ (see 1.2 for definition) is given by :

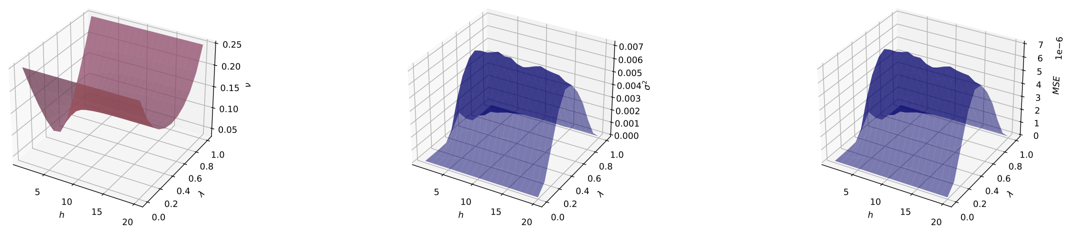
$$V_h(u, v) = \frac{1}{u}\Phi\left(\frac{a}{2} + \frac{1}{a}log(\frac{v}{u})\right) + \frac{1}{v}\Phi\left(\frac{a}{2} + \frac{1}{a}log(\frac{u}{v})\right). \quad (1.18)$$

Furthermore, for a max-stable process, the theoretical value of the λ-FMadogram is given by

$$\nu(h, \lambda) = \frac{V_h(\lambda, 1-\lambda)}{1 + V_h(\lambda, 1-\lambda)} - c(\lambda), \quad (1.19)$$

with $c(\lambda) = 3/\{2(1+\lambda)(1+1-\lambda)\}$ and for any $\lambda \in [0,1]$. This statement was shown in Proposition 1 of [Naveau et al., 2009]. The λ-FMadogram was estimated independently for each simulated field by using (1.10) with $T = 1024$. The z-axis correspond to the error of the estimator and the xy-space, $[0, 20] \times [0, 1]$, represent the distance h and parameter λ. In Smith's model, the pairwise dependence function between two locations s and $s+h$ decrease as the distance h between these two points increase.

The surface in Figure 5a provides the mean value of the estimated λ-FMadogram in blue, the true quantity is given by the surface in red. Figure 5c indicates the mean squared error between the estimated λ-FMadogram and the true value. As expected, the error is



(a) **Estimated value** (b) **variance** (c) **Mean Squared Error**

Figure 5: Simulation results obtained by generating 300 independently and identically distributed Smith random fields. The dependence structure is characterized by (1.17) with $\Sigma = 25 I_2$. Panel (5a) shows the estimated and the true λ-FMadogram. Panel (5b) shows the estimated variance ($\times 1024$) of the λ-FMadogram. Panel (5c) depicts the mean squared error between the true and estimated λ-FMadogram for all h and λ.

close to zero at the two boundary planes $\lambda = 0$ and $\lambda = 1$ by construction of the estimator. The largest mean squared errors are obtained where $\lambda = 0.5$, especially for very small distances, *i.e.* near $h = 0$. This behaviour is now well known from our discussion.

4.4 Simulation on block maxima

In this experiment, we derive the behavior of the asymptotic variance of componentwise maxima of i.i.d random vectors having a t copula distribution. A bivariate t copula is defined as :

$$C_{\chi,\theta}(u,v) = \int_{]-\infty, t_\chi^{\leftarrow}(u)]} \int_{]-\infty, t_\chi^{\leftarrow}(v)]} \frac{1}{2\pi(1-\theta^2)^{1/2}} \left(1 + \frac{x^2 - 2\theta xy + y^2}{\chi(1-\rho^2)}\right)^{-(\chi+2)/2} dydx, \tag{1.20}$$

where $\chi > 0$ is the number of degrees of freedom, $\theta \in [-1,1]$ is the linear correlation coefficient, t_χ is the distribution function of a t-distribution with χ degrees of freedom. According to [Demarta and McNeil, 2005] the bivariate t copula $C_{\chi,\theta}$ is attracted to the t extreme value copula. Hence, we simulate $X_{1j}, \dots, X_{Nj}$, $j \in \{1, \dots, n\}$, a block of N variables from a t copula and we take the maximum in this block. This step is repeated several times in order to form a sample $(\bigvee_{i=1}^N X_{i1}, \dots, \bigvee_{i=1}^N X_{in})$ of length n.
The result depicts on Figure 6 is what we waited for. As we increase the number of observations in block maxima, the empirical variance is well more fitted towards the asymptotic variance for a t-EV copula. Furthermore, as the sample's length increase, more the empirical variance fits the theoretical one.

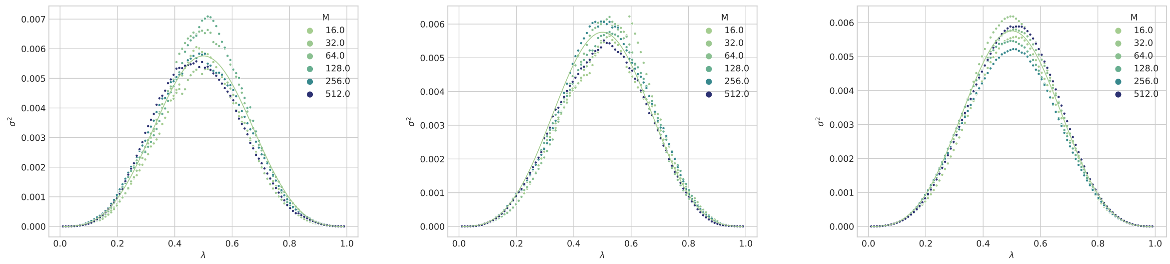


Figure 6: Simulation results obtained by generating $T \in \{128, 256, 512\}$ blocks maximas of length $M \in \{16, 32, 64, 128, 256, 512\}$ from t-copula. For each $\lambda \in \{i/100, i = 0.01, \dots, 0.99,$ we compute the empirical variance $(\times\ T)$ on 100 estimator of λ-FMadogram. The solid line is the theoretical asymptotic variance of t-EV copula.

5 Technical section

5.1 Study of the Pickands dependence function

Lemma 1. *Using the properties of the Pickands dependence function, we have that*

$$0 \leq \kappa(\lambda, A) \leq 1, \quad 0 \leq \zeta(\lambda, A) \leq 1, \quad 0 < u, v < 1.$$

Furthermore, if A admits a second derivative, $\kappa(\cdot, A)$ (resp. $\zeta(\cdot, A)$) is a decreasing function (resp. an increasing function).

Proof First, using that the graph of a (differentiable) convex function lies above all of its tangents and using that $A(t) \geq t$ gives, for $0 < t < 1$:

$$A'(t) \leq \frac{A(1) - A(t)}{1 - t} = \frac{1 - A(t)}{1 - t} \leq 1.$$

Same reasoning using $A(t) \geq 1 - t$ leads to:

$$A'(t) \geq \frac{A(t) - A(0)}{t - 0} = \frac{A(t) - 1}{t} \geq -1.$$

Let's fall back to κ and ζ. If we suppose that A admits a second derivative, the derivative of κ (resp ζ) with respect to λ gives:

$$\kappa'(\lambda, A) = -\lambda A''(\lambda) < 0, \quad \zeta'(\lambda, A) = (1 - \lambda) A''(\lambda) > 0, \quad \forall \lambda \in [0, 1].$$

Using $\kappa(0) = 1$, $\kappa(1) = 1 - A'(1) \geq 0$ gives $0 \leq \kappa(\lambda, A) \leq 1$. As $\zeta(0) = 1 + A'(0) \geq 0$ and $\zeta(1) = 1$, we have $0 \leq \zeta(\lambda, A) \leq 1$. That is the statement.

Now, we can obtain the same result while removing the hypothesis of A admits a second derivative. As A is a convex function, for $x, y \in [0, 1]$, we may have the following inequality:

$$A(x) \geq A(y) + A'(y)(x-y).$$

Take $x = 0$ and $y = \lambda$ gives $1 \geq A(\lambda) - \lambda A'(\lambda) = \kappa(\lambda)$. Now, using that $-\lambda A'(\lambda) \geq -\lambda$, clearly $A(\lambda) - \lambda A'(\lambda) \geq A(\lambda) - \lambda \geq 0$. As $A(\lambda) \geq max(\lambda, 1-\lambda)$. We thus obtain our statement.

Lemma 2. *If A admits a derivative, then $\lim\limits_{t\to 0^+} A'(t)$ and $\lim\limits_{t\to 1^-} A'(t)$ exists and are finite.*

Proof As A is convex and derivable, it follows that $A'(\cdot)$ is increasing. Furthermore, in the proof of Lemma 1, we showed that $-1 \leq A'(t) \leq 1$ for every $t \in (0,1)$ and therefore bounded. Then the two limits exist and are finite.
In the following of the section, we will refer to $A'(0)$ (resp. $A'(1)$) as the left limit of A at 0 (resp. as the right limit of A at 1).

5.2 Proof of Proposition 1

The first statement results directly from *(iii)*. The estimator $\hat{\nu}(\lambda)$ is strongly consistent since it holds

$$\begin{aligned}
&\left| \frac{1}{2T} \sum_{t=1}^{T} \left| \hat{F}_T^{\lambda}(X_t) - \hat{G}_T^{1-\lambda}(Y_t) \right| - \frac{1}{2}\mathbb{E} \left| F^{\lambda}(X) - G^{1-\lambda}(Y) \right| \right| \\
&\leq \left| \frac{1}{2T} \sum_{t=1}^{T} \left| \hat{F}_T^{\lambda}(X_t) - \hat{G}_T^{1-\lambda}(Y_t) \right| - \frac{1}{2T} \sum_{t=1}^{T} \left| F^{\lambda}(X_t) - G^{1-\lambda}(Y_t) \right| \right| \\
&+ \left| \frac{1}{2T} \sum_{t=1}^{T} \left| F^{\lambda}(X_t) - G^{1-\lambda}(Y_t) \right| - \frac{1}{2}\mathbb{E} \left| F^{\lambda}(X) - G^{1-\lambda}(Y) \right| \right|.
\end{aligned}$$

The second term converges almost surely to zero by the strong Law of Large Numbers. For the first term, we have

$$\begin{aligned}
&\left| \frac{1}{2T} \sum_{t=1}^{T} \left| \hat{F}_T^{\lambda}(X_t) - \hat{G}_T^{1-\lambda}(Y_t) \right| - \frac{1}{2T} \sum_{t=1}^{T} \left| F^{\lambda}(X_t) - G^{1-\lambda}(Y_t) \right| \right| \\
&\leq \frac{1}{2T} \sum_{t=1}^{T} \left| \left| \hat{F}_T^{\lambda}(X_t) - \hat{G}_T^{1-\lambda}(Y_t) \right| - \left| F^{\lambda}(X_t) - G^{1-\lambda}(Y_t) \right| \right|, \\
&\leq \frac{1}{2T} \sum_{t=1}^{T} \left| \hat{F}_T^{\lambda}(X_t) - F^{\lambda}(X_t) - \left(\hat{G}_T^{1-\lambda}(Y_t) - G^{1-\lambda}(Y_t) \right) \right|, \\
&\leq \frac{1}{2T} \sum_{t=1}^{T} \left| \hat{F}_T^{\lambda}(X_t) - F^{\lambda}(X_t) \right| + \left| \hat{G}_T^{1-\lambda}(Y_t) - G^{1-\lambda}(Y_t) \right|.
\end{aligned}$$

which converges almost surely to zero according to the strong Law of Large Numbers. To show *(iii)* we define the following function,

$$\nu_\lambda\colon [0,1]^2 \to [0,1], \quad (u,v) \mapsto u^\lambda \vee v^{1-\lambda} - \frac{1}{2}(u^\lambda + v^{1-\lambda}).$$

Using Lemma A.1 in Appendix and the equality $|u^\lambda - v^{1-\lambda}| = u^\lambda \vee v^{1-\lambda} - 2^{-1}(u^\lambda + v^{1-\lambda})$ $= \nu_\lambda(u,v)$ gives,

$$\begin{aligned}
\nu(\lambda) &= \frac{1}{2}\left(\int_{[0,1]} C(x^{\frac{1}{\lambda}},1)dx + \int_{[0,1]} C(1,x^{\frac{1}{1-\lambda}})dx\right) - \int_{[0,1]} C(x^{\frac{1}{\lambda}},x^{\frac{1}{1-\lambda}})dx, \\
&= \frac{1}{2}\left(\int_{[0,1]} x^{\frac{1}{\lambda}}dx + \int_{[0,1]} x^{\frac{1}{1-\lambda}}dx\right) - \int_{[0,1]} x^{\frac{A(\lambda)}{\lambda(1-\lambda)}}dx, \\
&= \frac{1}{2}\left(\frac{\lambda}{1+\lambda} + \frac{1-\lambda}{1+1-\lambda}\right) - \frac{\lambda(1-\lambda)}{A(\lambda)+\lambda(1-\lambda)}, \\
&= \frac{A(\lambda)}{A(\lambda)+\lambda(1-\lambda)} - \frac{1}{2}\left(\frac{1}{1+\lambda} + \frac{1}{1+1-\lambda}\right)
\end{aligned}$$

That is our statement.

□

5.3 A Lemma for Equation (1.13)

Lemma 3. *For all bounded-measurable function $f : [0,1]^2 \mapsto \mathbb{R}$, if $J(s,t) = 2^{-1}|s^\lambda - t^{1-\lambda}|$, then the following integral satisfies:*

$$\int_{[0,1]^2} f(x,y)dJ(x,y) = \frac{1}{2}\int_{[0,1]} f(0,y^{1/(1-\lambda)})dy + \frac{1}{2}\int_{[0,1]} f(x^{1/\lambda},0)dx - \int_{[0,1]} f(x^{1/\lambda},x^{1/(1-\lambda)})dx.$$

Proof Let $A = [0,s] \times [0,t]$, a closed pavement of $[0,1]^2$, where $s,t \in [0,1]$. Thus, $A \in \mathcal{B}([0,1])^2$. Let us introduce the following indicator function :

$$f_{s,t}(x,y) = \mathbb{1}_{\{(x,y)\in[0,1]^2, 0\le x\le s, 0\le y\le t\}}.$$

Then, for this function, we have in one hand :

$$\int_{[0,1]^2} f_{s,t}(x,y)dJ(x,y) = J(s,t) - J(0,0) = \frac{1}{2}|s^\lambda - t^{1-\lambda}|,$$

in other hand, using the equality $2^{-1}|x-y| = 2^{-1}(x+y) - x\wedge y$, one has to show

$$\begin{aligned}
\frac{1}{2}\left|s^\lambda - t^{1-\lambda}\right| &= \frac{s^\lambda}{2} + \frac{t^{1-\lambda}}{2} - s^\lambda \wedge t^{1-\lambda} \\
&= \frac{1}{2}\int_{[0,1]} f_{s,t}(x^{\frac{1}{\lambda}},0)dx + \frac{1}{2}\int_{[0,1]} f_{s,t}(0,y^{\frac{1}{1-\lambda}})dy - \int_{[0,1]} f_{s,t}(x^{\frac{1}{\lambda}},x^{\frac{1}{1-\lambda}})dx.
\end{aligned}$$

Notice that the class

$$\mathcal{E} = \Big\{ A \in \mathcal{B}([0,1]^2) : \int_{[0,1]^2} \mathbb{1}_A(x,y) dJ(x,y) = \int_{[0,1]} \mathbb{1}_A(x^{\frac{1}{\lambda}}, 0) dx + \int_{[0,1]} \mathbb{1}_A(0, y^{\frac{1}{1-\lambda}}) dy - \int_{[0,1]} \mathbb{1}_A(x^{\frac{1}{\lambda}}, x^{\frac{1}{1-\lambda}}) dx \Big\},$$

contains the class $\mathcal{P}$ of all closed pavements of $[0,1]^2$. It is otherwise a monotone class (or λ-system). Hence as the class $\mathcal{P}$ of closed pavement is a π-system, the class monotone theorem ensure that $\mathcal{E}$ contains the sigma-field generated by $\mathcal{P}$, that is $\mathcal{B}([0,1]^2)$.

This result holds for simple function $f(x,y) = \sum_{i=1}^n \lambda_i \mathbb{1}_{A_i}$ where $\lambda_i \in \mathbb{R}$ and $A_i \in \mathcal{B}([0,1]^2)$ for all $i \in \{1, \dots, n\}$. We then can prove the identity for positive measurable function by approximation with an increasing sequence of simple function using Beppo-Levy theorem. This identity is then extended to measurable function $f : [0,1]^2 \mapsto \mathbb{R}$ considering $f = f_+ - f_-$ with $f_+ = max(f, 0)$ and $f_- = min(-f, 0)$. We take the function bounded-measurable in order that the left hand size of the equality is well defined as a Lebesgue-Stieljes integral.

5.4 Proof of Theorem 1

This result comes from the proof of the Theorem 2.4 of [Marcon et al., 2017]. Following their approch, we express the empirical λ-FMadogram $\hat{\nu}_T(\lambda)$ in terms of the empirical copula and exploiting known results. Let us note :

$$\hat{\nu}_T(\lambda) = \frac{1}{T} \sum_{t=1}^T \nu_\lambda(\hat{F}_T(X_t), \hat{G}_T(Y_t)) = \int_{[0,1]^2} \nu_\lambda(u,v) d\hat{C}_T(u,v),$$

$$\nu(\lambda) = \int_{[0,1]^2} \nu_\lambda(u,v) dC(u,v).$$

Using Lemma A.1 in Appendix, we obtain :

$$\sqrt{T}(\hat{\nu}_T(\lambda) - \nu(\lambda)) = \frac{1}{2} \left(\int_{[0,1]} \mathbb{C}_T(x^{\frac{1}{\lambda}}, 1) dx + \int_{[0,1]} \mathbb{C}_T(1, x^{\frac{1}{1-\lambda}}) dx \right) - \int_{[0,1]} \mathbb{C}_T(x^{\frac{1}{\lambda}}, x^{\frac{1}{1-\lambda}}) dx.$$

Consider the function ϕ: $l^\infty([0,1]^2) \to l^\infty([0,1]]), f \mapsto \phi(f)$, defined by

$$(\phi(f))(\lambda) = \frac{1}{2} \left(\int_{[0,1]} f(x^{\frac{1}{\lambda}}, 1) dx + \int_{[0,1]} f(1, x^{\frac{1}{1-\lambda}}) dx \right) - \int_{[0,1]} f(x^{\frac{1}{\lambda}}, x^{\frac{1}{1-\lambda}}) dx.$$

This function is linear and bounded thus continuous. The continous mapping theorem (Theorem 1.3.6 of [van der Vaart and Wellner, 1996]) implies, as $T \to \infty$

$$\sqrt{T}(\hat{\nu}_T - \nu) = \phi(\mathbb{C}_T) \rightsquigarrow \phi(N_C),$$

in $l^\infty([0,1])$. Note that $N_C(u,1) = N_C(1,v) = 0$ almost surely for every $(u,v) \in [0,1]^2$. Indeed for the second one,

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

and $\partial C / \partial v(1,v)$ is well defined according to Lemma 2 and is equal to 1 almost surely. Then, we have

$$\sqrt{T}(\hat{\nu}_T - \nu) \rightsquigarrow \left(- \int_{[0,1]} N_C(x^{\frac{1}{\lambda}}, x^{\frac{1}{1-\lambda}}) dx \right).$$

That is our statement.

5.5 Asymptotic Behavior of the normalized errors of Equation (1.10)

It is readily verified that

$$\sqrt{T}\left(\hat{\nu}_T^*(\lambda) - \nu(\lambda)\right) = \\ \frac{1}{2}\left((1-\lambda) \int_{[0,1]} \mathbb{C}_T(x^{\frac{1}{\lambda}}, 1) dx + \lambda \int_{[0,1]} \mathbb{C}_T(1, x^{\frac{1}{1-\lambda}}) dx \right) - \int_{[0,1]} \mathbb{C}_T(x^{\frac{1}{\lambda}}, x^{\frac{1}{1-\lambda}}) dx.$$

Using the same argument as in the proof of Theorem 2, we can show that $(1-\lambda)\int_{[0,1]} \mathbb{C}_T(x^{\frac{1}{\lambda}}, 1) dx \rightsquigarrow \delta_{\{0\}}$ and $\lambda \int_{[0,1]} \mathbb{C}_T(1, x^{\frac{1}{1-\lambda}}) dx \rightsquigarrow \delta_{\{0\}}$, where $\delta_{\{0\}}$ refers to the Dirac measure at 0. We thus obtain that, by Slutsky's lemma

$$\sqrt{T}\left(\hat{\nu}_T^*(\lambda) - \nu(\lambda)\right) \rightsquigarrow - \int_{[0,1]} N_C(x^{\frac{1}{\lambda}}, x^{\frac{1}{1-\lambda}}) dx.$$

That's what we wanted to prove.

5.6 Proof of Theorem 2

We are able to compute the variance for each process and they are given by the following expressions :

$$Var\left(\int_{[0,1]} B_C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}}) du \right) = f(\lambda, A) \left(\frac{A(\lambda)}{A(\lambda) + 2\lambda(1-\lambda)} \right) = f(\lambda, A)\gamma_1^2,$$

$$Var\left(\int_{[0,1]} B_C(u^{\frac{1}{\lambda}}, 1) \frac{\partial C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}})}{\partial u} du \right) = f(\lambda, A) \left(\frac{\kappa^2(\lambda, A)(1-\lambda)}{2A(\lambda) - (1-\lambda) + 2\lambda(1-\lambda)} \right) = f(\lambda, A)\gamma_2^2,$$

$$Var\left(\int_{[0,1]} B_C(u^{\frac{1}{\lambda}}, 1) \frac{\partial C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}})}{\partial v} du \right) = f(\lambda, A) \left(\frac{\zeta^2(\lambda, A)\lambda}{2A(\lambda) - \lambda + 2\lambda(1-\lambda)} \right) = f(\lambda, A)\gamma_3^2.$$

We now compute the covariance :

$$\gamma_{12} := cov\left(\int_{[0,1]} B_C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}})du, \int_{[0,1]} B_C(v^{\frac{1}{\lambda}}, 1)\frac{\partial C(v^{\frac{1}{\lambda}}, v^{\frac{1}{1-\lambda}})}{\partial u}dv\right)$$
$$= \int_{[0,1]}\int_{[0,1]} \mathbb{E}[B_C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}})B_C(v^{\frac{1}{\lambda}}, 1)]\frac{\partial C(v^{\frac{1}{\lambda}}, v^{\frac{1}{1-\lambda}})}{\partial u}dudv,$$
$$= \int_{[0,1]}\int_{[0,v]} \mathbb{E}[B_C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}})B_C(u^{\frac{1}{\lambda}}, 1)]\frac{\partial C(v^{\frac{1}{\lambda}}, v^{\frac{1}{1-\lambda}})}{\partial u}dudv$$
$$+ \int_{[0,1]}\int_{[v,1]} \mathbb{E}[B_C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}})B_C(u^{\frac{1}{\lambda}}, 1)]\frac{\partial C(v^{\frac{1}{\lambda}}, v^{\frac{1}{1-\lambda}})}{\partial u}dudv.$$

For the first one, we have :

$$\int_{[0,1]}\int_{[0,v]} (C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}}) - C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}})v^{\frac{1}{\lambda}})\frac{\partial C(v^{\frac{1}{\lambda}}, v^{\frac{1}{1-\lambda}})}{\partial u}dudv =$$
$$\frac{\kappa(\lambda, A)}{2}f(\lambda, A)\left(\frac{1-\lambda}{2A(\lambda) + (2\lambda - 1)(1-\lambda)}\right).$$

For the second part, using Fubini, we have :

$$\int_{[0,1]}\int_{[0,u]} (C(v^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}}) - C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}})v^{\frac{1}{\lambda}})\frac{\partial C(v^{\frac{1}{\lambda}}, v^{\frac{1}{1-\lambda}})}{\partial u}dvdu.$$

For the right hand side of the minus sign, we may compute :

$$\int_{[0,1]}\int_{[0,u]} C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}})v^{\frac{1}{\lambda}}\frac{\partial C(v^{\frac{1}{\lambda}}, v^{\frac{1}{1-\lambda}})}{\partial u}dvdu = \frac{\kappa(\lambda, A)}{2}f(\lambda, A).$$

For the last one, some substitutions may be considered.

$$\int_{[0,1]}\int_{[0,u]} C(v^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}})\frac{\partial C(v^{\frac{1}{\lambda}}, v^{\frac{1}{1-\lambda}})}{\partial u}dvdu. \tag{1.21}$$

Following the proof of Proposition 3.3 from [Genest and Segers, 2009], the substitution $v^{\frac{1}{\lambda}} = x$ and $u^{\frac{1}{1-\lambda}} = y$ yields

$$\int_{[0,1]}\int_{[0,u]} C(v^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}})\frac{\partial C(v^{\frac{1}{\lambda}}, v^{\frac{1}{1-\lambda}})}{\partial u}dvdu$$
$$= \lambda(1-\lambda)\int_{[0,1]}\int_{[0,y^{\frac{1-\lambda}{\lambda}}]} C(x,y)\frac{\partial C(x, x^{\frac{\lambda}{1-\lambda}})}{\partial u}x^{\lambda-1}y^{-\lambda}dxdy$$
$$= \lambda(1-\lambda)\kappa(\lambda, A)\int_{[0,1]}\int_{[0,y^{\frac{1-\lambda}{\lambda}}]} C(x,y)x^{\frac{A(\lambda)}{1-\lambda}-(1-\lambda)-1}y^{-\lambda}dxdy.$$

Next, use the substitution $x = w^{1-s}$ and $y = w^s$. Note that $w = xy \in [0,1]$, $s = log(y)/log(xy) \in [0,1]$, $C(x,y) = w^{A(s)}$ and the Jacobian of the transformation is $-log(w)$. As the constraint $x < y^{-1+1/\lambda}$ reduces to $s < \lambda$, the integral becomes:

$$
\begin{aligned}
&-\lambda(1-\lambda)\kappa(\lambda,A)\int_{[0,\lambda]}\int_{[0,1]} w^{A(s)+(1-s)(A_2(\lambda)-(1-\lambda)-1)-s\lambda} log(w)dwds \\
&= \lambda(1-\lambda)\kappa(\lambda,A)\int_{[0,\lambda]} \left[A(s)+(1-s)(A_2(\lambda)-1-(1-\lambda))-s\lambda+1\right]^{-2} ds.
\end{aligned}
$$

Let's continue with computing the following integral :

$$
\begin{aligned}
\gamma_{13} &:= cov\left(\int_{[0,1]} B_C(u^{\frac{1}{\lambda}},1)\frac{\partial C(u^{\frac{1}{\lambda}},u^{\frac{1}{1-\lambda}})}{\partial u}du, \int_{[0,1]} B_C(1,v^{\frac{1}{1-\lambda}})\frac{\partial C(v^{\frac{1}{\lambda}},v^{\frac{1}{1-\lambda}})}{\partial v}dv\right) \\
&= \mathbb{E}\left[\int_{[0,1]}\int_{[0,1]} B_C(u^{\frac{1}{\lambda}},1)B_C(1,v^{\frac{1}{1-\lambda}})\frac{\partial C(u^{\frac{1}{\lambda}},u^{\frac{1}{1-\lambda}})}{\partial u}\frac{\partial C(v^{\frac{1}{\lambda}},v^{\frac{1}{1-\lambda}})}{\partial v}dudv\right] \\
&= \int_{[0,1]}\int_{[0,1]} \left(C(u^{\frac{1}{\lambda}},v^{\frac{1}{1-\lambda}}) - u^{\frac{1}{\lambda}}v^{\frac{1}{1-\lambda}}\right)\frac{\partial C(u^{\frac{1}{\lambda}},u^{\frac{1}{1-\lambda}})}{\partial u}\frac{\partial C(v^{\frac{1}{\lambda}},v^{\frac{1}{1-\lambda}})}{\partial v}dudv.
\end{aligned}
$$

The second term can be easily handled and its value is given by :

$$
\int_{[0,1]}\int_{[0,1]} u^{\frac{1}{\lambda}}v^{\frac{1}{1-\lambda}}\frac{\partial C(u^{\frac{1}{\lambda}},u^{\frac{1}{1-\lambda}})}{\partial u}\frac{\partial C(v^{\frac{1}{\lambda}},v^{\frac{1}{1-\lambda}})}{\partial v}dudv = f(\lambda,A)\kappa(\lambda,A)\zeta(\lambda,A).
$$

For the first, use the substitutions $u^{\frac{1}{\lambda}} = x$ and $v^{\frac{1}{1-\lambda}} = y$. This yields :

$$
\lambda(1-\lambda)\int_{[0,1]}\int_{[0,1]} C(x,y)\frac{\partial C(x,x^{\frac{\lambda}{1-\lambda}})}{\partial u}\frac{\partial C(y^{\frac{1-\lambda}{\lambda}},y)}{\partial v}x^{\lambda-1}y^{-\lambda}dxdy.
$$

Then, make the substitutions $x = w^{1-s}$, $y = w^s$ that were used for the preceding integral gives :

$$
\begin{aligned}
&-\lambda(1-\lambda)\kappa(\lambda,A)\zeta(\lambda,A)\int_{[0,1]}\int_{[0,1]} w^{A(s)+(1-s)(A_2(\lambda)-(1-\lambda)-1)+s(A_1(\lambda)-\lambda-1)} log(w)dwds \\
&= \lambda(1-\lambda)\kappa(\lambda,A)\zeta(\lambda,A)\int_{[0,1]} \left[A(s)+(1-s)(A_2(\lambda)-(1-\lambda)-1)+s(A_1(\lambda)-\lambda-1)+1\right]^{-2} ds.
\end{aligned}
$$

Similarly, the last covariance requires the same tools as used before, it is left to the reader. It then suffices to use the bilinearity of the covariance and to assemble the various terms to conclude.

5.7 Proof of Corollary 1

In independent case, we have $A(t) = 1$ for every $t \in [0,1]$, then $\kappa(\lambda, 1) = 1$ and $\zeta(\lambda, 1) = 1$ for each $\lambda \in [0,1]$. Then, Equation (1.21) equals :

$$f(\lambda, 1)\left(\frac{1+\lambda(1-\lambda)}{2-(1-\lambda)+2\lambda(1-\lambda)}\right).$$

Then, summing all the elements gives

$$cov\left(\int_{[0,1]} B_C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}})du, \int_{[0,1]} B_C(u^{\frac{1}{\lambda}}, 1)\frac{\partial C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}})}{\partial u}du\right) = f(\lambda, 1)\left(\frac{1-\lambda}{2-(1-\lambda)+2\lambda(1-\lambda)}\right).$$

Direct computations gives

$$cov\left(\int_{[0,1]} B_C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}})du, \int_{[0,1]} B_C(1, u^{\frac{1}{1-\lambda}})\frac{\partial C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}})}{\partial v}du\right) = f(\lambda, 1)\left(\frac{\lambda}{2-\lambda+2\lambda(1-\lambda)}\right).$$

In independence case, we have the following equality :

$$cov\left(\int_{[0,1]} B_C(u^{\frac{1}{\lambda}}, 1)\frac{\partial C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}})}{\partial u}du, \int_{[0,1]} B_C(1, u^{\frac{1}{1-\lambda}})\frac{\partial C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}})}{\partial v}du\right) = 0.$$

We thus obtain the equality.

5.8 A counter example against of variance's monotony with respect to positive dependence

First, notive that, under dependency condition, the variance of the λ-FMadogram evaluated in $\lambda = 0.5$ is equal to $1/150$.

Lemma 4. *Let us consider $A(t) = 1 - \theta t + \theta t^2$ where $\theta \in [0,1]$. If we take $\lambda = 0.5$, there exist $\theta \in (0,1)$ such that*

$$Var\left(\int_{[0,1]} N_C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}})du\right) > \frac{1}{150}. \tag{1.22}$$

Proof For this dependence function, we have immediately :

$$\kappa(\lambda, A) = 1 - \theta\lambda^2, \quad \zeta(\lambda, A) = 1 - \theta(1-\lambda)^2.$$

For $\lambda = 0.5$, we notice that $\kappa(0.5, A) = \zeta(0.5, A)$. By a simple change of variable, we notice that :

$$\int_0^{0.5}[A(s)+(1-s)(2A(0.5)-0.5-1)-0.5s+1]^{-2}ds = \int_{0.5}^1[A(s)+s(2A(0.5)-0.5-1)-0.5(1-s)+1]^{-2}ds.$$

By substitution, we have for the chosen copula that,

$$\int_{0.5}^1[A(s)+s(2A(0.5)-0.5-1)-0.5(1-s)+1]^{-2}ds = \int_{0.5}^1[\frac{3}{2}-s(\theta+1-2A(0.5))+s^2\theta]ds.$$

Let us take $\theta = 2A(0.5)-1$, which implies by direct computation that $\theta = 2/3 > 0$. Let us make of use of this lemma :

Lemma 5. *Let a, b be two reals. Note $I_n = \int_{\mathbb{R}}(ax^2+b)^n dx$, then :*

$$I_n = \frac{2n-3}{2b(n-1)}I_{n-1} + \frac{x}{2b(n-1)(ax^2+b)}.$$

Proof An integration by parts gives and some algebra gives:

$$\begin{aligned} I_{n-1} &= \frac{x}{(ax^2+b)^{n-1}} + 2(n-1)\int_{\mathbb{R}}\frac{ax^2}{(ax^2+b)^n}dx, \\ &= \frac{x}{(ax^2+b)^{n-1}} + 2(n-1)I_{n-1} - 2b(n-1)I_n. \end{aligned}$$

Solving the equation for I_n gives the result.

We want to compute the following quantity :

$$\int_{0.5}^1[\frac{3}{2}+s^2\frac{2}{3}]^{-2}ds.$$

The lemma gives :

$$\begin{aligned} \int_{0.5}^1[\frac{3}{2}+s^2\frac{2}{3}]^{-2}ds &= 36\int_{0.5}^1[4s^2+9]^{-2}ds \\ &= 2\left(\frac{7}{20}+\int_{0.5}^1(4s^2+9)^{-1}ds\right), \\ &= 2\left(\frac{7}{20}+\frac{1}{6}\int_{1/3}^{2/3}\frac{1}{u^2+1}du\right). \end{aligned}$$

Where we have made the substitution $u = 2s/3$ in the third line. Then :

$$\int_{0.5}^1[\frac{3}{2}+s^2\frac{2}{3}]^{-2}ds = 2\left[\frac{7}{20}+\frac{1}{6}(atan(2/3)-atan(1/3))\right] \approx 0.142596.$$

For the last integral, we have, by substitution for $\lambda = 0.5$ and $\theta = 2/3$:

$$\int_0^1 [A(s)+(1-s)(2A(0.5)-0.5-1)+s(2A(0.5)-0.5-1)+1]^{-2}ds = \int_0^1 [\frac{13}{6}-\frac{2}{3}s+\frac{2}{3}s^2]^{-2}ds.$$

Then, we are able to compute :

$$\begin{aligned}\int_0^1 [\frac{13}{6}-\frac{2}{3}s+\frac{2}{3}s^2]^{-2}ds = 36\int_0^1 (13-4s+4s^2)ds &\overset{u=(2s-1)}{=} 36\int_0^1 ((2s-1)^2+12)^2 ds,\\ = 18\int_{-1}^1 (u^2+12)^{-2}du &\overset{\text{Lemma}}{=} \frac{3}{4}\left(\frac{2}{13}+\int_{-1}^1 \frac{1}{u^2+12}du\right),\\ &\overset{v=u/(2\sqrt{3})}{=} \frac{6}{52}+\frac{3}{8\sqrt{3}}\int_{\frac{-1}{2\sqrt{3}}}^{\frac{1}{2\sqrt{3}}} \frac{1}{v^2+1}dv.\end{aligned}$$

$$\int_0^1 [\frac{13}{6}-\frac{2}{3}s+\frac{2}{3}s^2]^{-2}ds = \frac{\sqrt{3}}{8}\left(atan(\frac{1}{2\sqrt{3}})-atan(-\frac{1}{2\sqrt{3}})\right)+\frac{6}{52} \approx 0.23707.$$

Summing all the components of the variance gives $Var\left(\int_{[0,1]} N_C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}})du\right) \approx 0.00713 > 1/150$, which gives our counterexample.

Chapter 2

Madogram estimator with missing data

1 Introduction

As the volume of data expands, the problem of missing or contaminated data has been increasingly present in many fields of statistical applications. It frequently happens that all of the individuals of a sample of statistical data from a multivariate population are not observed. If a sample be represented in matrix form by allowing the rows to represend the individuals and the columns the variates, then the matrix of the type of sample with which we are concerned is incomplete in that some elements are not present. In dealing with fragmentary samples, it is important to have at hand techniques which will enable the statistician to extract as much information as possible from the data. This is especially true if the data are unique or expensive. A useful reference for general parametric statistical inferences with missing data was provided by [Little R.J.A., 1987]. A fundamental step of these studies begins with considering a random sample of incomplete data,

$$(X_t, Y_t, \delta_t), \quad t \in \{1, \dots, T\}, \tag{2.1}$$

where all the X_t's are observed and $\delta_t = 0$ if Y_t is missing, otherwise $\delta_t = 1$. The simple missing data pattern describe by (2.1) is basically created by the double sampling or two phase sampling (see chapter 12 of [Cochran, 2007]). The purpose of this sample is to furnish a good knowledge of the distribution of X. In a survey whose function is to make estimate for some other variate Y, it may pay to devote part of the ressources to this preliminary sample, although this means that the size of the sample in the main survey on Y may diminish.

Data (2.1) also arise in survival analysis, the study of the duration time preceding an event of interest is considered with series of random censors, which might prevent the capture of the whole survival time. This is known as the censoring mechanism and it arises from restrictions depending from the nature of the study. Typically, they may occur in medicine, with studies of the survival times before the recovery / decease from

a specific disease. Another important example is often realized in comparing treament effects of two educational programs. Individuals with lower scores on a preliminary test are more likely to receive the experimental treatment (*i.e.*, a composatory study program), whereas those with higher preliminary scores are more inclined to take the standard control. This phenomenon is well-known as the selection problem and we refer to Chapter 2 of [Angrist and Pischke, 2008] for more details.
In this chapter, we will consider the following sample

$$(I_t X_t, J_t Y_t, I_t, J_t), \quad t \in \{1, \dots, T\}, \tag{2.2}$$

where $I_t = 0$ (resp. $J_t = 0$) if X_t (resp. Y_t) is missing, otherwise $I_t = 1$ (resp. $J_t = 1$).
Robustness in estimation has known an important research activity developed in the 60's and 70's resulting in a large number of publications. For a summary, the interested reader is refered to [Huber, 2011]. Robustness can be seen as an estimation procedure in which both stochastic and approximation error are low (see Section 1.1 from [Baraud et al., 2016]). In other words, an estimator is said to be robust if our model provides a reasonable approximation of the true one and derive an estimator which remains close to the true distribution. In this Chapter, we mean by *robust* as *robust against outlier*, *e.g* the ϵ-contamination model (see [Huber, 1964]), or *robust again heavy-tailed data* where only low-order moments are assumed to be finite for the data distribution. There is no simple relation between the two definitions and the first framework of robustness that we have depicted. We want to propose a robust estimator of the Madogram. To achieve our goal, we leverage the idea of Median-Of-MeaNs (MoN). Intuitively, we replace the linear operator of expectation with the median of averages taken over non-overlaping blocks of the data, in order to get a robust estimate thanks to the median step (see [Lerasle et al., 2019] for a similar idea applied to Kernel). The MoN is one of the mean estimators that achieve a sub-Gaussian behavior under mild conditions. Introduced during the 1980s [Nemirovsky and Yudin, 1983] for the estimation of the mean of real-valued random variables, that is easy to compute, while exhibiting attractive robustness properties. In our perspective, we only know [Escobar-Bach et al., 2018] that include the contamination framework in their estimation of the Pickands dependence function in the extreme value theory.

2 Notations

Let (X, Y) be a bivariate random vector with joint distribution function $H(x, y)$ and continuous marginal distribution function $F(x)$ and $G(y)$. Its associated copula C is defined by $H(x, y) = C(F(x), G(y))$. Since F and G are continuous, the copula C is unique and we can write $C(u, v) = H(F^{\leftarrow}(u), G^{\leftarrow}(v))$ for $0 \leq u, v \leq 1$ and where $F^{\leftarrow}(u) = inf\{v \in \mathbb{R} | F(v) \geq u\}$ and $G^{\leftarrow}(u) = inf\{v \in \mathbb{R} | G(v) \geq u\}$ are the generalized inverse

functions of F and G respectively. Here, we use the abbreviation $Qf = \int f dQ$ for a given measurable function f and signed measure Q.
We suppose that we observe sequentilly a quadruple defined by Equation (2.2). At each $t \in \{1, \dots, T\}$, one of both entries may be missing. The observations consist of a sample of independent, identically distributed quadruple (I, J, X, Y) of law $\mathbb{P}$. The probability of observing a realisation partially or completely is denoted by $p_X = \mathbb{P}(I_t = 1) > 0$, $p_Y = \mathbb{P}(J_t = 1) > 0$ and $p_{XY} = \mathbb{P}(I_t = 1, J_t = 1) > 0$. Let us now define the empirical cumulative distribution of X (resp. Y and (X, Y)) in case of missing data,

$$\hat{F}_T(u) = \frac{\sum_{t=1}^T 1_{\{X_t \leq u\}} I_t}{\sum_{t=1}^T I_t}, \quad \hat{G}_T(v) = \frac{\sum_{t=1}^T 1_{\{Y_t \leq v\}} J_t}{\sum_{t=1}^T J_t}, \quad \hat{H}_T(u, v) = \frac{\sum_{t=1}^T 1_{\{X_t \leq u, Y_t \leq v\}} I_t J_t}{\sum_{t=1}^T I_t J_t}. \tag{2.3}$$

We have all tools in hand to define the *hybrid copula estomator* introduced by [Segers, 2014],

$$\hat{C}_T^{\mathcal{H}}(u, v) = \hat{H}_T(\hat{F}_T(u), \hat{G}_T(v)). \tag{2.4}$$

Given a rate $r_T > 0$ and $r_T \to \infty$ as $T \to \infty$, the normalized estimation error of the hybrid copula estimator is :

$$\mathbb{C}_T^{\mathcal{H}}(u, v) = r_T \left(\hat{C}_T^{\mathcal{H}}(u, v) - C(u, v) \right). \tag{2.5}$$

Assume that $(X_1, Y_1), \dots, (X_T, Y_T)$ is a sequence of independent random vectors with values in a measurable space $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ that represent the data available to the statistician. We will assume that the sample is partitioned into k disjoint subsets $B_1, \dots, B_K$ of cardinalities $n_j := card(B_j)$ respectively, where the partitioning scheme is independent of the data. Let f be a measurable function from $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ to $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$, we define the following estimator

$$\bar{\mathbb{P}}_{n_j} f = \frac{1}{n_j} \sum_{j \in B_j} f(X_j, Y_j).$$

We define the MoN estimator of f as solutions of the optimization problem

$$\hat{f}_{MoN} = \underset{z \in \mathbb{R}}{argmin} \sum_{j=1}^K \left| \bar{\mathbb{P}}_{n_j} f - z \right|, \tag{2.6}$$

which, if we note $med(\cdot)$ the usual univariate median

$$\hat{f}_{MoN} = med(\bar{\mathbb{P}}_{n_1} f, \dots, \bar{\mathbb{P}}_{n_K} f), \tag{2.7}$$

is a solution of Equation (2.6). We will denote by ν the FMadogram

$$\nu = \frac{1}{2} \mathbb{E}\left[|F(X) - G(Y)| \right]. \tag{2.8}$$

3 Weak convergence of the Madogram with missing data

As in the first chapter, we suppose that our copula function is of the extreme value type. In this study, we aim to analyze the asymptotic variance structure of the λ-FMadogram in case of missing data. Based on these identical and idenpendent copies $(I_1X_1, J_1Y_1, I_1, J_1), \dots, (I_TX_T, J_TY_T, I_T, J_T)$, we define the following estimator of the λ-FMadogram :

$$\hat{\nu}_T^{\mathcal{H}}(\lambda) = \frac{1}{2\sum_{t=1}^T I_tJ_t} \sum_{t=1}^T \left| \hat{F}_T^{\lambda}(X_t) - \hat{G}_T^{1-\lambda}(Y_t) \right| I_tJ_t. \tag{2.9}$$

We also make the following assumptions,

Assumption A. *(i) The bivariate distribution function H has continuous margins F, G and copula C.*

(ii) The derivative of the Pickands dependence function $A'(t)$ exists and is continuous on $(0,1)$.

We also make an assumption concerning the missing mechanism,

Assumption B. *We suppose for all $t \in \{1, \dots, T\}$, the pairs (I_t, J_t) and (X_t, Y_t) are independent, the data are missing completely at random (**MCAR**). Furthermore, we suppose that there exist at least one $t \in \{1, \dots, T\}$ such that $I_tJ_t \neq 0$.*

In the hybrid copula estimator, we need a technical assumption in order to guarantee the weak convergence of the process $\mathcal{C}_T^{\mathcal{H}}$ (see [Segers, 2014]),

Assumption C. *There exists $\gamma_t > 0$ and $r_t > 0$ such that $r_t \longrightarrow \infty$ as $t \to \infty$ such that in the space $l^{\infty}(\mathbb{R}^2) \otimes (l^{\infty}(\mathbb{R}), l^{\infty}(\mathbb{R}))$ equipped with the topology of uniform convergence, we have the joint weak convergence*

$$\left(r_t(\hat{H}_T - H); r_t(\hat{F}_t - F), r_t(\hat{G}_t - G) \right) \rightsquigarrow (\alpha \circ (F,G), \beta_1 \circ F, \beta_2 \circ G).$$

The stochastic processes α and β_j take values in $l^{\infty}([0,1]^2)$ and $l^{\infty}([0,1])$ respectively, and are such that $\alpha \circ F$ and $\beta_j \circ F_j$ have continuous trajectories on $[-\infty, \infty]^2$ and $[-\infty, \infty]$ almost surely.

Under condition A and C (see Theorem B.1 in Appendix), the stochastic process $\mathbb{C}_T^{\mathcal{H}}$ converges weakly to the tight Gaussian process S_C where we denote by $S_C(u,v)$ the process defined by,

$$S_C(u,v) = \alpha(u,v) - \frac{\partial C(u,v)}{\partial u} \beta_1(u) - \frac{\partial C(u,v)}{\partial v} \beta_2(v), \quad \forall (u,v) \in [0,1]^2. \tag{2.10}$$

Considering our statistical framework and missing mechanism, [Segers, 2014] shows (in example 3.5) that the processes α, β_1 and β_2 take the following closed form :

$$\begin{aligned}
\beta_1(u) &= p_X^{-1}\mathbb{G}\left(1_{X\le F^{\leftarrow}(u),I=1} - u1_{I=1}\right),\\
\beta_2(v) &= p_Y^{-1}\mathbb{G}\left(1_{Y\le G^{\leftarrow}(v),J=1} - v1_{J=1}\right),\\
\alpha(u,v) &= p_{XY}^{-1}\mathbb{G}\left(1_{X\le F^{\leftarrow}(u)}1_{Y\le G^{\leftarrow}(v),I=1,J=1} - C(u,v)1_{I=1,J=1}\right).
\end{aligned}$$

Furthermore, we are able to compute their covariance functions. This is summarised in the following lemma and we add some technical details available in last section.

Lemma 6. *The covariance function of the process $\beta_1(u)$, $\beta_2(v)$ and $\alpha(u,v)$ are : for $(u,u_1,u_2,v,v_1,v_2)\in[0,1]^6$,*

$$\begin{aligned}
cov\left(\beta_1(u_1),\beta_1(u_2)\right) &= p_X^{-1}\{u_1\wedge u_2 - u_1u_2\},\\
cov\left(\beta_2(v_1),\beta_2(v_2)\right) &= p_Y^{-1}\{v_1\wedge v_2 - v_1v_2\},\\
cov\left(\beta_1(u),\beta_2(v)\right) &= \frac{p_{XY}}{p_Xp_Y}\{C(u,v)-uv\},
\end{aligned}$$

and

$$\begin{aligned}
cov\left(\alpha(u_1,v_1),\alpha(u_2,v_2)\right) &= p_{XY}^{-1}\{C(u_1\wedge u_2, v_1\wedge v_2) - C(u_1,v_1)C(u_2,v_2)\},\\
cov\left(\alpha(u_1,v),\beta_1(u_2)\right) &= p_X^{-1}\{C(u_1\wedge u_2, v) - C(u_1,v)u_2\},\\
cov\left(\alpha(u,v_1),\beta_2(v_2)\right) &= p_Y^{-1}\{C(u, v_1\wedge v_2) - C(u,v_1)v_2\}.
\end{aligned}$$

Before going further, let us briefly talk about our estimator. As in Chapter 1, our estimator defined in (2.9) does not verify $\hat{\nu}_T^{\mathcal{H}}(0) = \hat{\nu}_T^{\mathcal{H}}(1) = 0.25$. But in addition, the variance of our estimator at $\lambda=0$ or $\lambda=1$ does not equal 0. Indeed, suppose that we evaluate our statistic at $\lambda=0$, we thus obtain the following quantity :

$$\hat{\nu}_T^{\mathcal{H}}(0) = \frac{1}{2\sum_{t=1}^T I_tJ_t}\sum_{t=1}^T\left(1-\hat{G}_T(Y_t)\right)I_tJ_t$$

In this situation, the sample $(X_t)_{t=1}^T$ is taken account through the indicator's sequence $(I_t)_{t=1}^T$ and induce a variance when estimating. Hence we can force our estimator as in [Naveau et al., 2009] to satisfy $\nu_T^{\mathcal{H}}(0) = \nu_T^{\mathcal{H}}(1) = 0.25$. This leads to the following definition :

$$\begin{aligned}
\hat{\nu}_T^{\mathcal{H}*}(\lambda) = &\frac{1}{2\sum_{t=1}^T I_tJ_t}\sum_{t=1}^T\left|\hat{F}_T^{\lambda}(X_t) - \hat{G}_T^{1-\lambda}(Y_t)\right|I_tJ_t - \frac{\lambda}{2\sum_{t=1}^T I_tJ_t}\sum_{t=1}^T\{1-\hat{F}_T^{\lambda}(X_t)\}I_tJ_t\\
&-\frac{1-\lambda}{2\sum_{t=1}^T I_tJ_t}\sum_{t=1}^T\{1-\hat{G}_T^{1-\lambda}(Y_t)\}I_tJ_t + \frac{1}{2}\frac{1-\lambda+\lambda^2}{(2-\lambda)(1+\lambda)} \qquad (2.11)
\end{aligned}$$

Nevertheless, the asymptotic behaviour of $\sqrt{T}\left(\hat{\nu}_T^{\mathcal{H}*}(\lambda)-\nu(\lambda)\right)$ is not the same as $\sqrt{T}\left(\hat{\nu}_T^{\mathcal{H}}(\lambda)-\nu(\lambda)\right)$ and they should be studied apart. This result is given by the theorem as follows,

Theorem 3. *Let $\lambda \in [0,1]$. Under Assumptions A, B, C we have the weak convergence in $l^\infty([0,1])$ for the hybrid estimator defined in (2.9), as $T \to \infty$,*

$$\sqrt{T}\left(\hat{\nu}_T^{\mathcal{H}}(\lambda)-\nu(\lambda)\right) \rightsquigarrow \left(\frac{1}{2}\int_{[0,1]} \alpha(x^{\frac{1}{\lambda}},1)-\beta_1(x^{\frac{1}{\lambda}})dx + \frac{1}{2}\int_{[0,1]} \alpha(1,x^{\frac{1}{1-\lambda}})-\beta_2(x^{\frac{1}{1-\lambda}})dx \right. \tag{2.12}$$
$$\left. - \int_{[0,1]} S_C(x^{\frac{1}{\lambda}}, x^{\frac{1}{1-\lambda}})dx\right)_{\lambda\in[0,1]},$$

$$\sqrt{T}\left(\hat{\nu}_T^{\mathcal{H}*}(\lambda)-\nu(\lambda)\right) \rightsquigarrow \left(\frac{1-\lambda}{2}\int_{[0,1]} \alpha(x^{\frac{1}{\lambda}},1)-\beta_1(x^{\frac{1}{\lambda}})dx + \frac{\lambda}{2}\int_{[0,1]} \alpha(1,x^{\frac{1}{1-\lambda}})-\beta_2(x^{\frac{1}{1-\lambda}})dx \right. \tag{2.13}$$
$$\left. - \int_{[0,1]} S_C(x^{\frac{1}{\lambda}}, x^{\frac{1}{1-\lambda}})dx\right)_{\lambda\in[0,1]}.$$

Proof is deferred at Section 6. Let us briefly discuss the difference between these processes and the process exhibited for the normalized estimation error of the λ-FMadogram without missing data. When $p_{XY} = p_X = p_Y = 1$, the processes $(\alpha(u,1)-\beta_1(u))_{u\in[0,1]}$ and $(\alpha(1,v)-\beta_2(v))_{v\in[0,1]}$ are equal to zero almost surely. We thus go back to our process as described by Theorem 1 in Chapter 1.

As an integral of a tight Gaussian process, we know that the two normalized estimation errors follows a centered Gaussian variable for a given $\lambda \in [0,1]$. We are able to give a closed form of the variance of the limiting Gaussian law. For notational convenience, we introduce the following notations :

$$\sigma_1^2 = \left(p_{XY}^{-1}-p_X^{-1}\right)\left(\frac{\lambda}{1+\lambda}\right)^2\frac{1}{1+2\lambda}, \quad \sigma_2^2 = \left(p_{XY}^{-1}-p_Y^{-1}\right)\left(\frac{1-\lambda}{1+1-\lambda}\right)^2\frac{1}{1+2(1-\lambda)},$$
$$\sigma_3^2 = f(\lambda,A)\left(p_{XY}^{-1}\gamma_1^2+p_X^{-1}\gamma_2^2+p_Y^{-1}\gamma_3^2\right)-2p_X^{-1}\gamma_{12}-2p_Y^{-1}\gamma_{13}+2\frac{p_{XY}}{p_Xp_Y}\gamma_{23}.$$

$$\sigma_{12} = \left(p_{XY}^{-1}-p_X^{-1}-p_Y^{-1}+\frac{p_{XY}}{p_Xp_Y}\right)\lambda(1-\lambda)\left(\int_{[0,1]}\left[A(s)-(1-s)(1-\lambda)-s\lambda+1\right]^{-2}ds - \frac{1}{(1+\lambda)(1+1-\lambda)}\right).$$

$$
\begin{aligned}
\sigma_{13} &= \left(p_{XY}^{-1} - p_X^{-1}\right) \lambda(1-\lambda)\Bigg(\int_{[0,\lambda]} [A(s) - (1-s)(1-\lambda) - s\lambda + 1]^{-2}\, ds \\
&+ \frac{\lambda}{A(\lambda + \lambda(1-\lambda)} \left[\frac{1-\lambda}{A(\lambda) + 2\lambda(1-\lambda)} - \frac{1}{1+\lambda} \right] \Bigg) \\
&- \left(p_Y^{-1} - \frac{p_{XY}}{p_X p_Y} \right) \zeta(\lambda, A)\lambda(1-\lambda)\Bigg(\int_{[0,1]} [A(s) + s(A_1(\lambda) - \lambda - 1) - (1-s)(1-\lambda) + 1]^{-2}\, ds \\
&- \frac{\lambda}{(1+\lambda)(A(\lambda) + \lambda(1-\lambda))} \Bigg).
\end{aligned}
$$

$$
\begin{aligned}
\sigma_{23} &= \left(p_{XY}^{-1} - p_Y^{-1}\right) \lambda(1-\lambda)\Bigg(\int_{[\lambda,1]} [A(s) - (1-s)(1-\lambda) - s\lambda + 1]^{-2}\, ds \\
&+ \frac{1-\lambda}{A(\lambda + \lambda(1-\lambda)} \left[\frac{\lambda}{A(\lambda) + 2\lambda(1-\lambda)} - \frac{1}{1+1-\lambda} \right] \Bigg) \\
&- \left(p_X^{-1} - \frac{p_{XY}}{p_X p_Y} \right) \kappa(\lambda, A)\lambda(1-\lambda)\Bigg(\int_{[0,1]} [A(s) + (1-s)(A_2(\lambda) - (1-\lambda) - 1) - s\lambda + 1]^{-2}\, ds \\
&- \frac{1-\lambda}{(1+1-\lambda)(A(\lambda) + \lambda(1-\lambda))} \Bigg).
\end{aligned}
$$

We thus may write,

Theorem 4. *For $\lambda \in (0,1)$, let $A_1(\lambda) = A(\lambda)/\lambda$, $A_2(\lambda) = A(\lambda)/(1-\lambda)$. Then, the asymptotic variance of $\sqrt{T}\left(\hat{\nu}_T^{\mathcal{H}}(\lambda) - \nu(\lambda)\right)$ and $\sqrt{T}\left(\hat{\nu}_T^{\mathcal{H}*}(\lambda) - \nu(\lambda)\right)$ has the following closed form*

$$
Var\left(\sqrt{T}\left(\hat{\nu}_T^{\mathcal{H}}(\lambda) - \nu(\lambda)\right)\right) = \frac{1}{4}\sigma_1^2 + \frac{1}{4}\sigma_2^2 + \sigma_3^2 + \frac{1}{2}\sigma_{12} - \sigma_{13} - \sigma_{23},
$$

$$
Var\left(\sqrt{T}\left(\hat{\nu}_T^{\mathcal{H}*}(\lambda) - \nu(\lambda)\right)\right) = \frac{(1-\lambda)^2}{4}\sigma_1^2 + \frac{\lambda^2}{4}\sigma_2^2 + \sigma_3^2 + \lambda(1-\lambda)\frac{1}{2}\sigma_{12} - (1-\lambda)\sigma_{13} - \lambda\sigma_{23}.
$$

Note that the variance of the limiting process of the normalized estimation error of $\hat{\nu}_T^{\mathcal{H}*}(\lambda)$ for a given λ is not always lower than that of $\hat{\nu}_T^{\mathcal{H}}(\lambda)$.

4 Concentration properties of the MoN-Madogram

In this section, we try to design a robust estimator of the FMadogram as defined in Equation 2.8. We are restricting our analysis to the FMadogram to avoid technical difficulties but the proof would be similar with a discussion according to the value of λ and using that $||x|^\lambda - |y|^\lambda| \leq |x-y|^\lambda$. Let $B_1, \dots, B_K$ a partition of the set $\{1, \dots, T\}$. Denote by $\bar{F}_{n_j}$ (resp. $\bar{G}_{n_j}$) the empirical cumulative distribution for the cumulative distribution of

X (resp. Y) computed within block B_j. We propose the following MoN-based madogram estimator

$$\hat{\nu}_{MoN} = \underset{z \in \mathbb{R}}{argmin} \sum_{j=1}^{K} |\bar{\nu}_j - z| , \tag{2.14}$$

where $\bar{\nu}_{n_j} = \frac{1}{2n_j} \sum_{t \in B_j} |\bar{F}_{n_j}(X_t) - \bar{G}_{n_j}(Y_t)|$. That is, in Equation (2.6), we take $f(x, y) = J(x, y) = |x - y|$ and $\bar{\mathbb{P}}_{n_j} = \bar{C}_{n_j}$ the empirical copula constructed on the block B_j.

Assumption 1. *The sample $((X_1, Y_1), \ldots, (X_T, Y_t))$ contains $T - T_o$ outliers drawn according to distribution H, and T_o outliers, upon which no assumption is made.*

In presence of outliers, the key point is to focus on sane blocks, *i.e* on blocks that does not contains a single outliers, since no inference can be made about blocks hit by an outlier. One way to ensure that sane blocks to be in majority is to consider twice more blocks than outliers. Indeed, in the worst case scenario, each outlier contaminate on block, but the sane blocks remains more numerous. Let K_s denote the total number of sane block containing no outliers. In other words, there exists $\delta \in (0, 1/2]$ such that $K_s \geq K(1/2+\delta)$. If the data are free from contaminations, then $K_s = K$ and $\delta = 1/2$.
We suppose without loss of generality that $n_j = \lceil T/K \rceil$ for every $j \in \{1, \ldots, K\}$. Using these notations, we can prove the following deviation bounds for our MoN-based estimator.

Theorem 5. *(Consistency & outlier-robustness of $\hat{\nu}$). Under Assumption 1, for any $\eta \in]0, 1[$ such that $K = \delta^{-1} log(1/\eta)$ it holds that with probability $1 - \eta$,*

$$|\hat{\nu}_{MoN} - \nu| \leq \frac{3}{\sqrt{2}} \frac{log\left(6e2^{\frac{1}{\delta}}\right)}{\delta} \sqrt{\frac{log\,(1/\eta)}{T}}.$$

Remark 1.
- *Dependence on T : These finite-sample guarantees show that estimator is robust to outliers, providing consistent estimates with high probability even under arbitrary contamination (affecting less than half of the samples).*
- *Dependence on δ : Recall that higher δ corresponds to less outliers,* i.e., *cleaner data in which case the bounds above become tighter.*
- *Dependence on η : An higher η gives a greater bound for which the estimator hold with an greater probability.*

5 Simulation

This section illustrates experimentally the take-home message of this chapter. The code for these experiments is available online [1]. We consider the same six extreme value copula

[1] `https://github.com/Aleboul/var_FMado/tree/main/missing_data`

from Chapter 1. In each experiment, we estimate the empirical variance ($\times T$) on several Monte Carlo simulation for both estimators in a finite sample setting, we then add the theoretical value of the asymptotic variance. For $p_X \in [0,1]$ and $p_Y \in [0,1]$ the missingness on variables X and Y is generated according to a Bernoulli distribution

$$I \sim \mathcal{B}(p_X), \quad J \sim \mathcal{B}(p_Y).$$

We also set that $p_{XY} = p_X p_Y$,*i.e.* I and J are independent.

5.1 Results for fixed sample sizes

Figure 1 presents the results for the six model defined in Chapter 1. Here, we take $p_X = p_Y = 0.75$. As waited, we directly see that both empirical and theoretical values of the variance of the normalized error of $\hat{\nu}^{\mathcal{H}}$ is difference from zero for each extremity of λ. Furthermore, in some models, we also lose the "parabolic" shape of the curve (see Figure (1a)). The introduction of the corrected estimator may us to recover the same pattern as noticed in Chapter 1. Notice that, in terms of variance, we do not have a strict dominance from one estimator to another as it was mentioned before.

6 Technical section

6.1 Proof of Lemma 6

Consider the following functions from $\{0,1\}^2 \times \mathbb{R}^2$ into $\mathbb{R}$: for $(x,y) \in \mathbb{R}^2$,

$$\begin{aligned} f_1(I,J,X,Y) &= \mathbb{1}_{\{I=1\}}, \quad g_{1,x} = \mathbb{1}_{\{X \le x, I=1\}}, \\ f_2(I,J,X,Y) &= \mathbb{1}_{\{I=1\}}, \quad g_{2,x} = \mathbb{1}_{\{X \le x, I=1\}}, \\ f_3 &= f_1 f_2, \quad g_{3,x,y} = g_{1,x} g_{2,y}. \end{aligned}$$

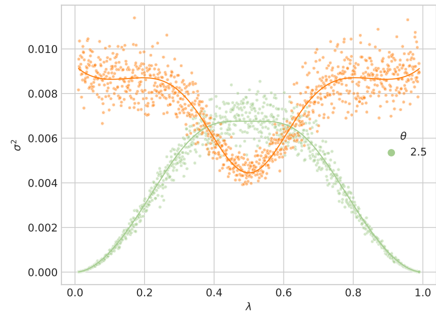
Let P denote the common distribution of the quadruples (I,J,X,Y). Consider the collection of functions

$$\mathcal{F} = \{f_1, f_2, f_3\} \cup \{g_{1,x} : x \in \mathbb{R}\} \cup \{g_{2,y} : y \in \mathbb{R}\} \cup \{g_{3,x,y} : (x,y) \in \mathbb{R}^2\}.$$

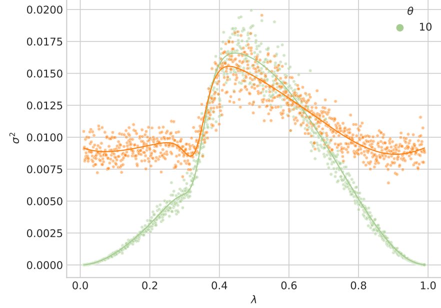
The empirical process $\mathbb{G}_T$ defined by

$$\mathbb{G}_T(f)\sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T} f(I_t,J_t,X_t,Y_t) - \mathbb{E}[f(I_t,J_t,X_t,Y_t)]\right), \quad f \in \mathcal{F},$$

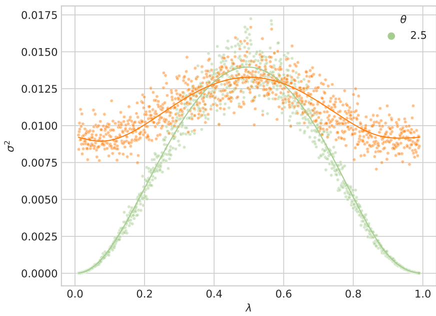
converge in $l^\infty(\mathcal{F})$ to a P-Brownian bridge $\mathbb{G}$ (see [Segers, 2014]). To establish such a statement, results on empirical processes based on the Thoery of Vapnik-Cervonenkis



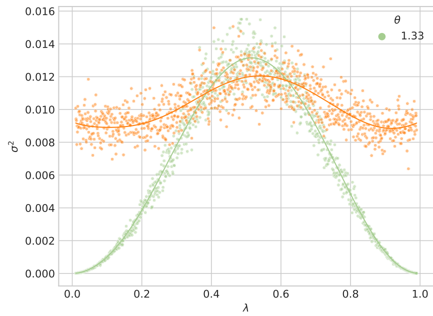
(a) **Asym. Neg. Logistic** ($\theta = 2.5$, $\psi_1 = 1.0$, $\psi_2 = 1.0$)



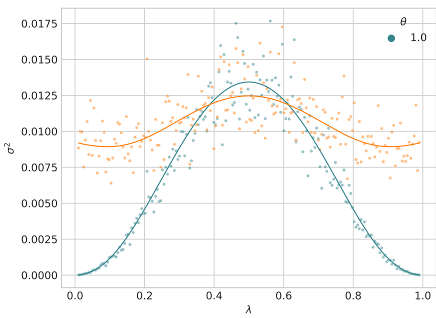
(b) **Asym. Neg. Logistic** ($\theta = 10$, $\psi_1 = 0.5$, $\psi_2 = 1.0$)



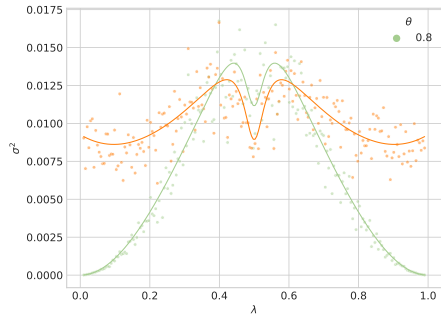
(c) **Asymmetric Logistic** ($\theta = \frac{5}{2}$, $\psi_1 = 0.1$, $\psi_2 = 1.0$)



(d) **Asymmetric Mixed** ($\theta = \frac{4}{3}$, $\kappa = -\frac{1}{3}$)



(e) **Husler-Reiss** ($\theta = 1.0$)



(f) **t-EV** ($\theta = 0.8$, $\chi = 0.2$)

Figure 1: Graph, as a function of λ, of the asymptotic variances of the estimators of the λ-FMadogram for six extreme-value copula models. The empirical of the variance ($\times 256$) based on 500 samples of size $T = 256$. For the fourth first panel, the red line correspond to $\hat{\nu}^{\mathcal{H}}$ while green line correspond to $\hat{\nu}^{\mathcal{H}*}$.

classes (VC-classes) of functions as formulated in [van der Vaart and Wellner, 1996] were used. We now add some line of algebra to establish the weak convergence of the processes $\hat{F}_T(x)$, $\hat{G}_T(y)$ and $\hat{H}_T(x, y)$. These lines are made for the first process as the method is similar for the others. For $x \in \mathbb{R}$,

$$\hat{F}_T(x) = \frac{p_X(x) + T^{-1/2}\mathbb{G}_T g_{1,x}}{p_X + T^{-1/2}\mathbb{G}_T f_1}$$

We may obtain :

$$\begin{aligned} p_X(\hat{F}_T(x) - F(x)) &= T^{-1/2}(\mathbb{G}_T(g_{1,x}) - \mathbb{G}_T(f_1)\hat{F}_T(x)) \\ &= T^{-1/2}(\mathbb{G}_T(g_{1,x} - F(x)f_1)) + T^{-1/2}\mathbb{G}_T(f_1)(F(x) - \hat{F}_T(x)) \end{aligned}$$

Multiplying by $\sqrt{T}$ and dividing by p_X gives :

$$\sqrt{T}(\hat{F}_T(x) - F(x)) = p_X^{-1}(\mathbb{G}_T(g_{1,x} - F(x)f_1)) + p_X^{-1}\mathbb{G}_T(f_1)(F(x) - \hat{F}_T(x))$$

Take a closer look at the second term in the right hand side. By the central limit theorem, we have that $\mathbb{G}_T(f_1) \xrightarrow{d} \mathcal{N}(0, \mathbb{P}(f_1 - \mathbb{P}f_1)^2)$, applying the law of the large number gives us that $(F(x) - \hat{F}_T(x)) = o_{\mathbb{P}}(1)$. With the help of Slutksy theorem, we must claim that :

$$\sqrt{T}(\hat{F}_T(x) - F(x)) = p_X^{-1}(\mathbb{G}_T(g_{1,x} - F(x)f_1)) + o_{\mathbb{P}}(1)$$

As a consequence, we obtain the following limiting process of the Lemma :

$$\beta_1(u) = p_X^{-1}\mathbb{G}(1_{X \leq F^{\leftarrow}(u), I=1} - u 1_{I=1})$$

We know that the covariance of a $\mathbb{P}$-Gaussian process is given by $\mathbb{E}[\mathbb{G}(f)\mathbb{G}(g)] = \mathbb{P}(fg) - \mathbb{P}(f)\mathbb{P}(g)$ where f, g are measurable functions. Now, using that, we have :

$$\begin{aligned} cov[\beta_1(u_1), \beta_1(u_2)] &= p_X^{-2}\mathbb{E}\left[\mathbb{G}(1_{X \leq F^{\leftarrow}(u_1), I=1} - u_1 1_{I=1})\mathbb{G}(1_{X \leq F^{\leftarrow}(u_2), I=1} - u_2 1_{I=1})\right] \\ &= p_X^{-2}(\mathbb{P}\left[(1_{X \leq F^{\leftarrow}(u_1), I=1} - u_1 1_{I=1})(1_{X \leq F^{\leftarrow}(u_2), I=1} - u_2 1_{I=1})\right]) \\ &= p_X^{-2}(\mathbb{P}(I = 1)\mathbb{P}(X \leq F^{\leftarrow}(u_1), X \leq F^{\leftarrow}(u_2)) - u_1 u_2 \mathbb{P}(I = 1)) \\ &= p_X^{-1}(u_1 \wedge u_2 - u_1 u_2) \end{aligned}$$

6.2 Proof of Theorem 3

We do the proof only for the normalized error of $\hat{\nu}^{\mathcal{H}*}$ as the proof of $\hat{\nu}^{\mathcal{H}}$ is clearly similar to this of Theorem 1 in Chapter 1. Using that $\mathbb{E}[F(X)^{\alpha}] = \frac{1}{1+\alpha}$ $(\alpha \neq 1)$, we can write

$\nu(\lambda)$ as :

$$\nu(\lambda) = \frac{1}{2}\mathbb{E}\left[|F^{\lambda}(X) - G^{1-\lambda}(Y)|\right] - \frac{\lambda}{2}\mathbb{E}\left[1 - F^{\lambda}(X)\right] - \frac{1-\lambda}{2}\mathbb{E}\left[1 - G^{1-\lambda}(Y)\right] + \frac{1}{2}\frac{1-\lambda-\lambda^2}{(1+\lambda)(1+1-\lambda)}.$$

Let us note by g_λ the function defined as:

$$g_\lambda \colon [0,1]^2 \to [0,1], \quad (u,v) \mapsto u^{\lambda} \vee v^{1-\lambda} - \frac{1}{2}\left((1-\lambda)u^{\lambda} + \lambda v^{1-\lambda}\right).$$

We are able to write our estimator of the λ-FMadogram (resp. the λ-FMadogram) in missing data framework as an integral with respect to the hybrid copula estimator (resp. the copula function). We then have :

$$\hat{\nu}_T^{\mathcal{H}*}(\lambda) = \frac{1}{\sum_{t=1}^{T} I_t J_t}\sum_{t=1}^{T} g_\lambda(\hat{F}_T(X_t), \hat{G}_T(Y_t)) I_T J_t + c_\lambda = \int_{[0,1]^2} g_\lambda(u,v) d\hat{C}_T^{\mathcal{H}}(u,v) + c_\lambda,$$

$$\nu(\lambda) = \int_{[0,1]^2} g_\lambda(u,v) dC(u,v) + c_\lambda.$$

Where c_λ a constant depending on λ. Using the same tools introduced to prove Lemma A.1, we are able to show that :

$$\sqrt{T}\left(\hat{\nu}_T^{\mathcal{H}*}(\lambda) - \nu(\lambda)\right) = \frac{1}{2}\left((1-\lambda)\int_{[0,1]} \mathbb{C}_T^{\mathcal{H}}(x^{\frac{1}{\lambda}}, 1) dx + \lambda \int_{[0,1]} \mathbb{C}_T^{\mathcal{H}}(1, x^{\frac{1}{1-\lambda}}) dx\right) - \int_{[0,1]} \mathbb{C}_T^{\mathcal{H}}(x^{\frac{1}{\lambda}}, x^{\frac{1}{1-\lambda}}) dx.$$

Consider the function $\phi \colon l^{\infty}([0,1]^2) \to l^{\infty}([0,1]), f \mapsto \phi(f)$, defined by

$$(\phi(f))(\lambda) = \frac{1}{2}\left((1-\lambda)\int_{[0,1]} f(x^{\frac{1}{\lambda}}, 1) dx + \lambda \int_{[0,1]} f(1, x^{\frac{1}{1-\lambda}}) dx\right) - \int_{[0,1]} f(x^{\frac{1}{\lambda}}, x^{\frac{1}{1-\lambda}}) dx.$$

This function is linear and bounded thus continuous. The continous mapping theorem (Theorem 1.3.6 of [van der Vaart and Wellner, 1996]) implies, as $T \to \infty$

$$\sqrt{T}(\hat{\nu}_T - \nu) = \phi(\mathbb{C}_T^{\mathcal{H}}) \rightsquigarrow \phi(S_C),$$

in $l^{\infty}([0,1])$. We note that $S_C(u,1) = \alpha(u,1) - \beta_1(u)$ and $S_C(1,v) = \alpha(1,v) - \beta_2(v)$. Indeed, just remark that for the first one we have $\beta_2(1) = 0$ and $\partial C / \partial u(u,1) = 1$ a.s. We thus obtain our statement.

6.3 Proof of Theorem 5

We will denote by S the index set of sane blocks. For the rest of the section, λ is a fixed constant between 0 and 1.

Lemma 7. *For every positive ϵ, it holds that*

$$\mathbb{P}\bigg\{\ |\hat{\nu}_{MoN}-\nu|>\epsilon\bigg\}\leq\mathbb{P}\bigg\{\ \big|\bar{\nu}_{n_j}-\nu\big|>\epsilon\bigg\}^{K\delta}2^K,\quad j\in S. \tag{2.15}$$

Proof For the first inequality, we have

$$\begin{aligned}\mathbb{P}\bigg\{\ |\hat{\nu}_{MoN}-\nu|>\epsilon\bigg\}&\leq\mathbb{P}\bigg\{\ \bigg|\sum_{k\in S}\mathbb{1}_{\{|\bar{\nu}_{n_j}-\nu|\geq\epsilon\}}\bigg|\geq\frac{K}{2}\bigg\},\\&\leq\mathbb{P}\bigg\{\ \bigg|\sum_{j\in S}\mathbb{1}_{\{|\bar{\nu}_{n_j}-\nu|\geq\epsilon\}}\bigg|\geq K_s-\frac{K}{2}\bigg\},\\&\leq\mathbb{P}\bigg\{\ \bigg|\sum_{j\in S}\mathbb{1}_{\{|\bar{\nu}_{n_j}-\nu|\geq\epsilon\}}\bigg|\geq K_s\left(1-\frac{1}{2}(\frac{1}{2}+\delta)^{-1}\right)\bigg\}.\end{aligned}$$

All these inequalities results from $K\geq K_s\geq K(2^{-1}+\delta)$ and that $K_s+K_o=K$. Notice that the random variable $\sum_{j\in S}\mathbb{1}_{\{|\bar{\nu}_{n_j}-\nu|\}}$ is distributed according to a binomial random variable with K_s trials and probability p_ϵ with

$$p_\epsilon=\mathbb{P}\left\{\big|\bar{\nu}_{n_j}-\nu\big|>\epsilon\right\}.$$

It can thus be upper bounded by

$$\begin{aligned}\sum_{n=\lceil K_s(1-\frac{1}{2}(\frac{1}{2}+\delta)^{-1})\rceil}^{K_s}\binom{K_s}{n}p_\epsilon^n\,(1-p_\epsilon)^{n-K_s}&\leq p_\epsilon^{K_s\left(1-\frac{1}{2}(\frac{1}{2}+\delta)^{-1}\right)}\sum_{n=1}^{K_s}\binom{K_s}{n},\\&\leq p_\epsilon^{K_s\left(1-\frac{1}{2}(\frac{1}{2}+\delta)^{-1}\right)}2^{K_s},\\&\leq p_\epsilon^{K\delta}2^K.\end{aligned}$$

That is our statement.

Lemma 8. *For every $j\in S$ and $\epsilon>0$, we have*

$$p_\epsilon\leq\mathbb{P}\left\{\left|\frac{1}{2n_j}\sum_{t\in B_j}|F(X_t)-G(Y_t)|-\frac{1}{2}\mathbb{E}\,|F(X)-G(Y)|\right|>\frac{\epsilon}{3}\right\} \tag{2.16}$$

$$+\,\mathbb{P}\left\{\underset{t\in B_j}{sup}\,\big|\bar{F}_{n_j}(X_t)-F(X_t)\big|>\frac{2\epsilon}{3}\right\}+\mathbb{P}\left\{\underset{t\in B_j}{sup}\,\big|\bar{G}_{n_j}(Y_t)-G(Y_t)\big|>\frac{2\epsilon}{3}\right\}. \tag{2.17}$$

Proof First, notice that we can obtain the following upper bound

$$\begin{aligned}\left|\bar{\nu}_{n_j}-\nu\right| &= \left|\frac{1}{2n_j}\sum_{t\in B_j}|\bar{F}_{n_j}(X_t)-\bar{G}_{n_j}(Y_t)|-\frac{1}{2}\mathbb{E}|F(X)-G(Y)|\right|,\\
&\leq \left|\frac{1}{2n_j}\sum_{t\in B_j}\left(|\bar{F}_{n_j}(X_t)-\bar{G}_{n_j}(Y_t)|-|F(X_t)-G(Y_t)|\right)\right|\\
&+\left|\frac{1}{2n_j}\sum_{t\in B_j}|F(X_t)-G(Y_t)|-\frac{1}{2}\mathbb{E}|F(X)-G(Y)|\right|.\end{aligned}$$

The first expression can be bounded by

$$\begin{aligned}&\left|\frac{1}{2n_j}\sum_{t\in B_j}\left(|\bar{F}_{n_j}(X_t)-\bar{G}_{n_j}(Y_t)|-|F(X_t)-G(Y_t)|\right)\right|,\\
&\overset{(a)}{\leq}\frac{1}{2n_j}\sum_{t\in B_j}\left||\bar{F}_{n_j}(X_t)-\bar{G}_{n_j}(Y_t)|-|F(X_t)-G(Y_t)|\right|,\\
&\overset{(b)}{\leq}\frac{1}{2n_j}\sum_{t\in B_j}\left|\bar{F}_{n_j}(X_t)-F(X_t)-\left(\bar{G}_{n_j}(Y_t)-G(Y_t)\right)\right||,\\
&\overset{(c)}{\leq}\frac{1}{2}\underset{t\in B_j}{sup}\left|\bar{F}_{n_j}(X_t)-F(X_t)\right|+\frac{1}{2}\underset{t\in B_j}{sup}\left|\bar{G}_{n_j}(Y_t)-G(Y_t)\right|.\end{aligned}$$

We used triangle inequality in (a), $||x|-|y||\leq|x-y|$ in (b) and both triangle inequality and that $\sum_{t=1}^{T}x_t\leq T\underset{t\in\{1,\ldots,T\}}{sup}x_t$ in (c). Since :

$$\begin{aligned}\left\{\left|\bar{\nu}_{n_j}-\nu\right|\leq\epsilon\right\}\supseteq&\left\{\frac{1}{2}\underset{t\in B_j}{sup}\left|\bar{F}_{n_j}(X_t)-F(X_t)\right|\leq\frac{\epsilon}{3}\right\}\cap\left\{\frac{1}{2}\underset{t\in B_j}{sup}\left|\bar{G}_{n_j}(Y_t)-G(Y_t)\right|\leq\frac{\epsilon}{3}\right\}\\
&\cap\left\{\left|\frac{1}{2n_j}\sum_{t\in B_j}|F(X_t)-G(Y_t)|-\frac{1}{2}\mathbb{E}|F(X)-G(Y)|\right|\leq\frac{\epsilon}{3}\right\}.\end{aligned}$$

We thus obtain our lemma.
We can write Equation 2.15 such as

$$exp\left(K\delta log\left(p_\epsilon 2^{\frac{1}{\delta}}\right)\right). \tag{2.18}$$

The DKW inequality (see page 384 in [Boucheron et al., 2013], [Massart, 1990] or in the proof of Theorem 1 in [Alquier et al., 2020] for a similar application) gives us an upper bound for Equation (2.17) in the following form :

$$4exp\left(-\frac{8}{9}n_j\epsilon^2\right)\leq 4exp\left(-\frac{2}{9}n_j\epsilon^2\right).$$

Clearly, we have that

$$\mathbb{E}\left[\frac{1}{2n_j}\sum_{t\in B_j}|F(X_t)-G(Y_t)|\right]=\frac{1}{2}\mathbb{E}|F(X)-G(Y)|,$$

and, for every $t \in B_j$

$$\frac{1}{2n_j}|F(X_t)-G(Y_t)|\leq\frac{1}{n_j}.$$

Applying Hoeffding's inequality permits us to bound Equation (2.16) by

$$2exp\left(-\frac{2}{9}n_j\epsilon^2\right).$$

Summing all these components and the use of Lemma 8 yields to

$$p_\epsilon\leq 6exp\left(-\frac{2}{9}n_j\epsilon^2\right).$$

Plugging this inequality in Equation (2.18) leads to

$$\mathbb{P}\left\{|\hat{\nu}_{MoN}-\nu|>\epsilon\right\}\leq exp\left(K\delta log\left(6e^{-\frac{2\epsilon^2 n_j}{9}}2^{\frac{1}{\delta}}\right)\right).$$

It can be set to η by choosing $K = log(1/\eta)\delta^{-1}$ and ϵ such that $6e^{-\frac{2\epsilon^2 n_j}{9}}2^{\frac{1}{\delta}} = 1/e$, or again

$$\epsilon=\frac{3}{\sqrt{n_j}}log\left(6e2^{\frac{1}{\delta}}\right)=\frac{3}{\sqrt{2}}\frac{log\left(6e2^{\frac{1}{\delta}}\right)}{\delta}\sqrt{\frac{log\left(1/\eta\right)}{T}}.$$

And we are done.

Bibliography

[Alquier et al., 2020] Alquier, P., Chérief-Abdellatif, B.-E., Derumigny, A., and Fermanian, J.-D. (2020). Estimation of copulas via maximum mean discrepancy.

[Angrist and Pischke, 2008] Angrist, J. D. and Pischke, J.-S. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

[Baraud et al., 2016] Baraud, Y., Birgé, L., and Sart, M. (2016). A new method for estimation and model selection: ρ-estimation. *Inventiones Mathematicae*.

[Beirlant et al., 2004] Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2004). *Statistics of Extremes: Theory and Applications*. Wiley. Pagination: 522.

[Bernard et al., 2013] Bernard, E., Naveau, P., Vrac, M., and Mestre, O. (2013). Clustering of Maxima: Spatial Dependencies among Heavy Rainfall in France. *Journal of Climate*, 26(20):7929–7937.

[Boucheron et al., 2013] Boucheron, S., Lugosi, G., and Massart, P. (2013). *Concentration inequalities. A nonasymptotic theory of independence*. Oxford University Press.

[Capéraà et al., 1997] Capéraà, P., Fougères, A.-L., and Genest, C. (1997). A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika*, 84(3):567–577.

[Cochran, 2007] Cochran, W. (2007). *Sampling Techniques, 3Rd Edition*. A Wiley publication in applied statistics. Wiley India Pvt. Limited.

[Coles et al., 1999] Coles, S., Heffernan, J., and Tawn, J. (1999). Dependence measures for extreme value analyses. *Extremes*, 2:339 – 365.

[Cooley et al., 2006] Cooley, D., Naveau, P., and Poncet, P. (2006). *Variograms for spatial max-stable random fields*, pages 373–390. Springer New York, New York, NY.

[Demarta and McNeil, 2005] Demarta, S. and McNeil, A. J. (2005). The t copula and related copulas. *International Statistical Review*, 73(1):111–129.

[Escobar-Bach et al., 2018] Escobar-Bach, M., Goegebeur, Y., and Guillou, A. (2018). Local robust estimation of the Pickands dependence function. *The Annals of Statistics*, 46(6A):2806 – 2843.

[Fermanian et al., 2004] Fermanian, J.-D., Radulovic, D., and Wegkamp, M. (2004). Weak convergence of empirical copula processes. *Bernoulli*, 10(5):847 – 860.

[Gaetan and Guyon, 2008] Gaetan, C. and Guyon, X. (2008). *Modélisation et statistique spatiales*. Mathématiques & applications. Springer, Berlin Heidelberg New York.

[Genest and Segers, 2009] Genest, C. and Segers, J. (2009). Rank-based inference for bivariate extreme-value copulas. *The Annals of Statistics*, 37(5B):2990 – 3022.

[Gudendorf and Segers, 2009] Gudendorf, G. and Segers, J. (2009). Extreme-value copulas.

[Guillou et al., 2014] Guillou, A., Naveau, P., and Schorgen, A. (2014). Madogram and asymptotic independence among maxima. *Revstat Statistical Journal*, 12:119–134.

[Huber, 1964] Huber, P. J. (1964). Robust Estimation of a Location Parameter. *The Annals of Mathematical Statistics*, 35(1):73 – 101.

[Huber, 2011] Huber, P. J. (2011). *Robust Statistics*, pages 1248–1251. Springer Berlin Heidelberg, Berlin, Heidelberg.

[Hüsler and Reiss, 1989] Hüsler, J. and Reiss, R.-D. (1989). Maxima of normal random vectors: Between independence and complete dependence. *Statistics & Probability Letters*, 7(4):283–286.

[Joe, 1990] Joe, H. (1990). Families of min-stable multivariate exponential and multivariate extreme value distributions. *Statistics & Probability Letters*, 9:75–81.

[Lerasle et al., 2019] Lerasle, M., Szabó, Z., Mathieu, T., and Lecué, G. (2019). MONK – Outlier-Robust Mean Embedding Estimation by Median-of-Means. In *ICML 2019 - 36th International Conference on Machine Learning*, Proceedings of Machine Learning Research, Long Beach, United States.

[Little R.J.A., 1987] Little R.J.A., R. D. (1987). *Statistical analysis with missing data.*

[Marcon et al., 2017] Marcon, G., Padoan, S., Naveau, P., Muliere, P., and Segers, J. (2017). Multivariate nonparametric estimation of the pickands dependence function using bernstein polynomials. *Journal of Statistical Planning and Inference*, 183:1–17.

[Massart, 1990] Massart, P. (1990). The Tight Constant in the Dvoretzky-Kiefer-Wolfowitz Inequality. *The Annals of Probability*, 18(3):1269 – 1283.

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

[Nemirovsky and Yudin, 1983] Nemirovsky, A. and Yudin, D. (1983). *Problem Complexity and Method Efficiency in Optimization*. John Wiley, New York.

[Oliveira and Galambos, 1977] Oliveira, J. D. T. and Galambos, J. (1977). The asymptotic theory of extreme order statistics. *International Statistical Review*, 47:230.

[Segers, 2014] Segers, J. (2014). Hybrid copula estimators.

[Sklar, 1959] Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, 8:229–231.

[Smith, 2005] Smith, R. L. (2005). Max-stable processes and spatial extremes.

[Tawn, 1988] Tawn, J. A. (1988). Bivariate extreme value theory: Models and estimation. *Biometrika*, 75(3):397–415.

[van der Vaart and Wellner, 1996] van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Process: With Applications to Statistics*.

Appendix A

Auxiliary results

Theorem A.1 (Theorem 3 of [Fermanian et al., 2004]). *Suppose that H has continuous marginal distribution functions and that the copula function $C(x, y)$ has continuous partial derivatives. Then the empirical copula process $\{\mathbb{C}_T(u, v), 0 \leq u, v \leq 1\}$ converges weakly to a Gaussian process $\{N_C(u, v), 0 \leq u, v \leq 1\}$ in $l^\infty([0, 1]^2)$.*

Under the assumptions defined in Assumption A, the following proposition from [Naveau et al., 2009] hold.

Proposition A.1 (Proposition 3 of [Naveau et al., 2009]). *Suppose that Assumptions A holds and let J be a function of bounded variation, continuous from above and with discontinuities of the first kind. Then:*

$$T^{-1/2} \sum_{t=1}^{T} (J(\hat{F}_T(X_t), \hat{G}_T(Y_t)) - \mathbb{E}[J(F(X), G(Y))])$$

converges in distribution to $\int_{[0,1]} N_C(u, v) dJ(u, v)$ where $N_C(u, v)$ is defined by equation (1.12) and the integral is well defined as a Lebesgue-Stieltjes integral. The special case, $J(x, y) = \frac{1}{2}|x^\lambda - y^{1-\lambda}|$ provide the weak of convergence of the λ-Madogram estimator defined by (1.9) :

$$T^{1/2}\{\hat{\nu}_T(\lambda) - \frac{1}{2}\mathbb{E}[|F^\lambda(X) - G^{1-\lambda}(Y)|]\}$$

converge in distribution to $\int_{[0,1]^2} N_C(u, v) dJ(u, v)$ where the latter integral satisfies :

$$\int_{[0,1]^2} f(x, y) dJ(x, y) = \frac{1}{2} \int_{[0,1]} f(0, y^{1/(1-\lambda)}) dy + \frac{1}{2} \int_{[0,1]} f(x^{1/\lambda}, 0) dx - \int_{[0,1]} f(x^{1/\lambda}, x^{1/(1-\lambda)}) dx \tag{A.1}$$

for all bounded-measurable function $f : [0, 1]^2 \mapsto \mathbb{R}$.

Lemma A.1. *(Lemma A.1 of [Marcon et al., 2017]) For $\lambda \in [0, 1]$, let H be any distri-*

bution function in $[0,1]^2$, let ν_λ be the function defined by

$$\nu_\lambda\colon [0,1]^2 \to [0,1], \quad (u,v) \mapsto u^\lambda \vee v^{1-\lambda} - \frac{1}{2}(u^\lambda + v^{1-\lambda}),$$

Then

$$\begin{aligned}\int_{[0,1]^2} \nu_\lambda(u,v) dH(u,v) = \frac{1}{2}\left(\int_{[0,1]} H(x^{\frac{1}{\lambda}}, 1) dx + \int_{[0,1]} H(1, x^{\frac{1}{1-\lambda}}) dx\right) \\ - \int_{[0,1]} H(x^{\frac{1}{\lambda}}, x^{\frac{1}{1-\lambda}}) dx.\end{aligned} \tag{A.2}$$

Proof We have,

$$u^\lambda \vee v^{1-\lambda} = 1 - \int_{[0,1]} \mathbb{1}_{\{u^\lambda \le x, v^{1-\lambda} \le x\}} dx = 1 - \int_{[0,1]} \mathbb{1}_{\{u \le x^{\frac{1}{\lambda}}, v \le x^{\frac{1}{1-\lambda}}\}} dx,$$

using the same technique, we may have,

$$\frac{1}{2}(u^\lambda + v^{1-\lambda}) = 1 - \frac{1}{2}\left(\int_{[0,1]} \mathbb{1}_{\{u \le x^{\frac{1}{\lambda}}\}} + \mathbb{1}_{\{v \le x^{\frac{1}{1-\lambda}}\}}\right) dx,$$

We obtain by substracting the two terms above and integration with respect to H,

$$\begin{aligned}\int_{[0,1]^2} v_\lambda(u,v) dH(u,v) = \frac{1}{2}\int_{[0,1]^2}\int_{[0,1]} \left(\mathbb{1}_{\{u \le x^{\frac{1}{\lambda}}\}} + \mathbb{1}_{\{v \le x^{\frac{1}{1-\lambda}}\}}\right) dH(u,v) dx \\ - \int_{[0,1]^2}\int_{[0,1]} \mathbb{1}_{\{u \le x^{\frac{1}{\lambda}}, v \le x^{\frac{1}{1-\lambda}}\}} dH(u,v) dx\end{aligned}$$

Applying Fubini lead us to the conclusion.

Appendix B

Auxiliary results

Theorem B.1 (Theorem 2.3 in [Segers, 2014]). *If conditions A and C holds, then uniformly in $u \in [0,1]^2$,*

$$\begin{aligned} r_T\{\hat{C}_T(u,v) - C(u,v)\} &= r_T\{\hat{H}_T((F,G)^{\leftarrow}(u,v) - C(u,v)\} \\ &\quad - \dot{C}_1(u,v) r_T\{\hat{F}_T(F^{\leftarrow}(u)) - u\} 1_{(0,1)}(u) \\ &\quad - \dot{C}_2(u,v) r_T\{\hat{G}_T(G^{\leftarrow}(v)) - v\} 1_{(0,1)}(v) + \circ_{\mathbb{P}}(1) \end{aligned}$$

as $T \to \infty$. Hence in $l^\infty([0,1]^2)$ equipped with the supremum norm, as $T \to \infty$,

$$(r_T\{\hat{C}_T(u,v) - C(u,v)\})_{u,v\in[0,1]^2} \rightsquigarrow (\alpha(u,v) - \dot{C}_1(u,v)\beta_1(u) - \dot{C}_2(u,v)\beta_2(v))_{u,v\in[0,1]^2} \quad \text{(B.1)}$$

The processes α, β_1 and β_2 have continuous trajectories almost surely. The right-hand side in (B.1) is well-defined because $\beta_j(0) = \beta_j(1) = 1$ almost surely with $j \in \{1,2\}$.