Introduction

0.1 Context

Management of environmental resources often requires the analysis of multivariate extreme values. In climate studies, extreme events such as heavy precipitation and record temperatures represent a major challenges due to their dire consequences. In the classical statistical 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. 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.

Also, inference methods for assessing the extremal dependence have been increasingly in demand. The most popular approach is based on second moment of the underlying random variables, the covariance. It is well known that only linear dependence can be captured by the covariance that it is characterizing only for a few special classes of distribution. As a beneficial alternative of dependence modeling, the concept of copulas going back to [Sklar, 1959]. The copula $C : [0,1]^2 \rightarrow [0,1]$ of a random vector (X, Y) allows us to separate the effect of dependence from the effects of the marginal distribution such as :

$$\mathbb{P}\left\{X \leq x, Y \leq y\right\} = C\left(\mathbb{P}\left\{X \leq x\right\}, \mathbb{P}\left\{Y \leq y\right\}\right).$$

The main consequence of this identity is that the copula completely characterizes the stochastic dependence between X and Y. Investigating the notion of copulas within the framework of multivariate extreme value theory leads to the so called extreme value copulas.

Some extreme events, such as heavy precipitation or wind speed has also spatial characteristics and geostatisticians are striving to better understand the physical process in hand. In geostatistic, we often consider $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space, S a set of locations and $(E, \mathcal{E})$ a measurable state space. One can 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 by [Cooley et al., 2006]

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

where $F(u) = \mathbb{P}(X(s) \leq u)$ is named as the FMadogram. His link to the pairwise extremal dependence function (Section 4.3 of [Coles et al., 1999]) or the Pickands dependence function ([Pickands, 1981]) make him an interesting quantity to capture the dependence between the extremas of stochastic processes or random variables. Furthermore, this quantity may be seen as a dissimilarity measure among bivariate maxima to be used for clustering time-series as shown by [Bernard et al., 2013] or [Bador et al., 2015].

The main drawback of this quantity is that she only focus on the value of the diagonal section of the pairwise extremal dependence function. In the bivariate case, the FMadogram characterize solely the extremal dependence coefficient for random variables X and Y (see Section 8.2.7 of [Beirlant et al., 2004]). To overpass this drawback, [Naveau et al., 2009] introduce the λ-FMadogram defined as,

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

for every $\lambda \in [0, 1]$.

This quantity characterize the pairwise extremal dependence function outside the diagonal section but also the whole Pickands dependence function ([Marcon et al., 2017]) and contribute to the vast litterature of the estimation of the Pickands dependence function for bivariate extreme value copulas (see for example [Pickands, 1981], [Deheuvels, 1991], [Hall and Tajvidi, 2000] or [Capéraà et al., 1997]). Statisticians may estimate this quantity but the classical results may applied only if the data in hands are clean as possible. This induce that the process of data collection has not been corrupted such as the data table is complete, each observation is independent from others and that the implicit law of the observations is still the same.

Nevertheless, as the volume of data expands, the problem of missing or contaminated data has been increasingly present in many fields. It frequently happens that some of the individuals of a sample from a multivariate population are not observed. If a sample be represented in matrix form by allowing the rows to represent the individuals and the columns the variables, then the matrix of the type of sample with which we are concerned is sparse. 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. A useful reference for general parametric statistical inferences with missing data was provided by [Little R.J.A., 1987].

Considering a sample from a random vector (X, Y) of incomplete data,

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

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 (1) is basically created by the double sampling or two phase sampling (see Chapter 12 of [Cochran, 2007]). Samples like (1) may 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. Beside of missing observations, the process of data might be disturb in a way that innerly deteriorate the quality of some data and one may ask that the estimation process should be robust.

The topic of 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 errors are low (see Section 1.1 from [Baraud et al., 2016]). In other words, an estimator is 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 report, 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. It is a main goal of this report to develop estimator of the (λ)-FMadogram under these sticky situation. To achieve our goal, we make of use of the hybrid copula estimator ([Segers, 2014]) for the missing data framework and we leverage the idea of Median-Of-MeaNs (MoN) for the contaminated data scheme.

0.2 Definitions and Notation

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 $C : [0,1]^2 \to [0,1]$ 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]

$$C(u,v) = (uv)^{A\left(\frac{log(v)}{log(uv)}\right)}, \tag{2}$$

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

$$C(u^t, u^{1-t}) = u^{A(t)}. \tag{3}$$

Let $(X_t, Y_t)_{t=1,\dots,T}$ 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)

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

If this relationship hold, H is said to be a multivariate extreme value distribution. We will call by FMadogram the following quantity

$$\nu = \frac{1}{2}\mathbb{E}\left[|F(X) - G(Y)| \right], \tag{4}$$

and the λ-FMadogram by the expression

$$\nu(\lambda) = \frac{1}{2}\mathbb{E}\left[\left|F^{\lambda}(X) - G^{1-\lambda}(Y)\right| \right]. \tag{5}$$

A classical estimator of the λ-FMadogram when the margins F, G are unknown is

$$\hat{\nu}(\lambda) = \frac{1}{2T}\sum_{t=1}^{T} \left| \hat{F}_T^{\lambda}(X_t) - \hat{G}_T^{1-\lambda}(Y_t) \right|, \tag{6}$$

with $\hat{F}_T$ (resp. $\hat{G}_T$) the empirical cumulative distribution function of X (resp. Y). We suppose that we observe sequentially a quadruple defined by

$$(I_t X_t, J_t Y_t, I_t, J_t), \quad t \in \{1, \dots, T\}, \tag{7}$$

where $X_t = 0$ (resp. $Y_t = 0$) if X_t (resp. Y_t) is missing, otherwise $I_t = 1$ (resp. $J_t = 1$), *i.e.* at each $t \in \{1, \dots, T\}$, one or both entries may be missing. The probability of observing a realization 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,

$$\hat{F}_T(u) = \frac{\sum_{t=1}^T 1_{\{X_t \leq u\}} I_t}{\sum_{t=1}^T I_t}, \quad \hat{G}_T(v) = \frac{\sum_{t=1}^T 1_{\{Y_t \leq v\}} J_t}{\sum_{t=1}^T J_t}, \quad \hat{H}_T(u,v) = \frac{\sum_{t=1}^T 1_{\{X_t \leq u, Y_t \leq v\}} I_t J_t}{\sum_{t=1}^T I_t J_t}. \quad (8)$$

Here, we weight the estimator by the number of observed data which is a natural estimator (if divided by T) of the probabilities of missing. We have all tools in hand to define the *hybrid copula estimator* introduced by [Segers, 2014],

$$\hat{C}_T^{\mathcal{H}}(u,v) = \hat{H}_T(\hat{F}_T(u), \hat{G}_T(v)). \quad (9)$$

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 :

$$\mathbb{C}_T^{\mathcal{H}}(u,v) = r_T \left(\hat{C}_T^{\mathcal{H}}(u,v) - C(u,v) \right). \quad (10)$$

In order to propose a robust estimator we will assume that the sample is partitioned into K disjoint subsets $B_1, \ldots, B_K$ 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 of $\mathbb{E}[f(X,Y)]$ by

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

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

$$\hat{f}_{MoN} = \underset{z \in \mathbb{R}}{argmin} \sum_{j=1}^K \left| \mathbb{P}_{n_j} f - z \right|, \quad (11)$$

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

$$\hat{f}_{MoN} = med(\mathbb{P}_{n_1} f, \ldots, \mathbb{P}_{n_K} f), \quad (12)$$

is a solution of Equation (11).

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 $f : \mathcal{X} \to \mathbb{R}$, let $||f||_\infty = sup_{x \in \mathcal{X}} |f(x)|$. Here, we use the abbreviation $Qf = \int f dQ$ for a given measurable function f and signed measure Q. The arrows $\overset{a.s.}{\to}$, $\overset{d}{\to}$ 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]).

This work is organized as follows : In Chapter 1 we state some results on the weak convergence of the estimator of the λ-FMadogram with missing data. We also propose a closed formula for the asymptotic variance of the λ-FMadogram for a fixed $\lambda \in [0,1]$. To propose a robust estimator of the FMadogram, we leverage the idea of Median-Of-MeaNs (MoN) and state a concentration inequality that this estimator does verify.

Chapter 2 will present our results in a finite-sample framework. The asymptotic variance of the normalized estimation error of several models would be drawn with their empirical counterpart obtained through simulation. We also propose a reproduction of the experiment of the λ-FMadogram with a Smith's process as find in [Naveau et al., 2009] and we will explain the augmentation of the Mean Squared Error while h is close to zero. This phenomenon would be also thoroughly explained through simulation and a counterexample.

For the ease of reading, we postponed all technical arguments and proofs in Chapter 3.

Chapter 1

Non parametric estimation of the Madogram on non-clean data

1 Definition of the estimator

For the rest of this report we will assume that the copula C is of extreme value type as defined in Equation (2). 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 2 in Section 1). Therefore if $v \searrow 0$, then $\partial C(u,v)/\partial u \to 0$ as required.

In the missing data framework given by Equation (7), based on a identical and independent copies $(I_1 X_1, J_1 Y_1, I_1, J_1), \dots, (I_T X_T, J_T Y_T, I_T, J_T)$, we defined the following estimator of the λ-FMadogram

$$\hat{\nu}_T^{\mathcal{H}}(\lambda) = \frac{1}{2\sum_{t=1}^T I_t J_t} \sum_{t=1}^T \left| \hat{F}_T^{\lambda}(X_t) - \hat{G}_T^{1-\lambda}(Y_t) \right| I_t J_t. \tag{1.1}$$

One may verify that outside the missing data framework, *i.e.* with $p_X = p_Y = p_{XY} = 1$, that $\hat{\nu}^{\mathcal{H}}(\lambda) = \hat{\nu}(\lambda)$. Before going further, let us briefly talk about our estimator. Our estimator defined in (1.1) 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 :

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

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 supplementary variance when estimating. Hence we can force our estimator as in [Naveau et al., 2009] to satisfay these endpoint corrections. This leads to the following definition :

$$\begin{aligned} \hat{\nu}_T^{\mathcal{H}*}(\lambda) = &\frac{1}{2\sum_{t=1}^T I_t J_t} \sum_{t=1}^T \left| \hat{F}_T^{\lambda}(X_t) - \hat{G}_T^{1-\lambda}(Y_t) \right| I_t J_t - \frac{\lambda}{2\sum_{t=1}^T I_t J_t} \sum_{t=1}^T \{1 - \hat{F}_T^{\lambda}(X_t)\} I_t J_t \\ &- \frac{1-\lambda}{2\sum_{t=1}^T I_t J_t} \sum_{t=1}^T \{1 - \hat{G}_T^{1-\lambda}(Y_t)\} I_t J_t + \frac{1-\lambda+\lambda^2}{2(2-\lambda)(1+\lambda)} \end{aligned} \tag{1.2}$$

Nevertheless, in missing data framework, 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.

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

Proposition 1. *Let (X, Y) a $\mathbb{R}^2$-valued random vector of distribution H. We have, for each $\lambda \in [0, 1]$,*

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

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

$$\nu(\lambda) = \frac{A(\lambda)}{A(\lambda) + \lambda(1-\lambda)} - \frac{1}{2}\left(\frac{1}{1+\lambda} + \frac{1}{1+1-\lambda}\right). \tag{1.3}$$

The proof is given in Section 2.

Remark 1. *The upper bound (resp. the lower bound) in (i) is exactly the value of the λ-FMadogram when X and Y are independent (resp. perfectly positive dependent),* i.e. *when $A(t) = 1$ (resp. $A(t) = t \vee (1-t)$).*

Now, we give some precisions under the missing mechanism,

Assumption B. *We suppose for all $t \in \{1, \ldots, T\}$, 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 $t \in \{1, \ldots, T\}$ such that $I_t J_t \neq 0$.*

Under this assumption, we state the strong consistency of our hybrid estimator of the λ-FMadogram.

Proposition 2. *Let* $(I_1X_1, J_1Y_1, I_1, J_1), \dots, (I_TX_T, J_TY_T, I_T, J_T)$ *an* i.i.d. *sample. We have, under Assumption B for a fixed* $\lambda \in [0,1]$, *as* $T \to \infty$

$$\hat{\nu}_T^{\mathcal{H}}(\lambda) \xrightarrow{a.s.} \nu(\lambda).$$

In order to propose a robust estimator of the FMadogram as defined in Equation (4). 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 the bound and λ and using that $||x|^\lambda - |y|^\lambda| \leq |x-y|^\lambda$.

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.

Let $B_1, \dots, B_K$ a partition of the set $\{1, \dots, T\}$. Denote by $\hat{F}_{n_j}$ (resp. $\hat{G}_{n_j}$) 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

$$\hat{\nu}_{MoN} = \underset{z \in \mathbb{R}}{argmin} \sum_{j=1}^{K} \left| \hat{\nu}_{n_j} - z \right|, \tag{1.4}$$

where $\hat{\nu}_{n_j} = \frac{1}{2n_j} \sum_{t \in B_j} |\hat{F}_{n_j}(X_t) - \hat{G}_{n_j}(Y_t)|$. That is, in Equation (11), we take $f(x,y) = |x-y|$ and $\mathbb{P}_{n_j} = C_{n_j}$ the empirical copula constructed on the block B_j.

Assumption 1. *The sample* $((X_1,Y_1), \dots, (X_T, Y_t))$ *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 $\delta = 1/2$.

2 Main results

Without missing data, the weak convergence of normalized estimation error of the empirical copula process has already been proved by [Fermanian et al., 2004] under the sole Assumption

A. This statement make use of previous results on the Hadamard differentiability of the map $\phi : D([0,1]^2) \to l^\infty([0,1]^2)$ which transforms the cumulative distribution function H into its copula function C (see lemma 3.9.28 from [van der Vaart and Wellner, 1996]). With the hybrid copula estimator, we need a technical assumption in order to guarantee the weak convergence of the process $\mathbb{C}_T^{\mathcal{H}}$ (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 1 in Appendix), the stochastic process $\mathbb{C}_T^{\mathcal{H}}$ converges weakly to the tight Gaussian process S_C the process defined by,

$$S_C(u,v) = \alpha(u,v) - \frac{\partial C(u,v)}{\partial u}\beta_1(u) - \frac{\partial C(u,v)}{\partial v}\beta_2(v), \quad \forall (u,v) \in [0,1]^2. \tag{1.5}$$

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 :

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

Furthermore, we are able to compute their covariance functions. This is summarised in the following lemma and we add some technical details available in Chapter 3.

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

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

and

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

We have all tools in hand to consider the weak convergence of the stochastic processes $(\sqrt{T}(\hat{\nu}^{\mathcal{H}}(\lambda)-\nu(\lambda)))_{\lambda\in[0,1]}$ and $(\sqrt{T}(\hat{\nu}^{\mathcal{H}*}(\lambda)-\nu(\lambda)))_{\lambda\in[0,1]}$. 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 Assumptions A, B, C we have the weak convergence in $l^{\infty}([0,1])$ for the hybrid estimator defined in (1.1) and (1.2), as $T \to \infty$,*

$$\sqrt{T}\left(\hat{\nu}_T^{\mathcal{H}}(\lambda)-\nu(\lambda)\right) \rightsquigarrow \left(\frac{1}{2}\int_{[0,1]} \alpha(x^{\frac{1}{\lambda}},1)-\beta_1(x^{\frac{1}{\lambda}})dx + \frac{1}{2}\int_{[0,1]} \alpha(1,x^{\frac{1}{1-\lambda}})-\beta_2(x^{\frac{1}{1-\lambda}})dx \right. \\ \left. - \int_{[0,1]} S_C(x^{\frac{1}{\lambda}},x^{\frac{1}{1-\lambda}})dx\right)_{\lambda\in[0,1]}, \tag{1.6}$$

$$\sqrt{T}\left(\hat{\nu}_T^{\mathcal{H}*}(\lambda)-\nu(\lambda)\right) \rightsquigarrow \left(\frac{1-\lambda}{2}\int_{[0,1]} \alpha(x^{\frac{1}{\lambda}},1)-\beta_1(x^{\frac{1}{\lambda}})dx + \frac{\lambda}{2}\int_{[0,1]} \alpha(1,x^{\frac{1}{1-\lambda}})-\beta_2(x^{\frac{1}{1-\lambda}})dx \right. \\ \left. - \int_{[0,1]} S_C(x^{\frac{1}{\lambda}},x^{\frac{1}{1-\lambda}})dx\right)_{\lambda\in[0,1]}. \tag{1.7}$$

We present also a concentration inequality that the MoN-based estimator of the FMadogram may verified. We suppose without loss of generality that $n_j = \lceil T/K \rceil$ for every $j \in \{1,\dots,K\}$. We can prove the following deviation bounds for our MoN-based estimator. We thus add some remark on the concentration bound.

Theorem 2. *(Consistency and 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 $1-\eta$,*

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

Details of the proof are available in Section 5 in Chapter 3.

Remark 2.
- *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.*

3 Some corrolaries

As an integral of a tight Gaussian process, we know that the two normalized estimation errors follows a centered Gaussian variable for a given $\lambda \in [0,1]$. Furthermore, a few computations are able to give a closed form of the variance of the limiting Gaussian law as an integral of the Pickands dependence function. This is summarized with the following proposition.

Proposition 3. *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*

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

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

The quantities $(\sigma_i^2)_{i\in\{1,2,3\}}$ and $(\sigma_{ij})_{i,j\in\{1,2,3\},i<j}$ are detailed in the corresponding Chapter. 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 $\hat{\nu}_T^{\mathcal{H}}(\lambda)$.

Outside the missing data framework when $p_X = p_Y = p_{XY} = 1$, the hybrid copula estimator become the empirical copula process. The limiting Gaussian process $\mathbb{C}_T$ of the normalized error of the empirical copula process is given by

$$N_C(u,v) = B_C(u,v) - \frac{\partial C}{\partial u}(u,v)B_C(u,1) - \frac{\partial C}{\partial u}(u,v)B_C(1,v), \tag{1.8}$$

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

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

In this setup, we can show that almost surely

$$B_C(u,1) - \frac{\partial C}{\partial u}(u,v)B_C(u,1) = 1, \quad \forall u \in]0,1[, \quad B_C(1,v) - \frac{\partial C}{\partial v}(u,v)B_C(1,v) = 1, \quad \forall v \in]0,1[.$$

Thus, proceding as in the proof of Theorem 1, we are able to show that

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

And we retrieve the asymptotic limit in law of the normalized estimation error of the λ-FMadogram as studied in [Marcon et al., 2017]. For a fixed $\lambda \in (0,1)$, [Naveau et al., 2009] has prove that the asymptotic law can be written as

$$\int_{[0,1]^2} f(x,y)dJ(x,y) = \frac{1}{2}\int_{[0,1]} f(1, y^{1/(1-\lambda)})dy + \frac{1}{2}\int_{[0,1]} f(x^{1/\lambda}, 1)dx - \int_{[0,1]} f(x^{1/\lambda}, x^{1/(1-\lambda)})dx, \tag{1.9}$$

for every measurable and bounded function $f : [0,1]^2 \to \mathbb{R}$. Some details explaining Equation (1.9) are given in Lemma 6 in Chapter 3. The special case $J(x,y) = 2^{-1}|x^\lambda - y^{1-\lambda}|$ satisfies the conditions, then some computations gives that :

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

Note that, without missing data, we are able to show that normalized estimation error of the estimator of the λ-FMadogram and his endpoint corrections are the same. We refer the reader to Section 8 for details.

We are able to infer the closed form without integral of the Pickands 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 Assumption A and if $C(u,v) = uv$, then the asymptotic variance of $\sqrt{T}(\hat{\nu}_T(\lambda) - \nu(\lambda))$ has the following form, for $\lambda \in (0,1)$*

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

Chapter 2

Finite setting results

1 Definition of the models

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

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

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.

2 Complete data framework

2.1 Simulation and experiment

A vast Monte Carlo study was used to illustrate Theorem 1 of Chapter 1 in finite-sample settings while none datas are missing. 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 we estimate the empirical version of the normalized estimation error's variance, namely

$$
\mathcal{E}_T(\lambda) := \widehat{Var}\left(\sqrt{T}\,(\hat{\nu}^*_T(\lambda) - \nu(\lambda))\right), \tag{2.1}
$$

and was computed by taking the variance over 500 samples. For each normalized estimation error, we represent its theoretical asymptotic variance using the form exhibits in Theorem 1.

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].

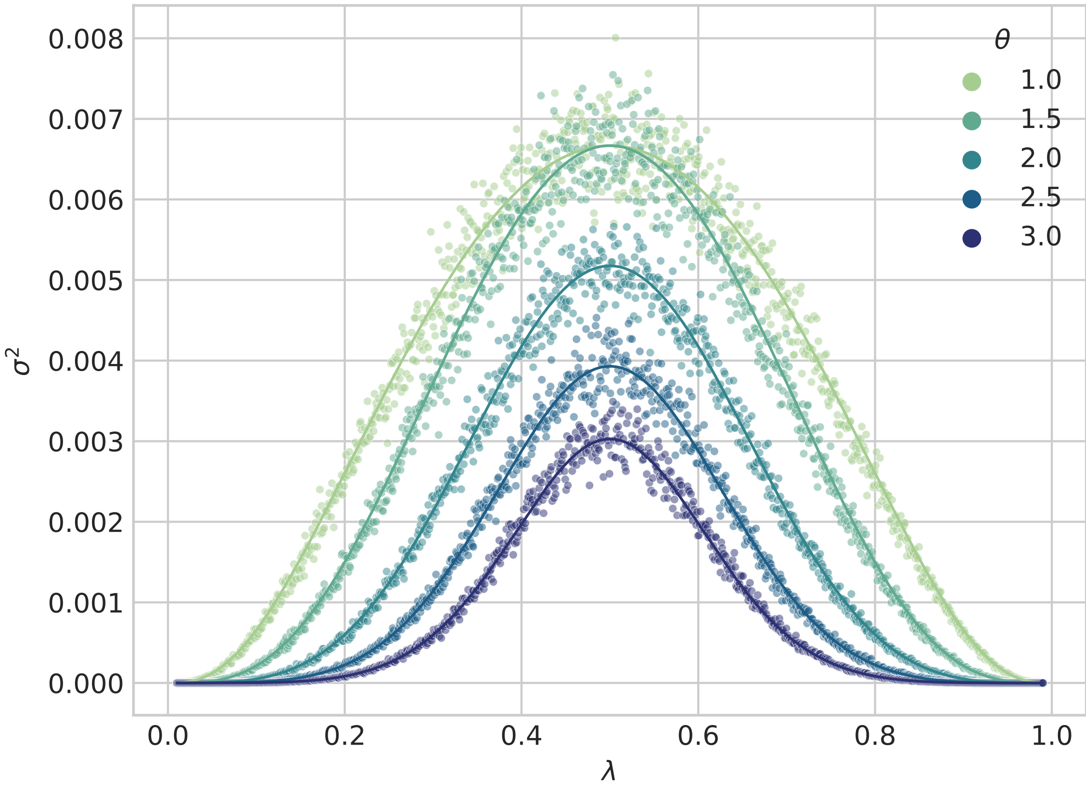
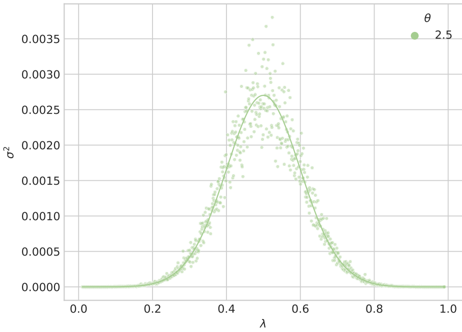
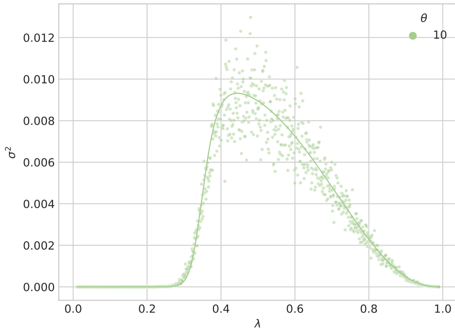


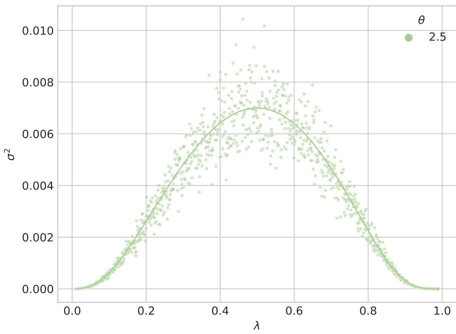
Figure 1: Variance of the normalized estimation error based on 500 samples of size $T = 256$ from the Gumbel copula with $\theta = \{1.0, 1.5, 2.0, 2.5, 3.0\}$ chosen in such a way that $\lambda \in \{i/1000 : i = 10, \ldots, 990\}$.



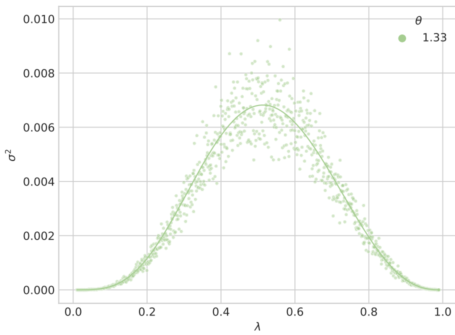
(a) **Asym. Neg. Logistic** ($\theta = 2.5$, $\psi_1 = 1.0$, $\psi_2 = 1.0$)



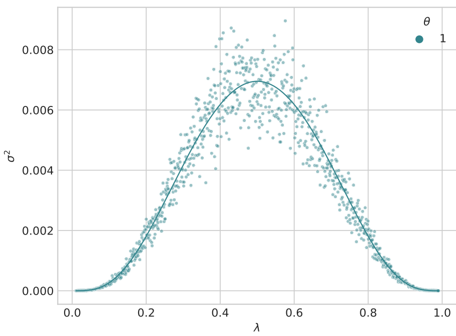
(b) **Asym. Neg. Logistic** ($\theta = 10$, $\psi_1 = 0.5$, $\psi_2 = 1.0$)



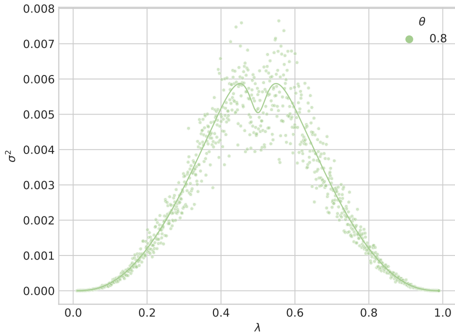
(c) **Asymmetric Logistic** ($\theta = \frac{5}{2}$, $\psi_1 = 0.1$, $\psi_2 = 1.0$)



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

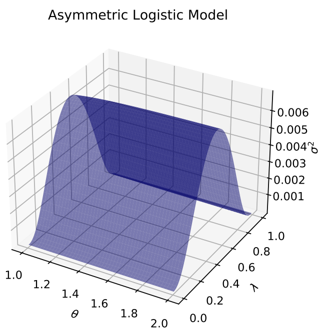


(e) **Hüsler-Reiss** ($\theta = 1$)

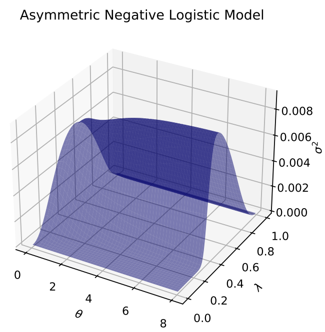


(f) **t-EV** ($\theta = 0.8$, $\chi = 0.2$)

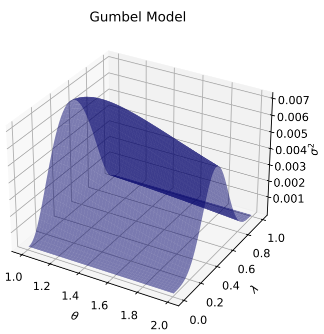
Figure 2: Equation (2.1), as a function of λ, of the asymptotic variances of the normalized estimation errors based on 500 samples of size $T = 256$ of the λ-FMadogram for six extreme-value copula models.



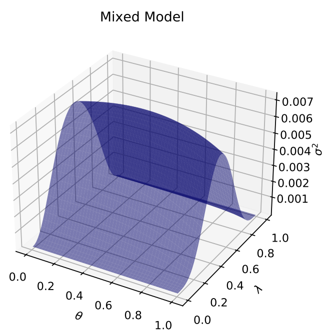
(a) **Asym. Logistic ($\psi_1 = 0.1$, $\psi_2 = 1.0$)**



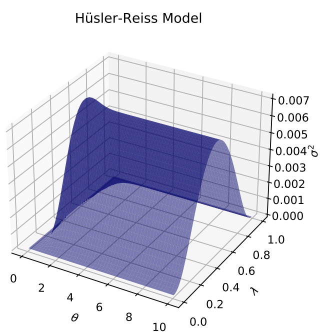
(b) **Neg. Logistic ($\psi_1 = 0.5$, $\psi_2 = 1.0$)**



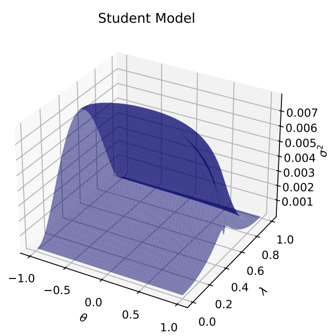
(c) **Gumbel ($\psi_1 = 0.1$, $\psi_2 = 1.0$)**



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



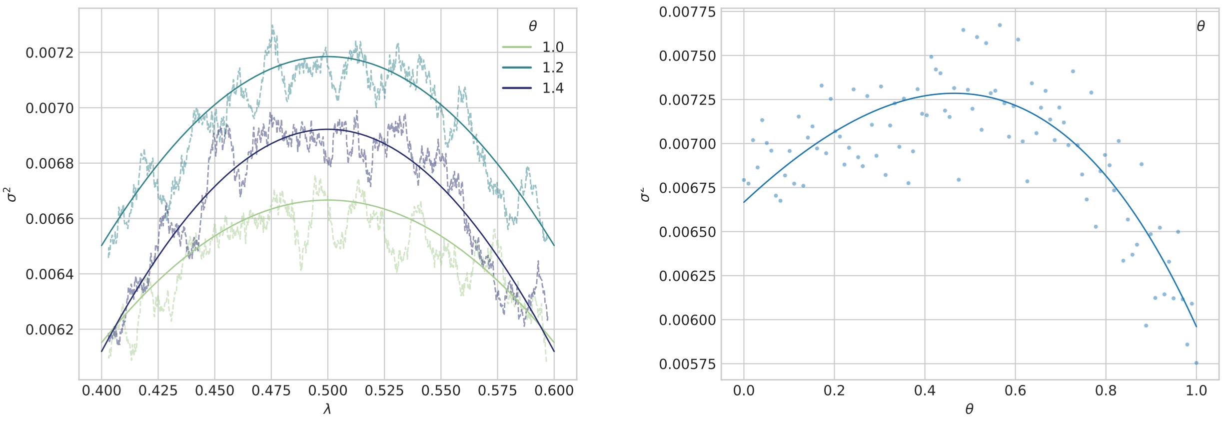
(e) **Hüsler-Reiss**



(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.

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



(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 $\lambda \in \{i/1000 : i = 400, \dots, 600\}$. 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 $\theta \in \{i/100 : i = 0, \dots 100\}$. The solid line is the asymptotic variance computed numerically using Theorem 1.

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 in Section 10 of Chapter 3 where we exhibit a counterexample, it is not always the case that the variance is lower as the variables are becoming positively dependent.

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.

2.3 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. We recall the bivariate distribution for the max-stable process model proposed by Smith is equal to :

$$\mathbb{P}\left(X(s) \leq u, X(s+h) \leq v\right) = exp\left[-\frac{1}{u}\Phi(\frac{a}{2} + \frac{1}{a}log(\frac{v}{u})) - \frac{1}{v}\Phi(\frac{a}{2} + \frac{1}{a}log(\frac{u}{v}))\right], \tag{2.2}$$

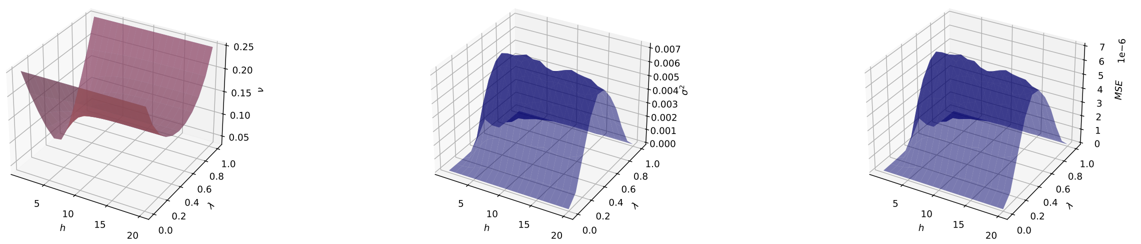
where Φ denotes the standard normal distribution function, with $a^2 = (h^\top \Sigma^{-1} h)$ and Σ is a covariance matrix. In case of isotropic field, we set $\Sigma = \sigma^2 I_2$. For this kind of process, the pairwise extremal dependence function $V_h(\cdot,\cdot)$ is given by :

$$V_h(u,v) = \frac{1}{u}\Phi\left(\frac{a}{2} + \frac{1}{a}log(\frac{v}{u})\right) + \frac{1}{v}\Phi\left(\frac{a}{2} + \frac{1}{a}log(\frac{u}{v})\right). \tag{2.3}$$

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

$$\nu(h,\lambda) = \frac{V_h(\lambda, 1-\lambda)}{1 + V_h(\lambda, 1-\lambda)} - c(\lambda), \tag{2.4}$$

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 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.



(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 (2.2) with $\Sigma = 25I_2$. 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.

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 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 discussions.

2.4 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 :

$$C_{\chi,\theta}(u,v) = \int_{]-\infty, t_\chi^{\leftarrow}(u)]} \int_{]-\infty, t_\chi^{\leftarrow}(v)]} \frac{1}{2\pi(1-\theta^2)^{1/2}} \left(1 + \frac{x^2 - 2\theta xy + y^2}{\chi(1-\rho^2)}\right)^{-(\chi+2)/2} dydx, \qquad (2.5)$$

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 $X_{1j}, \dots, X_{Mj}$, $j \in \{1, \dots, n\}$, a block of M 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 $(\bigvee_{i=1}^M X_{i1}, \dots, \bigvee_{i=1}^M X_{in})$ of length n.

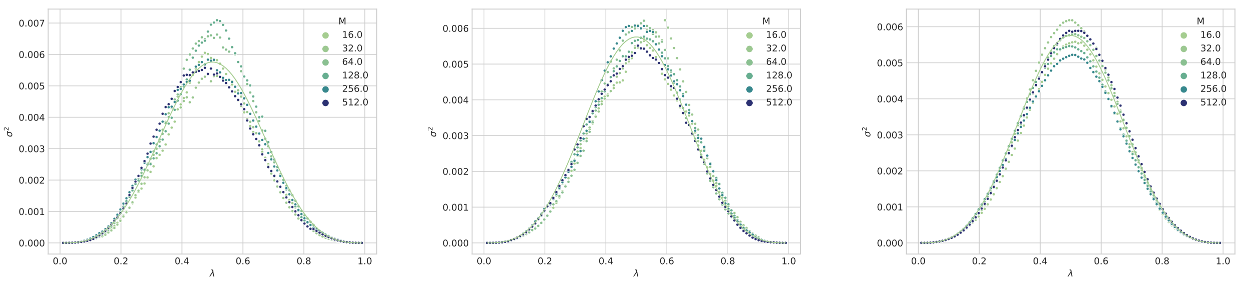


Figure 6: Simulation results obtained by generating $T \in \{128, 256, 512\}$ blocks maximas of length $M \in \{16, 32, 64, 128, 256, 512\}$ from t-copula. For each $\lambda \in \{i/100, i = 1, \dots, 99$, we compute the empirical normalized estimation error on 100 estimator of λ-FMadogram. The solid line is the theoretical asymptotic variance of t-EV copula.

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.

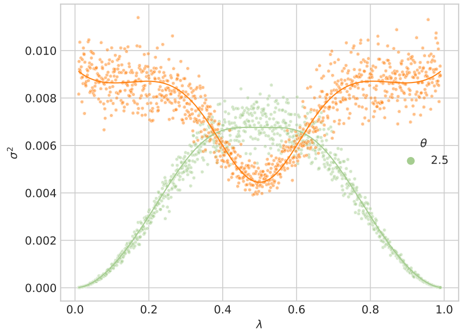
3 Missing data framework

In each experiment, we estimate the empirical variance 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

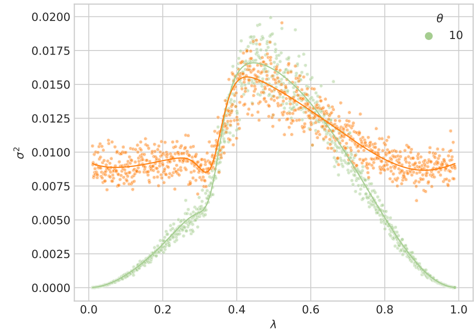
$$I \sim \mathcal{B}(p_X), \quad J \sim \mathcal{B}(p_Y).$$

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

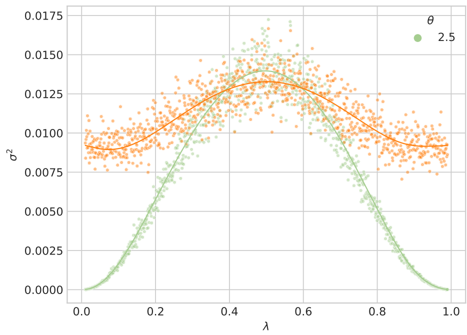
Figure 7 presents the results for the six models defined in Chapter 1. For each model, the green (resp. red) line presents the asymptotic variance of the normalized estimator error of $\hat{\nu}_T^{\mathcal{H}}$ (resp. $\hat{\nu}_T^{\mathcal{H}*}$ detailed by Proposition 3. For each $\lambda \in \{\frac{i}{1000}, i = \frac{10}{1000}, \cdots \frac{990}{1000}\}$, we estimate its empirical counterpart. 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 different from zero for each extremity of λ. Furthermore, in some models, we also lose the "parabolic" shape of the curve (see Figure (7a)). 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.



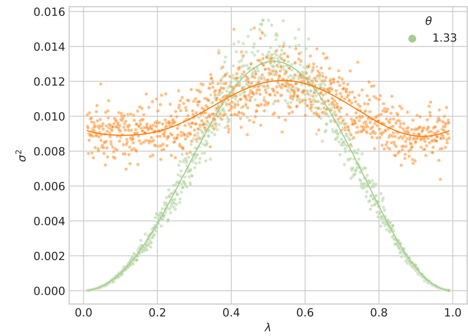
(a) **Asym. Neg. Logistic** ($\theta = 2.5$, $\psi_1 = 1.0$, $\psi_2 = 1.0$)



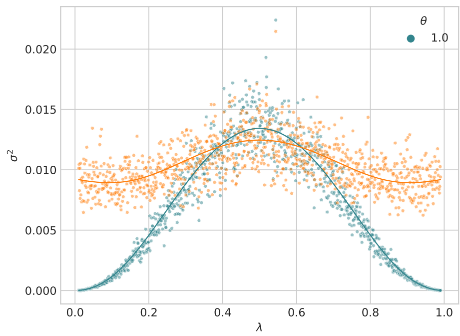
(b) **Asym. Neg. Logistic** ($\theta = 10$, $\psi_1 = 0.5$, $\psi_2 = 1.0$)



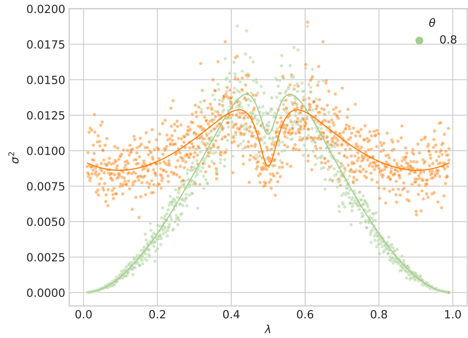
(c) **Asymmetric Logistic** ($\theta = \frac{5}{2}$, $\psi_1 = 0.1$, $\psi_2 = 1.0$)



(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 7: Equation (2.1), as a function of λ, of the asymptotic variances of the estimators of the λ-FMadogram for six extreme-value copula models. The empirical variance 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 $\hat{\nu}^{\mathcal{H}*}$.

4 Robustness on synthetic datas

We now compare our MoN-based estimator to the FMadogram. We designed two types of outliers :

- *top-left*: the outliers are drawn *i.i.d* from $\mathcal{U}([-3,-2] \times [2,3])$.
- *bottom-right*: the outliers are drawn *i.i.d* from $\mathcal{U}([2,3] \times [-3,-2])$.

In each case, sane data are sampled from the desired copula model thus inverted by the quantile function of a standard Gaussian distribution. Both types of outliers are considered in the two frameworks of contamination, *i.e.* *à la* Huber and adversarial one. In the adversarial contamination, our rule is to sample again points which are closer to the point $(0,0)$.

Figure 8 present an illustration of contaminated data for a given copula model for all types of contaminations and outliers we consider. We report the squared bias of each estimator for all the models considered in Figure 9. When there are no outliers, the MoN-based estimator and the FMadogram has the same bias, this is due that $\hat{\nu}_{MoN} = \hat{\nu}$ when $K = 1$.

Otherwise, our MoN-based estimator is not always better than the Madogram. This is the case when the Pickands dependence function is symmetric at $\lambda = 0.5$. When it is not, we can see that our MoN-based estimator become much more reliable when the Pickands is asymmetric. Indeed, when the asymmetry of the dependence model induces a concentration of the points on the bottom right, we see that the asymmetric negative logistic model is better than the FMadogram for the *bottom-right* types of outliers.

The same observation might be seen in the asymmetric logistic model which performs comparatively for the Huber's contamination and better for the adversarial one. For these asymmetric models, the MoN-based estimator is more robust than the FMadogram because the *bottom-right* types of outliers (resp. *top-left*) upsets the ranks of sane data for the asymmetric negative logistic model (resp. for the asymmetric logistic model).

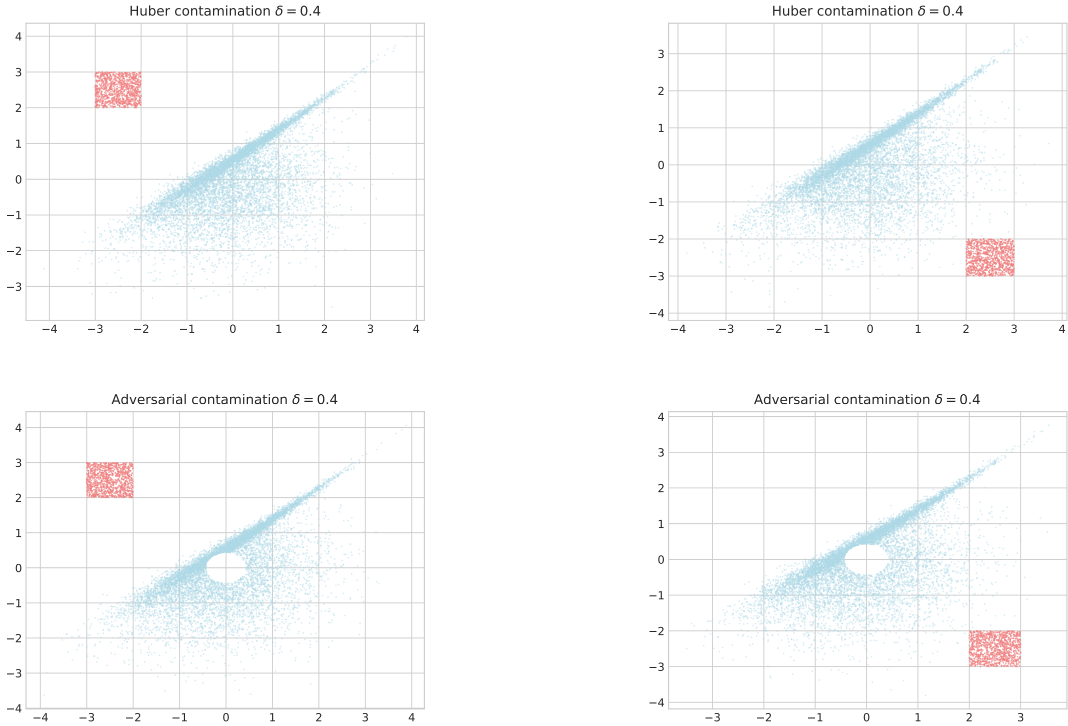


Figure 8: Sample of 10000 data with a fraction of 10% of outliers. Sane points are depicted in blue while contaminated ones are in red. In the first row (resp. second row), Huber's contamination (resp. adversarial contamination) are shown for *top-left* and *bottom-right* types of outliers. The model considered here is an extreme value copula with Pickands dependence function of the type asymmetric negative logistic with $\theta = 10$, $\psi_1 = 0.5$, $\psi_2 = 1.0$.

(a) **Gumbel** (b) **Asy. Log.** (c) **Asy. Neg. Log.**

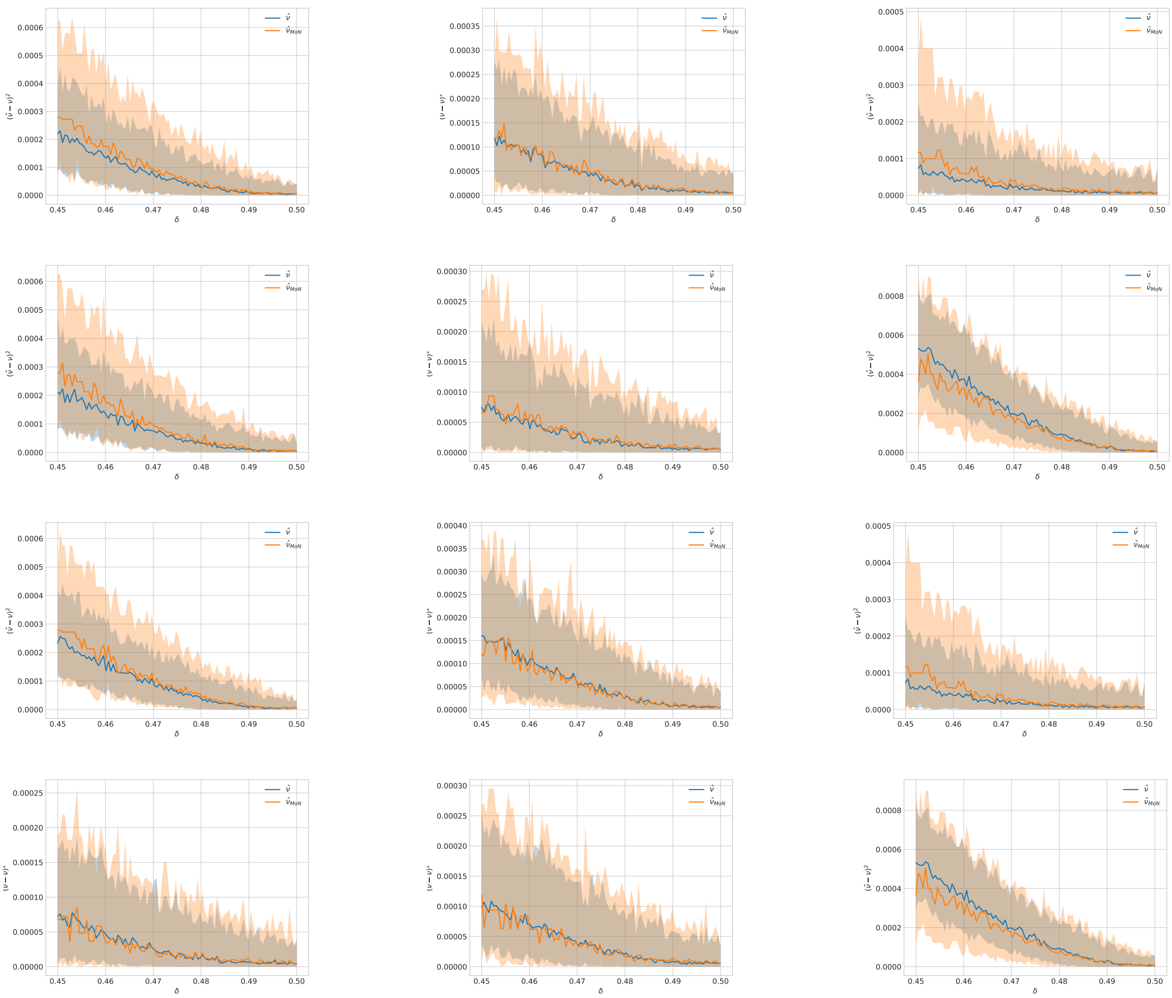


Figure 9: Squared bias of the MoN-based estimator (orange) and the FMadogram (blue) and the 90% confidence band computed on 100 estimators build on a sample of length $T = 1000$. The two first row is respectively the *top-left* and *bottom-right* types of outliers for the Huber's contamination model while the two next are for the adversarial contamination ones. For Gumbel's model, we took $\theta = 1.5$, for the asymmetric logistic model, we consider $\theta = 2.5$, $\psi_1 = 0.1$, $\psi_2 = 1.0$ while the asymmetric negative logistic model is defined as in Figure 8.

Chapter 3

Mathematical section

1 Study of the Pickands dependence function

Lemma 2. *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 3. *If A admits a derivative, then $\lim_{t \to 0^+} A'(t)$ and $\lim_{t \to 1^-} A'(t)$ exists and are finite.*

Proof As A is convex and derivable, it follows that $A'(\cdot)$ is increasing. Furthermore, in the proof of Lemma 2, 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).

2 Proof of Proposition 1 & Proposition 2

The first statement results directly from *(ii)*. To show *(ii)* 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.

We thus prove Proposition 2. The estimator $\hat{\nu}(\lambda)$ is strongly consistent since it holds

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

The second term converges almost surely to zero by the strong Law of Large Numbers and

Assumption B. For the first term, we have

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

which converges almost surely to zero according to the strong Law of Large Numbers and also with Assumption B.

□

3 Proof of Lemma 1

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

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

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 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 :

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

4 Proof of Theorem 1

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. Using that $\mathbb{E}[F(X)^\alpha] = \frac{1}{1+\alpha}$ $(\alpha \neq 1)$, we can write $\nu(\lambda)$ as :

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

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 ϕ: $l^\infty([0,1]^2) \to l^\infty([0,1]), f \mapsto \phi(f)$, 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.

5 Proof of Theorem 2

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 4. *For every positive ϵ, it holds that*

$$\mathbb{P}\left\{ |\hat{\nu}_{MoN} - \nu| > \epsilon \right\} \leq \mathbb{P}\left\{ \left|\hat{\nu}_{n_j} - \nu\right| > \epsilon \right\}^{K\delta} 2^K, \quad j \in S. \tag{3.1}$$

Proof For the first inequality, we have

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

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 $\sum_{j\in S} \mathbb{1}_{\{|\hat{\nu}_{n_j} - \nu|\}}$ is distributed according to a binomial random variable with K_s trials and probability p_ϵ with

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

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

$$
p_\epsilon \leq \mathbb{P}\left\{ \left| \frac{1}{2n_j} \sum_{t\in B_j} |F(X_t) - G(Y_t)| - \frac{1}{2}\mathbb{E}\,|F(X) - G(Y)| \right| > \frac{\epsilon}{3} \right\} \tag{3.2}
$$

$$
+ \mathbb{P}\left\{ \underset{t\in B_j}{sup} \left| \hat{F}_{n_j}(X_t) - F(X_t) \right| > \frac{2\epsilon}{3} \right\} + \mathbb{P}\left\{ \underset{t\in B_j}{sup} \left| \hat{G}_{n_j}(Y_t) - G(Y_t) \right| > \frac{2\epsilon}{3} \right\}. \tag{3.3}
$$

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

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

The first expression can be bounded by

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

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 :

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

We thus obtain our lemma.

We can write Equation 3.1 such as

$$
exp\left(K\delta log\left(p_\epsilon 2^{\frac{1}{\delta}}\right)\right). \tag{3.4}
$$

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 (3.3) 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 (3.2) by

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

Summing all these components and the use of Lemma 5 yields to

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

Plugging this inequality in Equation (3.4) 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.

6 Proof of Proposition 3

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 :

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

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.

$$\int_{[0,1]}\int_{[0,u]} C(v^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}})\frac{\partial C(v^{\frac{1}{\lambda}}, v^{\frac{1}{1-\lambda}})}{\partial u}dvdu. \tag{3.5}$$

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

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

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.

7 A Lemma for Equation (1.9)

Lemma 6. *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:*

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

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

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

Notice that the class

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

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.

8 Asymptotic Behavior of the normalized errors of Equation (1.2) under complete data

It is readily verified that

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

Using the same argument as in the proof of Theorem 1, we can show, with complete data, 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 extended Slutsky's lemma (example 1.4.7 of [van der Vaart and Wellner, 1996])

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

That's what we wanted to prove.

9 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 (3.5) equals :

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

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.

10 A counter example against variance's monotony with respect to an increasing 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 7. *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*

$$Var\left(\int_{[0,1]} N_C(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}})du\right) > \frac{1}{150}. \tag{3.6}$$

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 8. *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.

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.

[Bador et al., 2015] Bador, M., Naveau, P., Gilleland, E., Castellà, M., and Arivelo, T. (2015). Spatial clustering of summer temperature maxima from the cnrm-cm5 climate model ensembles & e-obs over europe. *Weather and Climate Extremes*, 9:17–24. The World Climate Research Program Grand Challenge on Extremes – WCRP-ICTP Summer School on Attribution and Prediction of Extreme Events.

[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*, 84:567–577.

[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.

[Deheuvels, 1991] Deheuvels, P. (1991). On the limiting behavior of the pickands estimator for bivariate extreme-value distributions. *Statistics & Probability Letters*, 12(5):429–439.

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

[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.

[Hall and Tajvidi, 2000] Hall, P. and Tajvidi, N. (2000). Distribution and dependence-function estimation for bivariate extreme-value distributions. *Bernoulli*, 6(6):835–844.

[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*. Wiley Series in Probability and Statistics.

[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.

[Pickands, 1981] Pickands, J. (1981). Multivariate extreme value distribution. *Proceedings 43th, Session of International Statistical Institution, 1981.*

[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.

[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*. Springer.

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:*

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

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

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

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 distribution 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

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

Proof We have,

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

using the same technique, we may have,

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

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

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

Applying Fubini lead us to the conclusion.

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

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

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

$$(r_T\{\hat{C}_T(u,v) - C(u,v)\})_{u,v\in[0,1]^2} \rightsquigarrow (\alpha(u,v) - \dot{C}_1(u,v)\beta_1(u) - \dot{C}_2(u,v)\beta_2(v))_{u,v\in[0,1]^2} \tag{A.3}$$

The processes α, β_1 and β_2 have continuous trajectories almost surely. The right-hand side in (A.3) is well-defined because $\beta_j(0) = \beta_j(1) = 1$ almost surely with $j \in \{1,2\}$.