1 Background and notations

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $\mathbf{X} = (X_1, \dots, X_d)$ be a d-dimensional random vector with values in $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. This random vector has a joint distribution function F and its margins are denoted by $F_j(x) = \mathbb{P}\{X_j \leq x\}$ for all $x \in \mathbb{R}$ and $j \in \{1, \dots, d\}$. We assume that C is an extreme value copula. We consider independent and identically distributed copies $\mathbf{X}_1, \dots, \mathbf{X}_n$ of $\mathbf{X}$. In presence of missing data, we do not observe a complete vector $\mathbf{X}_i$ for $i \in \{1, \dots, n\}$. We introduce $\mathbf{I}_i \in \{0,1\}^d$ which satisfies, $\forall j \in \{1, \dots, d\}$, $I_{i,j} = 0$ if $X_{i,j}$ is not observed. To formalize incomplete observations, we introduce the incomplete vector $\tilde{\mathbf{X}}_i$ with values in the product space $\bigotimes_{j=1}^d (\mathbb{R} \cup \{\mathrm{NA}\})$ such as

$$\tilde{X}_{i,j} = X_{i,j} I_{i,j} + \mathrm{NA}(1 - I_{i,j}), \quad i \in \{1, \dots, n\},\, j \in \{1, \dots, d\}.$$

Throughout the paper, we assume that we measured *i.i.d.* additional covariates on the i-th individual, denoted by $\mathbf{Z}_i = (Z_{i,1}, \dots, Z_{i,p})$, which are completely observed with law $\mathbf{Z}$ defined on the same probability space as $\mathbf{X}$. We thus suppose that we observe a $(2d+p)$-tuple such as :

$$\mathbf{W}_i \triangleq (\mathbf{I}_i, \tilde{\mathbf{X}}_i, \mathbf{Z}_i), \quad i \in \{1, \dots, n\},$$

i.e. at each $i \in \{1, \dots, n\}$ several entries of $\mathbf{X}_i$ may be missing. We also suppose that for all $i \in \{1, \dots, n\}$, $\mathbf{I}_i$ are *i.i.d.* copies from $\mathbf{I} = (I_1, \dots, I_d)$ where I_j is distributed according to a Bernoulli random variable $\mathcal{B}(p_j)$ with $p_j = \mathbb{P}\{I_j = 1 | \mathbf{X}, \mathbf{Z}\}$ for $j \in \{1, \dots, d\}$. We denote by $p_0 = \mathbb{P}\{I_1 = 1, \dots, I_d = 1 | \mathbf{X}, \mathbf{Z}\}$. Suppose we believe that the reason for missingness depends on $\mathbf{Z}_i$, and, moreover conditional on $\mathbf{Z}_i$, *i.e.* $\mathbf{X}_i$ has no additional effect on the probability of missing. Then, we state the following MAR assumption.

Assumption A. *We suppose that for all $i \in \{1, \dots, n\}$, the missing mechanism is Missing At Random (MAR), that is*

$$\begin{aligned}
&\mathbb{P}\{I_j = 1 | \boldsymbol{X}, \boldsymbol{Z}\} = \mathbb{P}\{I_j = 1 | \boldsymbol{Z}\} = \pi_j(\boldsymbol{Z}), \quad \forall j \in \{1, \dots, d\},\\
&\mathbb{P}\{I_1 = 1, \dots, I_d = 1 | \boldsymbol{X}, \boldsymbol{Z}\} = \mathbb{P}\{I_1 = 1, \dots, I_d = 1 \ | \boldsymbol{Z}\} = \eta(\boldsymbol{Z}).
\end{aligned}$$

The MAR assumption requires that, conditional on the observed individuals characteristics, the missing mechanism is independent of the potential value of $\mathbf{X}_i$. Under the Assumption A, F_j and F are identified by :

$$F_j(x) = \mathbb{E}\left[\frac{\mathbb{1}_{\{X_j \leq x_j\}} I_j}{\pi_j(\mathbf{Z})}\right], \quad \forall x \in \mathbb{R}, \quad F(\mathbf{x}) = \mathbb{E}\left[\frac{\mathbb{1}_{\{\mathbf{X} \leq \mathbf{x}\}} \Pi_{j=1}^d I_j}{\eta(\mathbf{Z})}\right], \quad \forall \mathbf{x} \in \mathbb{R}^d. \tag{1}$$

Based on (1), the inverse probability weighting estimator ([Horvitz and Thompson, 1952]) for F_j and F are :

$$F_{n,j}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{\mathbb{1}_{\{X_{i,j}\leq x\}}}{\pi_j(\mathbf{Z}_i)}I_{i,j}, \quad \forall x \in \mathbb{R}, \quad F_n(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n}\frac{\mathbb{1}_{\{\mathbf{X}_i\leq \mathbf{x}\}}}{\eta(\mathbf{Z}_i)}\Pi_{j=1}^{d}I_{i,j}.$$

The idea here is twofold : first, we estimate nonparametrically the margins using all available data of the corresponding series. Second, each observation is leverage by the inverse of the probability to observe it conditionally to $\mathbf{Z}_i$. Thus, an observation which are rarely to be observe will be adequately ponderate if so. Since $\pi_j(\cdot)$ and $\eta(\cdot)$ are unknown in practice, it is typically estimated either parametrically or non parametrically. Parametric methods are easy to implement, but will lead to erroneous results if the model is misspecified. Furthermore, parametric models such as logistic regression may estimate spurious estimation of the probability, *i.e.* an estimation close to zero, which produces important biais in the statistical analysis. This paper adopts the covariate principle of [Hamori et al., 2019] and [Hamori et al., 2020] to estimate π_j for every $j \in \{1,\ldots,d\}$ and η. The key insight is that the following equations holds for any continuous bounded function u and $j \in \{1,\ldots,d\}$

$$\mathbb{E}\left[\frac{u(\mathbf{Z}_i)}{\pi_j(\mathbf{Z}_i)}I_{i,j}\right] = \mathbb{E}\left[u(\mathbf{Z}_i)\right], \quad \forall i\{1,\ldots,n\}, \quad \mathbb{E}\left[\frac{u(\mathbf{Z}_i)}{\eta(\mathbf{Z}_i)}\Pi_{j=1}^{d}I_{i,j}\right] = \mathbb{E}\left[u(\mathbf{Z}_i)\right].$$

The estimator of π_j denoted as $\hat{\pi}_{jK}$, should satisfy the sample counterpart, *i.e.*

$$\frac{1}{n}\sum_{i=1}^{n}\frac{u_K(\mathbf{Z}_i)}{\hat{\pi}_{jK}(\mathbf{Z}_i)}I_{i,j} = \frac{1}{n}\sum_{i=1}^{n}u_K(\mathbf{Z}_i),$$

where $u_K(\cdot) = (u_{K,1}(\cdot),\ldots,u_{K,K}(\cdot))$ is a known sieve basis function that can approximate any suitable function u arbitrarily well, and $K\to\infty$ as $n\to\infty$. Common specifications of $u_K(\cdot)$ include orthonormal polynomials, B-splines and wavelets. Similarly, the estimator of η denoted as $\hat{\eta}_K$, should satisfy the sample counterpart

$$\frac{1}{n}\sum_{i=1}^{n}\frac{u_K(\mathbf{Z}_i)}{\hat{\eta}(\mathbf{Z}_i)}\Pi_{j=1}^{d}I_{i,j} = \frac{1}{n}\sum_{i=1}^{n}u_K(\mathbf{Z}_i).$$

For notational conveniency, we set that the dimension K is the same for all estimators. Observe that $F_{n,j}$ and F_n are increasing, takes positive values and are *càdlàg*. Their are not proper distribution functions, however, since in general the weigths $I_{i,j}\{\pi_j(\mathbf{Z}_i)\}^{-1}, 1\leq i\leq n$, do not sum up to one. Plug the estimator of π_j and η would be the same. Alternatively, we might consider the functions

$$\hat{F}_{n,j}(x) = \sum_{i=1}^{n} \frac{\mathbb{1}_{\{X_{i,j} \le x\}}}{\hat{\pi}_j(\mathbf{Z}_i)} I_{i,j} \Big/ \sum_{i=1}^{n} \frac{1}{\hat{\pi}_j(\mathbf{Z}_i)} I_{i,j}, \quad x \in \mathbb{R}, \tag{2}$$

$$\hat{F}_n(\mathbf{x}) = \sum_{i=1}^{n} \frac{\mathbb{1}_{\{\mathbf{X}_i \le x\}}}{\hat{\eta}(\mathbf{Z}_i)} \Pi_{j=1}^{d} I_{i,j} \Big/ \sum_{i=1}^{n} \frac{1}{\hat{\eta}(\mathbf{Z}_i)} \Pi_{j=1}^{d} I_{i,j}, \quad \forall \mathbf{x} \in \mathbb{R}^d. \tag{3}$$

We recall the definition of the hybrid copula estimator of [Segers, 2015]

$$\hat{C}_n^{\mathcal{H}}(\mathbf{u}) = \hat{F}_n\left(\hat{F}_{n,1}^{\leftarrow}(u_1), \dots, \hat{F}_{n,d}^{\leftarrow}(u_d)\right), \quad \forall \mathbf{u} \in [0,1]^d, \tag{4}$$

where $F^{\leftarrow}$ denotes the generalized inverse function of F, *i.e.* $F^{\leftarrow}(u) = \inf\{x \in \mathbb{R}, F(x) \ge u\}$ with $0 < u < 1$. The normalized estimation error of the hybrid copula estimator is :

$$\mathbb{C}_n^{\mathcal{H}}(\mathbf{u}) = \sqrt{n}\left(\hat{C}_n^{\mathcal{H}}(\mathbf{u}) - C(\mathbf{u})\right), \quad \mathbf{u} \in [0,1]^d. \tag{5}$$

The hybrid copula is an extension of the classical empirical copula which combines an estimator of the multivariate cumulative distribution function with estimators of the marginal cumulative distribution function that are not necessarily equal to the margins of the joint estimator. This feature is of main interest when dealing with missing data as we want to extract as much information as possible from the data and the classical framework would lead to listwise deletion.

On the condition that the first-order partial derivatives of the copula function C exists and are continuous on a subset of the unit hypercube, [Segers, 2012] obtained weak convergence of the normalized estimation error of the classical empirical copula process (see [Deheuvels, 1979]). To satisfy this condition, we introduce the following assumption as suggested in [Segers, 2012] (see Example 5.3).

Assumption B.

1. *The distribution function F has continuous margins $F_1, \dots, F_d$.*

2. *For every $j \in \{1, \dots, d\}$, the first-order partial derivative $\dot{\ell}_j$ of ℓ with respect to x_j exists and is continuous on the set $\{x \in [0,\infty)^d : x_j > 0\}$.*

The Assumption B1 guarantees that the representation $F(\mathbf{x}) = C(F_1(x_1), \dots, F_d(x_d))$ is unique on the range of $(F_1, \dots, F_d)$. Under the Assumption B2, the first-order partial derivatives of C with respect to u_j denoted as $\dot{C}_j$ exists and are continuous on the set $\{\mathbf{u} \in [0,1]^d : 0 < u_j < 1\}$. Without missing data, the weak convergence of the normalized estimation error of the empirical copula process has been proved by [Fermanian et al., 2004] under a more restrictive condition than Assumption B. The difference being that C should be continuously differentiable on the closed cube. This statement makes use of previous results on the Hadamard differentiability of the map $\phi : D([0,1]^2) \to \ell^\infty([0,1]^2)$, where $D([0,1]^2)$ denotes the Skorohod space of all right continuous function on $[0,1]$ with left limits, which transforms the cumulative distribution function H into its copula function C (see also 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 $\mathbb{C}_n^{\mathcal{H}}$ in (5) (see [Segers, 2015]). We note for convenience marginal distributions and quantile functions into vector valued functions $\mathbf{F}_d$ and $\mathbf{F}_d^{\leftarrow}$:

$$\mathbf{F}_d(\mathbf{x}) = (F_1(x_1), \ldots, F_d(x_d)), \quad \mathbf{x} \in \mathbb{R}^d, \quad \mathbf{F}_d^{\leftarrow}(\mathbf{u}) = (F_1^{\leftarrow}(u_1), \ldots, F_d^{\leftarrow}(u_d)), \quad \mathbf{u} \in [0,1]^d.$$

Assumption C. *In the space $\ell^\infty(\mathbb{R}^d) \otimes (\ell^\infty(\mathbb{R}), \ldots, \ell^\infty(\mathbb{R}))$ equipped with the topology of uniform convergence, we have the joint weak convergence*

$$\left(\sqrt{n}(\hat{F}_n - F); \sqrt{n}(\hat{F}_{n,1} - F_1), \ldots, \sqrt{n}(\hat{F}_{n,d} - F_d)\right) \rightsquigarrow (\alpha \circ \boldsymbol{F}_d, \beta_1 \circ F_1, \ldots, \beta_d \circ F_d)\,,$$

where the stochastic processes α and $\beta_j, j \in \{1, \ldots, d\}$ take values in $\ell^\infty([0,1]^d)$ and $\ell^\infty([0,1])$ respectively, and are such that $\alpha \circ F$ and $\beta_j \circ F_j$ have continuous trajectories on $[-\infty, \infty]^d$ and $[-\infty, \infty]$ almost surely.

A smoothness condition is required to establish the large-sample behavior of $\hat{F}_{n,j}$ for $j \in \{1, \ldots, d\}$ and $\hat{F}_n$.

Assumption D. *For any $j \in \{1, \ldots, d\}$, the conditional distribution fuctions $F_j(x|z) = \mathbb{P}\{X_j \leq x | \boldsymbol{Z} = z\}$ and $F(\mathbf{x}|z) = \mathbb{P}\{\boldsymbol{X} \leq \mathbf{x} | \boldsymbol{Z} = z\}$ are Lipschitz continuous in x and $\mathbf{x}$.*

We now state a Lemma that guarantees that our estimators defined in (2) and (3) does verify Assumption C.

Lemma 1. *Under Assumptions A, B, D and regularity conditions 2-7 provided from [Hamori et al., 2019], the vector $(\sqrt{n}(\hat{F}_n - F); \sqrt{n}(\hat{F}_{n,1} - F_1), \ldots, \sqrt{n}(\hat{F}_{n,d} - F_d))$ where*

$\hat{F}_n$ and $\hat{F}_{n,j}$ are defined in (2) and (3) verifies :

$$\sup_{x\in\mathbb{R}}\left|\sqrt{n}\{\hat{F}_{n,j}(x)-F_j(x)\}-\sqrt{n}\left\{\frac{1}{n}\sum_{i=1}^{n}\psi_j(\mathbf{W}_i,x)\right\}\right|=o_{\mathbb{P}}(1),$$
$$\sup_{\mathbf{x}\in\mathbb{R}^d}\left|\sqrt{n}\{\hat{F}_{n}(\mathbf{x})-F_j(\mathbf{x})\}-\sqrt{n}\left\{\frac{1}{n}\sum_{i=1}^{n}\Psi(\mathbf{W}_i,\mathbf{x})\right\}\right|=o_{\mathbb{P}}(1),$$

where,

$$\psi_j(\mathbf{W}_i,\cdot)=\frac{\mathbb{1}_{\{X_{i,j}\leq\cdot\}}}{\pi_j(\boldsymbol{Z}_i)}I_{i,j}-\frac{F_j(\cdot|\boldsymbol{Z}_i)}{\pi_j(\boldsymbol{Z}_i)}(I_{i,j}-\pi_j(\mathbf{Z}_i))-F_j(\cdot),$$
$$\Psi(\mathbf{W}_i,\cdot)=\frac{\mathbb{1}_{\{\mathbf{X}_i\leq\cdot\}}}{\eta(\boldsymbol{Z}_i)}I_{i,j}-\frac{F(\cdot|\boldsymbol{Z}_i)}{\eta(\boldsymbol{Z}_i)}(I_{i,j}-\eta(\mathbf{Z}_i))-F(\cdot).$$

As a consequence, our estimators does verify Assumption C with

$$\beta_j(u_j)=\mathbb{G}\left(\frac{\mathbb{1}_{\{X_j\leq F_j^{\leftarrow}(u_j)\}}}{\pi_j(\boldsymbol{Z})}I_j-\frac{F_j(F_j^{\leftarrow}(u_j)|\boldsymbol{Z})}{\pi_j(\boldsymbol{Z})}(I_j-\pi_j(\mathbf{Z}))-u_j\right),$$
$$\alpha(\mathbf{u})=\mathbb{G}\left(\frac{\mathbb{1}_{\{\mathbf{X}_i\leq\mathbf{F}_d^{\leftarrow}(\mathbf{u})\}}}{\eta(\boldsymbol{Z})}\Pi_{j=1}^{d}I_j-\frac{F(\mathbf{F}_d^{\leftarrow}(\mathbf{u})|\boldsymbol{Z})}{\eta(\boldsymbol{Z})}(\Pi_{j=1}^{d}I_j-\eta(\mathbf{Z}))-C(\mathbf{u})\right),$$

where $\mathbb{G}$ is a tight Gaussian process.

The proof is postponed in Appendix A. Note that, under the MCAR assumption, we have that $\eta(\mathbf{Z})=p_0\in[0,1]$, $\pi_j(\mathbf{Z})=p_j\in[0,1]$, $F(\mathbf{x}|Z)=F(\mathbf{x})$ and $F_j(x|Z)=F_j(x|Z)$ for $j\in\{1,\ldots,d\}$. We thus obtain :

$$\beta_j(u_j)=p_j^{-1}\mathbb{G}\left(\mathbb{1}_{\{X_j\leq F_j^{\leftarrow}(u_j)\}}I_j-u_jI_j\right),\quad j\in\{1,\ldots,d\}$$
$$\alpha(\mathbf{u})=p_0^{-1}\mathbb{G}\left(\mathbb{1}_{\{\boldsymbol{X}_i\leq\mathbf{F}_d^{\leftarrow}(\mathbf{u})\}}\Pi_{j=1}^{d}I_j-C(\mathbf{u})\Pi_{j=1}^{d}I_j\right).$$

We thus retrieve the result of [Segers, 2015] (see example 3.5) in arbitrary dimension.

A Proofs

Proof of Lemma 1 In the proof of Theorem 2 of the online supplementary materials of [Hamori et al., 2019], it is shown that :

$$\sup_{x\in\mathbb{R}}\left|\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\mathbb{1}_{\{X_{i,j}\leq x\}}I_{i,j}}{\hat{\pi}_j(\mathbf{Z}_i)}-F_j(x)\right)-\sqrt{n}\left\{\frac{1}{n}\sum_{i=1}^{n}\psi_j(\mathbf{W}_i,x)\right\}\right|=o_{\mathbb{P}}(1).$$

Also, replacing $\mathbb{1}_{\{X_{i,j}\leq x\}}$ with 1 in the proof of Theorem 2 of *loc. sit.*, we obtain :

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{I_{i,j}}{\hat{\pi}(\mathbf{Z}_i)}-1\right)=o_{\mathbb{P}}(1),$$

which implies that

$$\sqrt{n}\left(n\Big/\sum_{i=1}^{n}\frac{I_{i,j}}{\hat{\pi}(\mathbf{Z}_i)}-1\right)=o_{\mathbb{P}}(1).$$

Given this, we have :

$$\underset{x\in\mathbb{R}}{sup}\left|\sqrt{n}\left(\hat{F}_{n,j}(x)\right)-\sqrt{n}\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{\mathbb{1}_{\{X_{i,j}\leq x\}}}{\hat{\pi}_j(\mathbf{Z}_i)}\right\}\right|=o_{\mathbb{P}}(1).$$

Applying the triangle inequality, we thus have :

$$\sup_{x\in\mathbb{R}}\left|\sqrt{n}\{\hat{F}_{n,j}(x)-F_j(x)\}-\sqrt{n}\left\{\frac{1}{n}\sum_{i=1}^{n}\psi_j(\boldsymbol{W}_i,x)\right\}\right|=o_{\mathbb{P}}(1).$$

Using the same arguments in the proof of Theorem 2 of [Hamori et al., 2019], we can show for the cumulative distribution function of $\mathbf{X}$ that the following holds

$$\underset{\mathbf{x}\in\mathbb{R}^d}{sup}\left|\sqrt{n}\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{\mathbb{1}_{\{\mathbf{X}_i\leq\mathbf{x}\}}\Pi_{j=1}^{d}I_j}{\hat{\eta}(\mathbf{Z})}-F(\mathbf{x})\right\}-\sqrt{n}\left\{\frac{1}{n}\sum_{i=1}^{n}\Psi(\mathbf{W}_i,\mathbf{x})\right\}\right|=o_{\mathbb{P}}(1).$$

Using the same arguments as above gives :

$$\sup_{\mathbf{x}\in\mathbb{R}^d}\left|\sqrt{n}\{\hat{F}_n(\mathbf{x})-F(\mathbf{x})\}-\sqrt{n}\left\{\frac{1}{n}\sum_{i=1}^{n}\Psi(\mathbf{W}_i,\mathbf{x})\right\}\right|=o_{\mathbb{P}}(1).$$

Hence the first statement. Now, the collection of

$$\mathcal{F}=\left\{\Psi(\cdot,\mathbf{x}):\mathbf{x}\in\mathbb{R}^d\right\}\cup\{\psi_j(\cdot,x):x\in\mathbb{R},1\leq j\leq d\}$$

is $\mathbb{P}$-Donsker by Lemma 2 in the space $\ell^\infty(\mathcal{F})$. Remark 2.5 of [Segers, 2015] applies and

$$\mathbb{C}_n^{\mathcal{H}}(\mathbf{u})\rightsquigarrow\alpha(\mathbf{u})-\sum_{j=1}^{d}\dot{C}_j(\mathbf{u})\beta_j(u_j),$$

with α and β describes in the statement. We conclude. $\square$

B Proof of Lemmatas

Lemma 2. *The class of functions* $\{\Psi(\cdot,\mathbf{x}):\mathbf{x}\in\mathbb{R}^d\}$ *and* $\{\psi_j(\cdot,x):x\in\mathbb{R},1\le j\le d\}$ *are* $\mathbb{P}$*-Donsker.*

Proof Let $\{\mathbb{1}_{\{\mathbf{X}\le\mathbf{x}\}}:\mathbf{x}\in\mathbb{R}^d\}$ and this class of functions is well known to be $\mathbb{P}$-Donsker as $\mathbb{1}_{\{\mathbf{X}\le\mathbf{x}\}}$ shatters $\mathbb{R}^d$. By Lemma 3, $\{F(\mathbf{x}|\mathbf{Z}):\mathbf{x}\in\mathbb{R}^d\}$ is also $\mathbb{P}$-Donsker. As $g(\mathbf{W})=(\Pi_{j=1}^d I_j-\eta(\mathbf{Z}))\{\eta(\mathbf{Z})\}^{-1}$ is a uniformly bounded and measurable function as the support of $\mathbf{Z}$ is supposed compact it results that $\{\Psi(\cdot,\mathbf{x}):\mathbf{x}\in\mathbb{R}^d\}$ is $\mathbb{P}$-Donsker (see example 2.10.10 of [van der Vaart and Wellner, 1996]). Using the same arguments state the result for $\{\psi_j(\cdot,x):x\in\mathbb{R},1\le j\le d\}$. $\square$

Lemma 3. *The collection of sets* $\{F_j(x|\boldsymbol{Z}):x\in\mathbb{R},1\le j\le d\}$ *and* $\{F(\mathbf{x}|\boldsymbol{Z}),\mathbf{x}\in\mathbb{R}^d\}$ *are* $\mathbb{P}$*-Donsker.*

Proof Denoting by $\mathbb{P}$ the law of the random vector $(\mathbf{I},\mathbf{X},\mathbf{Z})$. Define the expectation under $\mathbb{P}$, the empirical version and empirical process as follows :

$$\mathbb{P}f=\int f d\mathbb{P},\quad \mathbb{P}_n f=\frac{1}{n}\sum_{i=1}^n f(\mathbf{I}_i,\mathbf{X}_i,\mathbf{Z}_i),\quad \mathbb{G}_n f=\sqrt{n}\left(\mathbb{P}_n-\mathbb{P}\right)f,$$

for any real valued measurable function f. We prove the theorem for $d=1$ and $d=2$. The general case can be treated similarly. Throughout the paper, we write for all $\mathbf{z}\in\mathbb{R}^p$, $\mathbf{x}\in\mathbb{R}^d$

$$\eta_{\mathbf{x}}(\mathbf{z})=F(\mathbf{x}|\mathbf{z}).$$

We want to prove that the class of functions

$$\mathcal{H}\triangleq\left\{\mathbf{z}\mapsto\eta_{\mathbf{x}}(\mathbf{z})\,;\,\mathbf{x}\in\mathbb{R}^d\right\},$$

is $\mathbb{P}$-Donsker, that is, there exists a Gaussian process $\mathbb{G}$ such that

$$\sup_{f\in\mathcal{H}}\{\mathbb{E}|f|(d\mathbb{G}_n-d\mathbb{G})\}\underset{n\to\infty}{\longrightarrow}0.$$

For $d=1$, we consider the class of functions $\mathcal{F}=\{\mathbb{1}_{\{]-\infty,x]\}},x\in\bar{\mathbb{R}}\}\in[0,1]$. We may construct an ϵ-net $V_\epsilon=\{u_1,\dots,u_{N(\epsilon)}\}$ with $|V_\epsilon|\le 3\epsilon^{-1}$. Let us proceed as follows : we select points in V_ϵ so that at step k, we have $u_k\in[0,1]\setminus\cup_{r=1}^{k-1}B(u_r,\epsilon)$. This means that $V_\epsilon=\{u_1,\dots,u_{N(\epsilon)}\}$ is an ϵ-net for $[0,1]$. As $B(u_r,\epsilon)\cap B(u_s,\epsilon)=\emptyset$ for $r,s\in\{1,\dots,N(\epsilon)\}$ with $r\neq s$, we have that $N(\epsilon)\le 3\epsilon^{-1}$.

Let $[0,1]=\bigcup_{r=1}^{K_1(\epsilon)}[l_{\epsilon,r}^{(1)},u_{\epsilon,r}^{(1)}]$ be a ϵ^2-covering of $[0,1]$ by intervals with length $u_{\epsilon,r}^{(1)}-l_{\epsilon,r}^{(1)}$ (for $\mathbb{L}_1(\mathbb{P}_1)$) at most ϵ^2 (using Lemma 4) where $\mathbb{P}_1$ is the marginal law of X_1. Indeed, by the construction detailed above we can pick out $-\infty=x_{1,1}<x_{2,1}<\cdots<x_{K_1(\epsilon),1}=\infty$ and consider $l_{\epsilon,r}^{(1)}=\mathbb{1}_{\{]-\infty,x_{r-1,1}]\}}$ and $u_{\epsilon,r}^{(1)}=\mathbb{1}_{\{]-\infty,x_{r,1}]\}}$ so that $F(x_r)-F(x_{r-1})\leq\epsilon^2$. This can always be done in such a way that $K_1(\epsilon)\leq 1+3/\epsilon^2$.

Now, consider

$$\mathcal{G}=\{[l_{\epsilon,r},u_{\epsilon,r}],r\in 1<r\leq K_1(\epsilon)\},$$

with $l_{\epsilon,r}=\eta_{x_{r-1}}(\mathbf{z})$ and $u_{\epsilon,r}=\eta_{x_r}(\mathbf{z})$. Set $x\in\bar{\mathbb{R}}$, then we can find $x_{r-1},x_r\in\bar{\mathbb{R}}$ such that $x_{r-1}\leq x\leq x_r$ and $\forall\mathbf{z}\in\mathbb{R}^p$ and $\eta_{x_{r-1}}(\mathbf{z})\leq\eta_x(\mathbf{z})\leq\eta_{x_r}(\mathbf{z})$. Furthermore, $||u_{r,\epsilon}-l_{r,\epsilon}||_{\mathbb{P},2}\leq(F(x_{r-1})-F(x_r))^{1/2}\leq\epsilon$ for all $1<r\leq K(\epsilon)$. Thus, the bracketing integral is bounded (see [Kosorok, 2008] page 17), then $\mathcal{H}$ is is $\mathbb{P}$-Donsker.

For $d=2$, denoting by $\mathbb{P}_1$ and $\mathbb{P}_2$ the marginal law X_1 and X_2 respectively, let $[0,1]=\bigcup_{r=1}^{K_1(\epsilon)}[l_{\epsilon,r}^{(1)},u_{\epsilon,r}^{(1)}]$ and $[0,1]=\bigcup_{r=1}^{K_2(\epsilon)}[l_{\epsilon,r}^{(2)},u_{\epsilon,r}^{(2)}]$ be two covering of $[0,1]$ with lengths $u_{\epsilon,r}^{(1)}-l_{\epsilon,r}^{(1)}$ (for $\mathbb{L}_1(\mathbb{P}_1)$) and $u_{\epsilon,r}^{(2)}-l_{\epsilon,r}^{(2)}$ (for $\mathbb{L}_1(\mathbb{P}_2)$) at most ϵ^2. We can find $-\infty=x_{1,1}<x_{2,1}<\cdots<x_{K_1(\epsilon),1}=\infty$ such that $l_{\epsilon,r}^{(1)}=\mathbb{1}_{\{]-\infty,x_{r-1,1}]\}}$ and $u_{\epsilon,r}^{(1)}=\mathbb{1}_{\{]-\infty,x_{r,1}]\}}$ and $||u_{\epsilon,r}^{(1)}-l_{\epsilon,r}^{(1)}||_{\mathbb{P}_1,1}\leq 2^{-1}\epsilon^2$ by the use of Lemma 4. Similarly, we can dispose of $-\infty=x_{1,2}<x_{2,2}<\cdots<x_{K_2(\epsilon),2}=\infty$ and $l_{\epsilon,r}^{(2)}=\mathbb{1}_{\{]-\infty,x_{r-1,2}]\}}$, $u_{\epsilon,r}^{(2)}=\mathbb{1}_{\{]-\infty,x_{r,2}]\}}$ such that $||u_{\epsilon,r}^{(2)}-l_{\epsilon,r}^{(2)}||_{\mathbb{P}_2,1}\leq 2^{-1}\epsilon^2$. By denoting $\mathcal{M}_2(\epsilon)=\{1,\ldots,K_1(\epsilon)\}\times\{1,\ldots,K_2(\epsilon)\}\setminus(1,1)$, let $\mathbf{r}=(r_1,r_2)\in\mathcal{M}_2(\epsilon)$, we now consider the following rectangle :

$$\mathcal{V}_2(\epsilon,\mathbf{r})=[x_{r_1-1,1},x_{r_1,1}]\times[x_{r_2-1,2},x_{r_2,2}]\,.$$

We thus have $\bar{\mathbb{R}}^2=\bigcup_{\mathbf{r}\in\mathcal{M}_2(\epsilon)}\mathcal{V}_2(\epsilon,\mathbf{r})$ and $|\mathcal{M}_2(\epsilon)|\leq(1+6/\epsilon^2)^2$. We obtain that the set

$$\mathcal{G}=\left\{[\eta_{\mathbf{x}_{\mathbf{r}_1}},\eta_{\mathbf{x}_{\mathbf{r}_2}}]\,;\,\mathbf{r}_1,\mathbf{r}_2\in\mathcal{M}_2(\epsilon)\right\}$$

is a covering of ϵ-bracket of the set $\mathcal{H}$, that is for all $\eta\in\mathcal{H}$, there exists $\mathbf{r}_1$ and $\mathbf{r}_2\in\mathcal{M}_2(\epsilon)$ such that

$$\eta_{\mathbf{x}_{\mathbf{r}_1}}\leq\eta_{\mathbf{x}}\leq\eta_{\mathbf{x}_{\mathbf{r}_2}},\quad\text{and}\quad\mathbb{E}\left[(\eta_{\mathbf{x}_{\mathbf{r}_1}}-\eta_{\mathbf{x}_{\mathbf{r}_1}})^2\right]^{1/2}\leq\epsilon. \tag{6}$$

To this aim, set $\mathbf{x}\in\mathbb{R}^2$ and choose $\mathbf{x}_{\mathbf{r}_1}$ and $\mathbf{x}_{\mathbf{r}_2}$ with $\mathbf{r}_1,\mathbf{r}_2\in\mathcal{M}_2(\epsilon)$ such that $\mathbf{x}_{\mathbf{r}_1}\leq\mathbf{x}\leq\mathbf{x}_{\mathbf{r}_2}$ (see Figure 1). Note that, by monotonocity of conditional expectation we have for $\mathbf{z}\in\mathbb{R}^p$,

$$\eta_{\mathbf{x}_{\mathbf{r}_1}}\leq\eta_{\mathbf{x}}\leq\eta_{\mathbf{x}_{\mathbf{r}_2}}.$$

To prove the second statement of (6), observe that

$$
\begin{aligned}
&\mathbb{E}\left[(\eta_{\mathbf{x}_{\mathbf{r}_2}}(\mathbf{Z})-\eta_{\mathbf{x}_{\mathbf{r}_1}}(\mathbf{Z}))^2\right] \leq \mathbb{E}\left[\eta_{\mathbf{x}_{\mathbf{r}_2}}(\mathbf{Z})-\eta_{\mathbf{x}_{\mathbf{r}_1}}(\mathbf{Z})\right] = \mathbb{P}\left\{X \leq \mathbf{x}_{\mathbf{r}_2} \setminus X \leq \mathbf{x}_{\mathbf{r}_1}\right\} \\
&\leq \mathbb{P}\left\{X_1 \in]\mathbf{x}_{\mathbf{r}_1}^{(1)}, \mathbf{x}_{\mathbf{r}_2}^{(1)}]\right\} + \mathbb{P}\left\{X_2 \in]\mathbf{x}_{\mathbf{r}_1}^{(2)}, \mathbf{x}_{\mathbf{r}_2}^{(2)}]\right\} = ||l_{\epsilon,\mathbf{r}_1}^{(1)} - u_{\epsilon,\mathbf{r}_1}^{(1)}||_{\mathbb{P}_1,1} + ||l_{\epsilon,\mathbf{r}_1}^{(2)} - u_{\epsilon,\mathbf{r}_1}^{(2)}||_{\mathbb{P}_2,1} \\
&\leq \frac{\epsilon^2}{2} + \frac{\epsilon^2}{2} = \epsilon^2.
\end{aligned}
$$

Furthermore, the bracketing integral is finite and thus $\mathcal{H}$ is $\mathbb{P}$-Donsker, that's what we wanted to prove.

□

Lemma 4. *Let $(\mathcal{F}, d)$ be a connected bounded metric space and $V = \{u_1, \ldots, u_N\}$ an ϵ-net of $(\mathcal{F}, d)$, then $\forall i \in \{1, \ldots, N\}$, $\exists j \in \{1, \ldots, N\}$ with $j \neq i$ such that*

$$B(u_i, \epsilon) \cap B(u_j, \epsilon) \neq \emptyset.$$

Proof As $\mathcal{F}$ is bounded, we have that $N < \infty$. By absurd, suppose that $\exists i \in \{1, \ldots, N\}$ such that $\forall j \in \{1, \ldots, N\}$ with $j \neq i$ such that

$$B(u_i, \epsilon) \cap B(u_j, \epsilon) = \emptyset.$$

Without loss of generality, take $i = 1$, we then have $\forall j \geq 2$

$$B(u_1, \epsilon) \cap B(u_j, \epsilon) = \emptyset.$$

Note that $E_1 \triangleq B(u_1, \epsilon) \cap \mathcal{F}$ is a closed set as the intersection of two closed ones. Furthermore, $E_2 \triangleq (\cup_{j=1}^{d} B(u_j, \epsilon)) \cap \mathcal{F}$ is also closed. We clearly have

$$E_1 \cup E_2 = \mathcal{F}.$$

We thus have find a partition of $\mathcal{F}$ with two nonempty disjoint closed sets. Hence the contradiction. □

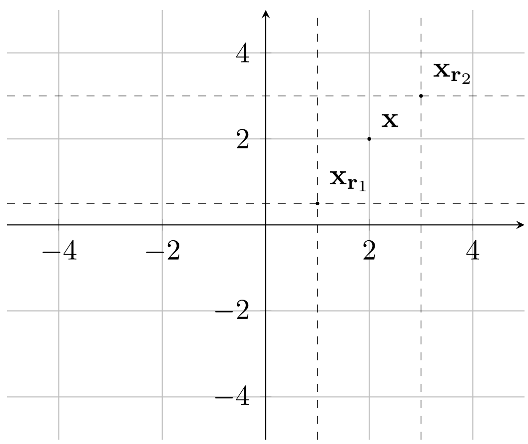

Figure 1: Respective positions of $\mathbf{x}$, $\mathbf{x}_{\mathbf{r}_1}$ and $\mathbf{x}_{\mathbf{r}_2}$ with $d = 2$.

References

[Deheuvels, 1979] Deheuvels, P. (1979). La fonction de dépendance empirique et ses propriétés. un test non paramétrique d'indépendance. *Bulletin Royal Belge de L'Académie des sciences, 65, 274-292.*

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

[Hamori et al., 2019] Hamori, S., Motegi, K., and Zhang, Z. (2019). Calibration estimation of semiparametric copula models with data missing at random. *Journal of Multivariate Analysis*, 173:85–109.

[Hamori et al., 2020] Hamori, S., Motegi, K., and Zhang, Z. (2020). Copula-based regression models with data missing at random. *Journal of Multivariate Analysis*, 180:104654.

[Horvitz and Thompson, 1952] Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47(260):663–685.

[Kosorok, 2008] Kosorok, M. R. (2008). *Introduction to empirical processes and semiparametric inference*. Springer, New York ; London.

[Segers, 2012] Segers, J. (2012). Asymptotics of empirical copula processes under nonrestrictive smoothness assumptions. *Bernoulli*, 18(3):764 – 782.

[Segers, 2015] Segers, J. (2015). Hybrid copula estimators. *J. Statist. Plann. Inference*, 160:23–34.

[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.