Eckhard Liebscher

Semiparametric estimation of the parameters of multivariate copulas

Kybernetika, Vol. 45 (2009), No. 6, 972--991

Persistent URL: `http://dml.cz/dmlcz/140022`

KYBERNETIKA — VOLUME *45* (2009), NUMBER 6, PAGES 972 – 991

SEMIPARAMETRIC ESTIMATION OF THE PARAMETERS OF MULTIVARIATE COPULAS

Eckhard Liebscher

In the paper we investigate properties of maximum pseudo-likelihood estimators for the copula density and minimum distance estimators for the copula. We derive statements on the consistency and the asymptotic normality of the estimators for the parameters.

Keywords: multivariate density estimation, copula, maximum likelihood estimators, minimum distance estimators

AMS Subject Classification: 62H12

1. INTRODUCTION

Let $X = (X^{(1)}, \dots, X^{(d)})^T$ be a d-dimensional random vector. We denote the marginal density and the marginal distribution function of $X^{(m)}$ by f_m and F_m, respectively ($m = 1, \dots, d$). H and h denote the joint distribution function and the joint density of X, respectively. According to Sklar's theorem, we have

$$H(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)) \quad (x_i \in \mathbb{R})$$

where $C : [0,1]^d \to [0,1]$ is the d-dimensional copula. If H is absolutely continuous with density h, then copula C is uniquely determined, and the density h satisfies the following formula

$$h(x_1, \dots, x_d) = \varphi(F_1(x_1), \dots, F_d(x_d))\, f_1(x_1) \, \dots \, f_d(x_d) \tag{1}$$

where $\varphi(u_1, \dots, u_d) = \frac{\partial^d}{\partial u_1 \dots \partial u_d} C(u_1, \dots, u_d)$ is the so-called *copula density.* Concerning the detailed theory of copulas, we refer to the monographs by Joe [12] and by Nelsen [17]. In these monographs the reader also finds surveys of most usable copulas which are symmetric. Who is interested in asymmetric families of copulas may consult the author's paper [16].

In this paper we consider the parametric family $\mathcal{C} = (C_\theta)_{\theta \in \Theta}$ of copulas on $[0,1]^d$. Here $\Theta \subset \mathbb{R}^q$ is the parameter space. The symbol φ_θ denotes the density of C_θ. The aim of this paper is to analyse asymptotic properties of semiparametric estimation procedures for the copula and the copula density. We consider maximum pseudo-likelihood estimators (MPLE) and minimum distance estimators (MDE) where no model for the marginal distribution is needed. The asymptotic behaviour of MPLE has been studied in Genest and Rivest [7], Oakes [18], Genest et al. [5], Shih and Louis [20], and Chen and Fan [1]. The efficiency of such estimators has been analysed in Genest and Werker [8]. An efficient estimation method for parametric classes of copula densities is introduced and investigated in Chen et al. [3]. The asymptotic behaviour of two-stage estimation procedures is studied in Joe [12]. Tsukahara [22] examined minimum distance estimators for the parameters of copulas. The paper by Chen and Fan [2] deals with the asymptotic behaviour of MPLE in the time series framework.

In the present paper we provide statements on the strong consistency and the asymptotic normality of MPLE and MDE for the parameters of the copula. We extend the definition of the estimators in the way that the estimator can be an output of a numerical algorithm solving the corresponding optimisation problem approximately. Because of the complexity of the multivariate distribution, it is frequently not possible to find an appropriate model for a given dataset. In these cases we have to be satisfied with reasonable approximation for the true model. Then we estimate not the true model but an approximation of it, and efficiency of the estimators is not well-defined. This situation is often called misspecification and is covered by our results. The parametric families copulas often include cases of non-identifiable distributions. In this specific situation we do not obtain the consistency of the estimator but a convergence to the set of minimisers of the corresponding nonstochastic optimisation problem. Although we assume that the sample contains i.i.d. random variables, the proof techniques allow a straightforward extension to stationary sequences of dependent random variables. In comparison to other papers on estimation the reader can therefore notice the following points of novelty of the present paper:

- We consider approximate estimators.
- It is not assumed that the underlying copula of the sample items belongs to the parametric family or approximates a member of it.
- The situation of non-identifiability is incorporated.
- The definition of the minimum distance estimators differs significantly from that in the paper Tsukahara [22].

In this paper we focus on the semiparametric estimation of the copula density or the copula based on a parametric model. The advantage of parametric and semi-parametric methods over nonparametric ones is that the latter methods show a bad performance especially in higher dimensions (see Gijbels and Mielniczuk [9]) concerning kernel estimators for copula densities). In the context of parametric copula models the model selection problem arises. Goodness-of-fit tests are studied in the papers by Fermanian [4] and Genest et al. [6] among others. The problem of model selection is discussed in Wang and Wells [23] and Chen and Fan [1].

The problem of estimating copulas appears frequently in the context of the estimation of multivariate distributions. For this reason one can use formula (1) to fit the joint density h to a given sample. Especially in the situation of high dimensions, it is convenient to split the estimation problem into two steps. In our settings it is possible to estimate the marginal densities and the copula density separately. This makes the estimation procedure more tractable for the implementation on the computer. Otherwise it will lead to a high-dimensional estimation problem. The marginal distributions can be fitted using parametric or nonparametric standard methods. Applying nonparametric kernel estimators for the marginal density, we obtain a semiparametric estimator for h which is studied in Hall and Neumayer [10] and Liebscher [15].

The paper is organised as follows: In Section 2 we provide the results concerning the maximum-likelihood estimators. Section 3 is devoted to minimum distance estimators. The reader finds the proof of the results in Sections 4 and 5.

2. ASYMPTOTIC PROPERTIES OF MAXIMUM LIKELIHOOD ESTIMATORS

In this section we consider a family $\mathcal{F} = (\varphi_\theta)_{\theta\in\Theta}$ of models for the copula density with continuous functions $\theta \rightsquigarrow \varphi_\theta(x) = \varphi(x \mid \theta)$ for all $x \in [0,1]^d$. Let $X_1,\ldots,X_n$ with $X_i = (X_i^{(1)},\ldots,X_i^{(d)})^T$ be a sample of random d-dimensional vectors having a distribution with copula density φ. Obviously, $\bar{Y}_l = (F_1(X_l^{(1)}),\ldots,F_d(X_l^{(d)}))^T$ has the joint density φ. Because of the complexity of the multivariate distribution, we do not assume $\varphi \in \mathcal{F}$. Thus the aim is to estimate the vector $\theta_0 \in \Theta$ of parameters maximising

$$\Phi(\theta) = \mathrm{E}\ln\left(\varphi(\bar{\mathrm{Y}}_{\mathrm{i}} \mid \theta)\right) = \int_{[0,1]^{\mathrm{d}}} \ln\left(\varphi(\mathrm{u} \mid \theta)\right)\varphi(\mathrm{u})\,\mathrm{du}.$$

Vector θ_0 represents the vector with the property that $\varphi(\cdot \mid \theta_0)$ approximates best the true copula density φ concerning the Kullback–Leibler divergence. We denote the empirical marginal distribution function of $X_i^{(j)}$ by $\hat{F}_{nj}$. Let $F_{nj} = \frac{n}{n+1}\hat{F}_{nj}$ be the rescaled empirical marginal distribution function. We introduce $Y_{nji} = F_{nj}(X_i^{(j)})$ for $i = 1,\ldots,n, j = 1,\ldots,d$, $Y_{ni} = (Y_{n1i},\ldots,Y_{ndi})^T$, and

$$\Phi_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} \ln\left(\varphi(Y_{ni} \mid \theta)\right). \tag{2}$$

The maximum pseudo-likelihood estimator $\hat{\theta}_n$ (MPLE, sometimes called canonical maximum likelihood estimator) of the parameter θ of the copula is determined by

$$\hat{\theta}_n \in \arg\max_{\theta\in\Theta} \Phi_n(\theta)$$

(see Genest et al. [5]). In many situations an explicit formula for $\hat{\theta}_n$ is not available. In these cases we have to perform a numerical algorithm which gives out an approximate value instead of the exact one. Here the consideration of approximate MPLE can be useful. This kind of estimators $\hat{\theta}_n$ is defined by

$$\Phi_n(\hat{\theta}_n) \geq \max_{\theta\in\Theta} \Phi_n(\theta) - \varepsilon_n, \tag{3}$$

Φ_n as in (2) and $\{\varepsilon_n\}$ is a sequence of random variables with $\varepsilon_n \to 0$ a.s. Hess [11] introduced this kind of estimators and studied it thoroughly. The following two examples show that identifiability problems occur in the context of copulas and cannot be avoided in a simple way without excluding important cases.

Example 1. Cook–Johnson product copula (see Liebscher [16])

$$C(u_1,\dots,u_d) = \left(1+\sum_{i=1}^{d}(u_i^{-\gamma\tau_i}-1)\right)^{-1/\gamma}\left(1+\sum_{i=1}^{d}(u_i^{-\delta(1-\tau_i)}-1)\right)^{-1/\delta}.$$

This copula has $d+2$ parameters: $\gamma,\delta\in[0,+\infty),\tau_1,\dots,\tau_d$, where $\tau_i\in[0,1]$. In the case $\gamma=\delta$ there are models which are not identifiable: parameters $(\gamma,\gamma,\tau_1,\dots,\tau_d)$ and $(\gamma,\gamma,1-\tau_1,\dots,1-\tau_d)$ lead to the same distribution of $\bar{Y}_l$.

Example 2. Convex combination of Cook–Johnson copulas

$$C(u_1,\dots,u_d) = \lambda\left(1+\sum_{i=1}^{d}(u_i^{-\gamma}-1)\right)^{-1/\gamma}+(1-\lambda)\left(1+\sum_{i=1}^{d}(u_i^{-\delta}-1)\right)^{-1/\delta}$$

with parameters $\lambda\in[0,1],\gamma,\delta\in[0,+\infty)$. Here an identifiability problem occurs: parameters $(0,\gamma,\delta)$ and $(1,\delta,\gamma)$ lead to the same distribution of $\bar{Y}_l$ which is the Cook–Johnson copula.

Theorem 2.1 provides the consistency result for the estimator $\hat{\theta}_n$.

Theorem 2.1. Assume that Θ is compact, and $\ln\varphi(\cdot\mid\cdot)$ is continuous on $D\times\Theta$ where $D\subset[0,1]^d$ and the interior of D has Lebesgue measure 1. Suppose that the estimator $\hat{\theta}_n$ satisfies (3), and $F_1,\dots,F_d$ are continuous.

a) Then

$$\lim_{n\to\infty} d(\hat{\theta}_n,\Psi)=0 \quad \text{a.s.}$$

where $\Psi=\arg\max_{\theta\in\Theta}\Phi(\theta)$, $d(x,A)=\inf_{y\in A}\|x-y\|$, $\|.\|$ is the Euclidean norm.

b) If in addition the condition

$$\Phi(\theta)<\Phi(\theta_0) \;\text{ for all } \theta\in\Theta\backslash\{\theta_0\} \tag{4}$$

(i. e. $\Psi=\{\theta_0\}$) is satisfied, then

$$\lim_{n\to\infty}\hat{\theta}_n=\theta_0 \quad \text{a.s.} \tag{5}$$

In the case $\varepsilon_n=0$, strong consistency of $\hat{\theta}_n$ has been proven in the paper by Chen and Fan ([1], Proposition 1), but the proof of Lemma 1(c) is not correct. One can construct a function h such that the assumptions of Lemma 1(c) are fulfilled and the conclusion fails to hold. An assumption on the function h like continuity is missing there. Theorem 2.1 fills the gap in Chen and Fan's paper and incorporates the case of nonidentifiability (part a)) as well as approximate estimators. The property (5) is called minimum Kullback–Leibler information consistency (cf. Suzukawa et al. [21]). The assumption on compactness of Θ may be weakened in some way at a price of higher technical complexity in the proofs. Note that discontinuity of $\ln \varphi(\cdot \mid \theta)$ is allowed here on a set having a closure with Lebesgue measure 0.

In the classical case $\varphi \in \mathcal{F}$; i.e. $\varphi = \varphi(\cdot \mid \theta_0)$ for some $\theta_0 \in \Theta$, the identifiability condition

$$\varphi(\cdot \mid \theta) \neq \varphi(\cdot \mid \theta_0) \text{ for all } \theta \neq \theta_0$$

is sufficient for the assumption (4) which means that θ_0 is the unique maximiser of Φ.

Example 1, continued. Let $\theta_0 = (\gamma_0, \gamma_0, \tau_{01}, \ldots, \tau_{0d})^T$ with $\tau_{0i} \neq 0.5$ for at least one i. Then $\hat{\theta}_{1n} \to \gamma_0$ and $\hat{\theta}_{2n} \to \gamma_0$ a.s. where $\hat{\theta}_n = (\hat{\theta}_{in})_{i=1,\ldots,d+2}$.

In the asymptotic normality results, we need the following assumptions:

Assumption $\mathcal{D}$. The derivatives $\frac{\partial^3}{\partial\theta_i \partial\theta_j \partial\theta_k} \ln \varphi(u \mid \theta)$, $\frac{\partial^3}{\partial\theta_i \partial\theta_j \partial u_l} \ln \varphi(u \mid \theta)$ and $\frac{\partial^3}{\partial\theta_i \partial u_l \partial u_m} \ln \varphi(u \mid \theta)$ exist for $\theta \in \Theta, u \in [0,1]^d$ and all possible i, j, k, l, m. Let $G_{il}(u) = \frac{\partial^2}{\partial\theta_i \partial u_l} \ln \varphi(u \mid \theta)\Big|_{\theta=\theta_0}$ and $\bar{G}_i(u) = \frac{\partial}{\partial\theta_i} \ln \varphi(u \mid \theta)\Big|_{\theta=\theta_0}$. Let θ_0 be an interior point of Θ. There is a function $M : [0,1]^d \to [0,\infty)$, an $\bar{\varepsilon} > 0$ and a neighbourhood $U(\theta_0) \subset \Theta$ of θ_0 such that

$$\begin{aligned}
|G_{il}(u)| \leq M(u), \quad \bar{G}_i^2(u) &\leq M(u), \\
\sup_{\theta \in U(\theta_0)} \sup_{v:\|u-v\|<\bar{\varepsilon}} \left| \frac{\partial^3}{\partial\theta_i \partial\theta_j \partial\theta_k} \ln \varphi(v \mid \theta) \right| &\leq M(u), \\
\sup_{v:\|u-v\|<\bar{\varepsilon}} \left| \frac{\partial^3}{\partial\theta_i \partial\theta_j \partial u_l} \ln \varphi(v \mid \theta_0) \right| &\leq M(u), \\
\sup_{v:\|u-v\|<\bar{\varepsilon}} \left| \frac{\partial^3}{\partial\theta_i \partial u_l \partial u_m} \ln \varphi(v \mid \theta_0) \right| &\leq M(u)
\end{aligned}$$

for $i, j, k = 1, \ldots, q$, $l, m = 1, \ldots, d$, $u \in [0,1]^d$ and $\mathrm{E}\, M(\bar{Y}_l) < +\infty$. Moreover, there is a positively definite matrix $I(\theta_0) = (I_{ij}(\theta_0))_{i,j=1,\ldots,q}$ such that

$$\mathrm{E}\ \frac{\partial^2}{\partial\theta_i \partial\theta_j} \ln \varphi(\bar{Y}_l \mid \theta)\Big|_{\theta=\theta_0} = -I_{ij}(\theta_0).$$

In the case $\varphi = \varphi(\cdot \mid \theta_0)$, this matrix $I(\theta_0)$ is usually called the information matrix. □

Now we give the asymptotic normality result for estimators $\hat{\theta}_n$.

Theorem 2.2. Assume that $\varepsilon_n = o(n^{-1})$. Let the assumptions of Theorem 2.1 b) be satisfied. Then, under condition $\mathcal{D}$, we have

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma).$$

Here $\Sigma = I(\theta_0)^{-1}\Sigma_1 I(\theta_0)^{-1}$, and

$$\begin{aligned} \Sigma_1 &= \left(\operatorname{cov}\left(\Gamma_i(X_{11}, \dots, X_{d1}), \Gamma_k(X_{11}, \dots, X_{d1})\right)\right)_{i,k=1,\dots,q}, \\ \Gamma_i(z_1, \dots, z_d) &= \iint\limits_{[0,1]^d} \sum_{j=1}^{d} G_{ij}(t)\mathbf{1}(F_j(z_j) \le t_j)\,\mathrm{d}C(t) + \bar{G}_i(F_1(z_1), \dots, F_d(z_d)). \end{aligned}$$

This result was provided in Chen and Fan ([1], Proposition 2) for $\varepsilon_n = 0$, but the proof uses Lemma 1(c) which is problematic as mentioned above. Similarly to Genest et al. [5], one can use the following estimator for Σ_1:

$$\begin{aligned} \left(\hat{\Sigma}_1\right)_{i,j} &= \frac{1}{n-1} \sum_{k,l=1}^{n} \left(\hat{\Gamma}_i(Y_{nk}) - \bar{\Gamma}_i\right)\left(\hat{\Gamma}_j(Y_{nl}) - \bar{\Gamma}_j\right), \\ \hat{\Gamma}_i(z_1, \dots, z_d) &= \frac{1}{n} \sum_{k=1}^{n} \sum_{j=1}^{d} G_{ij}(Y_{nk} \mid \hat{\theta}_n)\, I(z_j \le Y_{njk}) + \bar{G}_i(F_1(z_1), \dots, F_d(z_d)). \end{aligned}$$

3. MINIMUM DISTANCE ESTIMATORS

Let $X_1, \dots, X_n$ be the sample of random vectors as in the previous section. Consider the family $\overset{\circ}{\mathcal{F}} = (C_\theta)_{\theta \in \Theta}$ of copulas where $\theta \rightsquigarrow C_\theta = C(u \mid \theta)$ is continuous for all $u \in [0,1]^d$. We denote the marginal empirical joint distribution function by $\hat{H}_n$. Let $F_n(x) = (F_{1n}(x_1), \dots, F_{dn}(x_d))^T$ with F_{nj} as in the previous section, $F(x) = (F_1(x_1), \dots, F_d(x_d))^T$ for $x = (x_1, \dots, x_d)^T \in \mathbb{R}^d$, and

$$\begin{aligned} \Phi_n(\theta) &= \frac{1}{n} \sum_{i=1}^{n} \left(\hat{H}_n(X_i) - C(F_n(X_i) \mid \theta)\right)^2 \\ &= \int_{\mathbb{R}^d} \left(\hat{H}_n(x) - C(F_n(x) \mid \theta)\right)^2 \mathrm{d}\hat{H}_n(x). \end{aligned}$$

Further we introduce

$$\Phi(\theta) = \int_{\mathbb{R}^d} \left(H(x) - C(F(x) \mid \theta)\right)^2 \mathrm{d}H(x).$$

The aim is to estimate the vector θ_0 which minimises Φ. In this section we want to study minimum distance estimators (MDE) $\hat{\theta}_n$ satisfying

$$\Phi_n(\hat{\theta}_n) \le \min_{\theta \in \Theta} \Phi_n(\theta) + \varepsilon_n,$$

where $\{\varepsilon_n\}$ is a sequence of random variables with $\varepsilon_n \to 0$ a.s. This estimator $\hat{\theta}_n$ is not the exact minimiser of Φ_n but an approximate minimiser. The definition of the

MDE differs significantly from that in the paper by Tsukahara [22] in the respect that the calculation of multiple integrals is avoided. Tsukahara's [22] estimator is given by

$$\breve{\theta} = \arg\min_{\theta\in\Theta} \int_{[0,1]^d} \left(C_n(u) - C(u,\theta)\right)^2 \, \mathrm{d}u.$$

The following theorem provides the result about consistency of the MDE.

Theorem 3.1. Let Φ be continuous.

a) Then

$$\lim_{n\to\infty} d(\hat{\theta}_n, \Psi) = 0 \quad \text{a.s.},$$

where $\Psi = \arg\min_{\theta\in\Theta} \Phi(\theta)$, $d(\cdot,\cdot)$ as in Theorem 2.1.

b) If in addition the condition

$$\Phi(\theta) > \Phi(\theta_0) \;\; \text{for all } \theta \in \Theta \backslash \{\theta_0\} \tag{6}$$

(i. e. $\Psi = \{\theta_0\}$) is satisfied, then

$$\lim_{n\to\infty} \hat{\theta}_n = \theta_0 \quad \text{a.s.}$$

If $C = C(\cdot \mid \theta_0) \in \mathcal{F}^{\circ}$ is continuous, then the identifiability condition

$$C(\cdot \mid \theta) \neq C(\cdot \mid \theta_0) \text{ for all } \theta \neq \theta_0$$

is sufficient for the assumption (6). The next Theorem 3 states that $\hat{\theta}_n$ is asymptotically normally distributed in the case $\Psi = \{\theta_0\}$. The following assumption on partial derivatives of the copula is needed in this theorem.

Assumption $\mathcal{A}$. $\bar{C}_k(\cdot \mid \cdot), \bar{C}_{kl}(\cdot \mid \cdot), \tilde{C}_j(\cdot \mid \cdot), \tilde{C}_{jk}(\cdot \mid \cdot)$ denote the partial derivatives $\frac{\partial}{\partial\theta_k}C(. \mid \theta)$, $\frac{\partial^2}{\partial\theta_j\partial\theta_k}C(\cdot \mid \theta)$, $\frac{\partial}{\partial u_j}C(u \mid \cdot)$, $\frac{\partial^2}{\partial\theta_k\partial u_j}C(u \mid \theta)$, respectively. We assume that these derivatives exist, and for $k,l = 1,\ldots,q, j = 1,\ldots,d$, the functions $(u,t) \rightsquigarrow \bar{C}_{kl}(u \mid t)$, $(u,t) \rightsquigarrow \tilde{C}_{jk}(u \mid t)$ are continuous on $[0,1]^d \times U(\theta_0)$, where $U(\theta_0) \subset \Theta$ is a neighbourhood of θ_0. θ_0 is an interior point of Θ.

Theorem 3.2. Assume that $\varepsilon_n = o(n^{-1})$, Assumption $\mathcal{A}$ and the assumptions of Theorem 3.1 b) are satisfied. Then

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0,\Sigma).$$

Here $\Sigma = \Sigma_2^{-1}\Sigma_1\Sigma_2^{-1}$, $\Sigma_1 = \operatorname{cov}(Z_i)$, $Z_i = (Z_{ij})_{j=1,\ldots,q}$,

$$\gamma_{kj}(x \mid \theta_0) = (H(x) - C(F(x) \mid \theta_0))\,\tilde{C}_{jk}\,(F(x) \mid \theta_0) - \tilde{C}_j\,(F(x) \mid \theta_0)\,\bar{C}_k(F(x) \mid \theta_0),$$

$$\begin{aligned} Z_{ij} &= \sum_{j=1}^{d} \int_{\mathbb{R}^d} I\left(X_1^{(j)} \le x_j\right) \gamma_{kj}(x \mid \theta_0)\,\mathrm{d}H(x) + \int_{\mathbb{R}^d} I\left(X_1 \le x\right) \bar{C}_k(F(x) \mid \theta_0)\,\mathrm{d}H(x) \\ &\quad + \left(H(X_1) - C(F(X_1) \mid \theta_0)\right) \bar{C}_k(F(X_1) \mid \theta_0), \end{aligned}$$

$\Sigma_2 = (\mathcal{H}_{ij})_{i,j=1,\dots,q}$,

$$\begin{aligned} \mathcal{H}_{ij} &= \int_{\mathbb{R}^d} \big((H(x) - C(F(x) \mid \theta_0))\, \bar{C}_{ij}(F(x) \mid \theta_0) \\ &\quad + \bar{C}_i(F(x) \mid \theta_0) \bar{C}_j(F(x) \mid \theta_0) \big)\ \mathrm{d}H(x). \end{aligned}$$

Tsukahara [22] proved consistency and asymptotic normality for the estimator $\check{\theta}$ in the case where the copula C of X_i belongs to a small neighbourhood of a member of the parametric family. Since the covariance structure of the estimator is rather complicated and the covariances are hard to estimate directly, it is recommended to use alternative techniques like bootstrap to estimate the covariances.

4. PROOFS OF THE RESULTS OF SECTION 2

Let

$$\bar{\Phi}_n(\theta) = \int_E \psi(t, x)\,\mathrm{d}P_n(x), \quad \bar{\Phi}(t) = \int_E \psi(t, x)\,\mathrm{d}P(x)$$

for $t \in \Theta$ with a measurable function $\psi : \Theta \times E \to \mathbb{R}$. We assume that these Lebesgue integrals exist. P_n and P are a random and respective a nonrandom probability measure on E. Let $\hat{\theta}_n$ be an estimator satisfying

$$\bar{\Phi}_n(\hat{\theta}_n) \le \min_{\theta \in \Theta} \bar{\Phi}_n(\theta) + \varepsilon_n$$

and $\{\varepsilon_n\}$ is a sequence of random variables with $\varepsilon_n \to 0$ a.s. The following theorem is a direct consequence of Theorem 2.2 in Lachout et al. [14].

Proposition 4.1. Suppose that Θ is a compact set, and for every θ, function $\psi(\cdot, x)$ is lower semicontinuous at θ for all x except a set V_θ of P-measure zero. Assume that $\int_E \inf_{t \in \Theta} \psi(t, x)\,\mathrm{d}P(x) > -\infty$, $\int_E \psi(\theta, x)\,\mathrm{d}P(x) < +\infty$ for all $\theta \in \Theta$. Let $B(\theta, R) = \{t \in \Theta : \|t - \theta\| \le R\}$. Assume also that for any $\theta \in \Theta \backslash \{\theta_0\}, R > 0$,

$$\liminf_{n \to \infty} \int_E \inf_{t \in B(\theta, R)} \psi(t, x)\,\mathrm{d}P_n(x) \ \ge \ \int_E \inf_{t \in B(\theta, R)} \psi(t, x)\,\mathrm{d}P(x) \quad \text{a.s.}, \tag{7}$$

$$\limsup_{n \to \infty} \int_E \psi(\theta_0, x)\,\mathrm{d}P_n(x) \ \le \ \int_E \psi(\theta_0, x)\,\mathrm{d}P(x) \quad \text{a.s.} \tag{8}$$

(a) Then

$$\lim_{n \to \infty} d(\hat{\theta}_n, \Psi) = 0 \quad \text{a.s.}$$

where $\Psi = \arg\min_{\theta\in\Theta} \int_E \psi(\theta,x)\,\mathrm{d}P(x)$, $d(\cdot,\cdot)$ as above.

(b) Moreover, if in addition, $\bar\Phi(\theta) > \bar\Phi(\theta_0)$ holds for all $\theta \in \Theta\backslash\{\theta_0\}$, then

$$\lim_{n\to\infty} \hat\theta_n = \theta_0 \quad \text{a.s.}$$

Throughout the remainder of this section, let the settings of Section 2 be valid.

Proof of Theorem 2.1. Here P_n is a discrete measure with $P_n(\{Y_{ni}\}) = n^{-1}$, $Y_{ni} = (F_{n1}(X_i^{(1)}),\dots,F_{nd}(X_i^{(d)}))^T)$ and $P_n([0,1]^d) = 1$. Moreover, P is the distribution measure of $\bar Y_i = (F_1(X_i^{(1)}),\dots,F_d(X_i^{(d)}))^T$. Let $\psi(t,x) = -\ln(\varphi(x\mid t))$. By the Glivenko–Cantelli theorem, we have

$$\sup_{z\in\mathbb{R}} |F_{nj}(z) - F_j(z)| \to 0 \quad \text{for } j = 1\dots d, \omega\in\Omega^*, \mathrm{P}(\Omega^*) = 1. \tag{9}$$

Let $\varepsilon, R > 0$ and $\theta \in \Theta\backslash\{\theta_0\}$. We use the notation $\tilde D := D\backslash[0,1]^d$ and $E_\eta = \{x \in [0,1]^d : \exists y \in E : \|x-y\| \le \eta\}$. By the continuity of the Lebesgue measure there exist an open set $A \subset D$ such that $\mathrm{P\{\bar Y_i \in A^c\}} \le \varepsilon$ with $A^c := [0,1]^d\backslash A$, and $\ln\varphi$ is continuous on the closed set $A_\eta \times \Theta$ for some $\eta > 0$. In the following we show that the assumptions of Proposition 4.1 are satisfied. Now $\ln\varphi$ is uniformly continuous on $A_1 \times \Theta$ and $\sup_{x\in[0,1]^d} \sup_{\theta\in\Theta} \ln(\varphi(x\mid\theta)) = C_1 < +\infty$. Now there is a $\delta : 0 < \delta < \eta$ such that

$$\begin{aligned}
\sup_{t\in B(\theta,R)} \ln(\varphi(u_1\mid t)) &< \ln(\varphi(u_1\mid t_0(u_1))) + \frac{\varepsilon}{2}\\
&< \ln(\varphi(u_2\mid t_0(u_1))) + \varepsilon\\
&\le \sup_{t\in B(\theta,R)} \ln(\varphi(u_2\mid t)) + \varepsilon, \quad t_0(u_1)\in B(\theta,R)
\end{aligned}$$

for all $u_1 \in [0,1]^d, u_2 \in A$ with $\|u_1-u_2\| < \delta$. Hence, by the strong law of large numbers and (9),

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \sup_{t\in B(\theta,R)} \ln(\varphi(Y_{ni}\mid t))\\
\le\ &\limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \sup_{t\in B(\theta,R)} \ln(\varphi(Y_{ni}\mid t))\, I(\bar Y_i \in A)\\
&+C_1 \limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^n I(\bar Y_i \in A^c)\\
\le\ &\limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \sup_{t\in B(\theta,R)} \ln\big(\varphi(\bar Y_i\mid t)\big)\, I(\bar Y_i \in A)\\
&+C_1 \mathrm{P}\{\bar Y_i \in A^c\} + \varepsilon\\
\le\ &\mathrm{E} \sup_{t\in B(\theta,R)} \ln\big(\varphi(\bar Y_i\mid t)\big) + \varepsilon(C_1+1)
\end{aligned}$$

for $n \geq n_0(\omega)$, $\omega \in \Omega^{**} \subset \Omega^*$, $\mathrm{P}(\Omega^{**}) = 1$. Now we obtain (7) by letting $\varepsilon \to 0$. Inequality (8) can be shown in a similar way. Theorem 2.1 follows now from Proposition 4.1. $\square$

Next we prove asymptotic normality of the MLE. In this proof we need the asymptotic normality of the random vector $\nabla_\theta \Phi_n(\theta_0) = \left(\frac{1}{n}\sum_{\ell=1}^n \frac{\partial}{\partial\theta_i} \ln\left(\varphi(Y_{nl} \mid \theta)\right)\right)_{i=1,\dots,q}$ which will be shown in the next lemma.

Lemma 4.1. Under the conditions of Theorem 2.2, we have

$$\sqrt{n}\nabla_\theta \Phi_n(\theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma_1)$$

with the covariance matrix Σ_1 introduced in Theorem 2.2.

P r o o f. Remember that $\bar{G}_i(u) = \frac{\partial}{\partial\theta_i} \ln\left(\varphi(u \mid \theta)\right)\Big|_{\theta=\theta_0}$. Obviously, we have

$$\mathrm{E}\,\bar{G}_i(\bar{Y}_l) = \int \frac{\partial}{\partial\theta_i} \ln\left(\varphi(y \mid \theta)\right)\varphi(y)\,\mathrm{d}y\Big|_{\theta=\theta_0} = \frac{\partial}{\partial\theta_i}\Phi(\theta_0) = 0.$$

We obtain

$$\sqrt{n}\frac{\partial}{\partial\theta_i}\Phi_n(\theta_0) = \frac{1}{\sqrt{n}}\sum_{l=1}^n \frac{\partial}{\partial\theta_i} \ln\left(\varphi(Y_{nl} \mid \theta)\right)\Big|_{\theta=\theta_0} = A_{ni} + B_{ni} \quad \text{for } i = 1,\dots,q,$$

where

$$A_{ni} = \frac{1}{\sqrt{n}}\sum_{l=1}^n \left(\bar{G}_i(Y_{nl}) - \bar{G}_i(\bar{Y}_l)\right), \quad B_{ni} = \frac{1}{\sqrt{n}}\sum_{l=1}^n \bar{G}_i(\bar{Y}_l).$$

By the strong law of large numbers,

$$\frac{1}{n^\delta}\sum_{l=1}^n \left|M(\bar{Y}_l)\right| \to 0 \quad \text{a.s. for } \delta > 1.$$

Thus

$$\begin{aligned} A_{ni} &= \frac{1}{\sqrt{n}}\sum_{l=1}^n\sum_{j=1}^d G_{ij}(\bar{Y}_l)\left(F_{nj}(X_l^{(j)}) - F_j(X_l^{(j)})\right) \\ &\quad + \frac{1}{\sqrt{n}}\sum_{l=1}^n\sum_{j=1}^d \left(G_{ij}(Y_{nl}^*) - G_{ij}(\bar{Y}_l)\right)\left(F_{nj}(X_l^{(j)}) - F_j(X_l^{(j)})\right) \\ &= \frac{1}{n\sqrt{n}}\sum_{j=1}^d\sum_{l=1}^n\sum_{k=1}^n G_{ij}(\bar{Y}_l)\left(I(X_k^{(j)} \leq X_l^{(j)}) - F_j(X_l^{(j)})\right) + o(1) \quad \text{a.s.,} \end{aligned}$$

where $G_{ij}(u) = \frac{\partial^2}{\partial\theta_i \partial u_j} \ln\left(\varphi(u \mid \theta)\right)\Big|_{\theta=\theta_0}$ and $Y^*_{nl} = \bar{Y}_l + \eta(Y_{nl} - \bar{Y}_l)$, $\eta \in (0,1)$. Now we consider the convergence in distribution of

$$\begin{aligned}
D_n &= \sqrt{n} \sum_{i=1}^{q} \lambda_i \frac{\partial}{\partial\theta_i} \Phi_n(\theta_0) \\
&= \frac{1}{n\sqrt{n}} \sum_{i=1}^{q} \lambda_i \left(\sum_{j=1}^{d} \sum_{l=1}^{n} \sum_{k=1}^{n} G_{ij}(\bar{Y}_l) \left(I(X_k^{(j)} \le X_l^{(j)}) - F_j(X_l^{(j)}) \right) \right. \\
&\qquad \left. + (n-1) \sum_{l=1}^{n} \left(\bar{G}_i(\bar{Y}_l) - \mathrm{E}\, \bar{G}_i(\bar{Y}_l) \right) \right) + o(1) \\
&= \frac{1}{n\sqrt{n}} \sum_{l=1}^{n-1} \sum_{k=l+1}^{n} \kappa(X_k, X_l) + o(1) \quad \text{a.s.} \quad \text{for } \lambda_i \in \mathbb{R},
\end{aligned}$$

where

$$\begin{aligned}
\kappa(\bar{x}_1, \bar{x}_2) &= \sum_{i=1}^{q} \lambda_i \left(\sum_{j=1}^{d} \left(G_{ij}(F_1(\bar{x}_{12}), \ldots, F_d(\bar{x}_{d2})) \left(\mathbf{1}(\bar{x}_{j1} \le \bar{x}_{j2}) - F_j(\bar{x}_{j2}) \right) \right. \right. \\
&\qquad \left. + G_{ij}(F_1(\bar{x}_{11}), \ldots, F_d(\bar{x}_{d1})) \left(\mathbf{1}(\bar{x}_{j2} \le \bar{x}_{j1}) - F_j(\bar{x}_{j1}) \right) \right) \\
&\qquad + \bar{G}_i(F_1(\bar{x}_{11}), \ldots, F_d(\bar{x}_{d1})) - \mathrm{E}\, \bar{G}_i(\bar{Y}_1) \\
&\qquad \left. + \bar{G}_i(F_1(\bar{x}_{12}), \ldots, F_d(\bar{x}_{d2})) - \mathrm{E}\, \bar{G}_i(\bar{Y}_1) \right)
\end{aligned}$$

and $\bar{x}_k = (\bar{x}_{1k}, \ldots, \bar{x}_{dk})^T$. Obviously, κ is symmetric,

$$\begin{aligned}
\mathrm{E}\, \kappa(\bar{x}_1, X_1) &= \sum_{i=1}^{d} \lambda_i \left(\int_{[0,1]^d} \sum_{j=1}^{d} \left(G_{ij}(t) \left(\mathbf{1}(F_j(\bar{x}_{j1}) \le t_j) - t_j \right) \right) \mathrm{d}C(t) \right. \\
&\qquad \left. + \bar{G}_i(F_1(\bar{x}_{11}), \ldots, F_d(\bar{x}_{d1})) - \mathrm{E}\, \bar{G}_i(\bar{Y}_1) \right)
\end{aligned}$$

and $\mathrm{E}\, \kappa(X_1, X_2) = 0$. Applying a central limit theorem for U-statistics (cf. Theorem A in Serfling [19], p. 192) and the Cramér–Wold device, we obtain the asymptotic normality of $\sqrt{n} \nabla_\theta \Phi_n(\theta_0)$. □

Proof of Theorem 2.2. Let $\tilde{\theta}_n = \arg\max_{\theta \in \Theta} \Phi_n(\theta)$. The Taylor expansion leads immediately to

$$\tilde{\theta}_n - \hat{\theta}_n = o_{\mathrm{P}}(n^{-1/2}). \tag{10}$$

We define the $q \times q$-matrix

$$W_n(\theta) = \frac{1}{n}\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j}\sum_{l=1}^{n}\ln\left(\varphi(\bar{Y}_l \mid \theta)\right)\right)_{i,j=1,\dots,q}$$

with $\bar{Y}_l$ as above. By the definition of the estimator, $\nabla_\theta\Phi_n(\tilde{\theta}_n) = 0$. Using the mean value theorem, we obtain

$$-\nabla_\theta\Phi_n(\theta_0) = W_n^*\left(\tilde{\theta}_n - \theta_0\right),$$

where $W_n^* = \frac{1}{n}\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j}\sum_{l=1}^{n}\ln\left(\varphi(Y_{nl} \mid \theta)\right)\Big|_{\theta=\theta^*_{in}}\right)_{i,j=1,\dots,q}$. Here, $\theta^*_{n1},\dots,\theta^*_{nq}$ are random variables with $\theta^*_{nj} = \theta_0 + (\tilde{\theta}_n - \theta_0)\eta_j$, $\eta_j \in (0,1)$. By Theorem 2.1, $\theta^*_{nj} \to \theta_0$ a.s.for all j. Further

$$\sqrt{n}\left(\tilde{\theta}_n - \theta_0\right) = -W_n^{*-1}\left(\sqrt{n}\,\nabla_\theta\Phi_n(\theta_0)\right). \tag{11}$$

In the sequel we show that

$$W_n^* - W_n(\theta_0) \xrightarrow{\mathrm{P}} 0 \quad \text{and} \tag{12}$$

$$W_n^* \xrightarrow{\mathrm{P}} -I(\theta_0). \tag{13}$$

The quantity $(W_n^*)_{ij}$ denotes the entry of matrix W_n^* in the ith row and the jth column. Let $\varepsilon > 0$ and $i,j \in \{1,\dots,q\}$ arbitrary. Applying the triangle inequality, we can derive

$$\begin{aligned}
&|(W_n^*)_{ij} - (W_n(\theta_0))_{ij}| \\
\le\;& \frac{1}{n}\sum_{l=1}^{n}\left|\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ln\left(\varphi(Y_{nl} \mid \theta)\right)\Big|_{\theta=\theta^*_{ni}} - \frac{\partial^2}{\partial\theta_i\partial\theta_j}\ln\left(\varphi(Y_{nl} \mid \theta)\right)\Big|_{\theta=\theta_0}\right)\right| \\
&+\frac{1}{n}\sum_{l=1}^{n}\left|\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ln\left(\varphi(Y_{nl} \mid \theta)\right)\Big|_{\theta=\theta_0} - \frac{\partial^2}{\partial\theta_i\partial\theta_j}\ln\left(\varphi(\bar{Y}_l \mid \theta)\right)\Big|_{\theta=\theta_0}\right)\right| \\
\le\;& \sqrt{d}\,\frac{1}{n}\sum_{l=1}^{n}H(\bar{Y}_l)\,\|\theta^*_{ni} - \theta_0\| + \sqrt{d}\,\frac{1}{n}\sum_{l=1}^{n}H(\bar{Y}_l)\,\|Y_{nl} - \bar{Y}_l\|.
\end{aligned}$$

The right hand side of this inequality tends to zero almost surely in view of (9). Hence (12) is valid. We apply the law of large numbers to obtain $W_n(\theta_0) \xrightarrow{\mathrm{P}} -I(\theta_0)$. Together with (12), it follows that (13) holds true. Since $I(\theta_0)$ is positive definite, its inverse matrix $I(\theta_0)^{-1}$ exists and we have

$$W_n^{*\,-1} \xrightarrow{\mathrm{P}} -I(\theta_0)^{-1}. \tag{14}$$

Let $Y \sim \mathcal{N}(0,\Sigma_1)$. Combining Lemma 4.1, (10), (11) and (14), we obtain

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow{\mathcal{D}} -I(\theta_0)^{-1}Y$$

which completes the proof. □

5. PROOFS OF THE RESULTS OF SECTION 3

Let $\bar{\Phi}_n : \Theta \to \mathbb{R}$, and $\hat{\theta}_n$ be an estimator satisfying

$$\bar{\Phi}_n(\hat{\theta}_n) \leq \min_{\theta \in \Theta} \bar{\Phi}_n(\theta) + \varepsilon_n.$$

Here $\{\varepsilon_n\}$ is a sequence of random variables with $\varepsilon_n \to 0$ a.s. Theorem 2.2 of the paper Lachout et al. [14] leads to the following proposition.

Proposition 5.1. Assume that Θ is compact, and $\lim_{n\to\infty} \sup_{t\in\Theta} \left|\bar{\Phi}_n(t) - \bar{\Phi}(t)\right| = 0$ a.s. holds for a continuous function $\bar{\Phi}$.

(a) Then

$$\lim_{n\to\infty} d(\hat{\theta}_n, \Psi) = 0 \quad \text{a.s.},$$

where $\Psi = \arg\min_{t\in\Theta} \bar{\Phi}(t)$, $d(\cdot,\cdot)$ as above.

(b) Moreover, if in addition, $\bar{\Phi}(\theta) > \bar{\Phi}(\theta_0)$ holds for all $\theta \in \Theta\backslash\{\theta_0\}$, then

$$\lim_{n\to\infty} \hat{\theta}_n = \theta_0 \quad \text{a.s.}$$

Let $\bar{\Phi}_n = \Phi_n$ and $\bar{\Phi} = \Phi$ with Φ_n and Φ as in Section 3. The following lemma concerns the justification of the assumptions of Proposition 5.1.

Lemma 5.2. Assume that Θ is compact. Then

$$\lim_{n\to\infty} \sup_{t\in\Theta} |\Phi_n(t) - \Phi(t)| = 0 \text{ a.s.}$$

Proof. Utilising the Lipschitz continuity of copulas, we obtain

$$\begin{aligned}
&\sup_{t\in\Theta} |\Phi_n(t) - \Phi(t)| \\
\leq\ & \sup_{t\in\Theta} \int_{\mathbb{R}^d} \Big(\hat{H}_n(x) + H(x) + C(F_n(x) \mid t) + C(F(x) \mid t)\Big) \\
&\qquad \Big|\hat{H}_n(x) - H(x) - C(F_n(x) \mid t) + C(F(x) \mid t)\Big| \,\mathrm{d}\hat{H}_n(x) + R_n \\
\leq\ & 4 \sup_{x\in\mathbb{R}^d} \Big|\hat{H}_n(x) - H(x)\Big| + 4 \sup_{t\in\Theta, x\in\mathbb{R}^d} |C(F_n(x) \mid t) - C(F(x) \mid t)| + R_n \\
\leq\ & 4 \sup_{x\in\mathbb{R}^d} \Big|\hat{H}_n(x) - H(x)\Big| + 4 \sum_{i=1}^{d} \sup_{x\in\mathbb{R}^d} |F_{ni}(x) - F_i(x)| + R_n,
\end{aligned}$$

where $R_n = \sup_{t\in\Theta} \left| \int_{\mathbb{R}^d} (H(x) - C(F(x) \mid t))^2 \, \mathrm{d}(\hat{H}_n(x) - H(x)) \right|$. By the strong uniform ergodic theorem,

$$\begin{aligned} \sup_{x\in\mathbb{R}^d} \left| \hat{H}_n(x) - H(x) \right| &\to 0, \quad \sup_{x\in\mathbb{R}^d} |F_{ni}(x) - F_i(x)| \to 0, \\ R_n &\to 0 \ \text{ a.s.} \end{aligned}$$

This completes the proof. □

Proof of Theorem 3.1. Theorem 3.1 is a direct consequence of Proposition 5.1 and Lemma 5.2. □

Let $\tilde{\theta}_n = \arg\min_{\theta\in\Theta} \Phi_n(\theta)$. Throughout the remainder of this section we suppose that Assumption $\mathcal{A}$ is satisfied. Since $\nabla_\theta f_n(\tilde{\theta}_n) = 0$, we can derive

$$\nabla_\theta f_n(\theta_0) = -\mathcal{H}_n(t^*) \left(\tilde{\theta}_n - \theta_0 \right), \tag{15}$$

where $t^* = \theta_0 + \eta \left(\tilde{\theta}_n - \theta_0 \right), \eta \in (0,1)$, $\mathcal{H}_n(\theta) = (\mathcal{H}_{nij}(\theta))_{i,j=1,\ldots,q}$,

$$\begin{aligned} \nabla_\theta f_n(\theta) &= 2 \int_{\mathbb{R}^d} \left(\hat{H}_n(x) - C(F_n(x) \mid \theta) \right) \nabla_\theta C(F_n(x) \mid \theta) \, \mathrm{d}\hat{H}_n(x), \\ \mathcal{H}_{nij}(\theta) &= 2 \int_{\mathbb{R}^d} \left(\hat{H}_n(x) - C(F_n(x) \mid \theta) \right) \bar{C}_{ij}(F_n(x) \mid \theta) \, \mathrm{d}\hat{H}_n(x) \\ &\quad + 2 \int_{\mathbb{R}^d} \bar{C}_i(F_n(x) \mid \theta) \bar{C}_j(F_n(x) \mid \theta) \, \mathrm{d}\hat{H}_n(x). \end{aligned}$$

Next we show the asymptotic normality of $\nabla_\theta f_n(\theta_0)$ and that $\mathcal{H}_n(t^*)$ converges in probability to a certain matrix. The following four lemmas are used in the proof of asymptotic normality.

Lemma 5.3. We have

$$Q_{nk} := \sqrt{n} \int_{\mathbb{R}^d} \left(\hat{H}_n(x) - H(x) \right) \bar{C}_k(F(x) \mid \theta_0) \, \mathrm{d} \left(\hat{H}_n(x) - H(x) \right) \xrightarrow{\mathrm{P}} 0$$

for $k = 1, \ldots, q$.

Proof. Let

$$\begin{aligned} \kappa(x,y) &= (\mathbf{1}(y \le x) - H(x)) \, \bar{C}_k(F(x) \mid \theta_0) \\ &\quad - \int_{\mathbb{R}^d} (\mathbf{1}(y \le z) - H(z)) \, \bar{C}_k(F(z) \mid \theta_0) \, \mathrm{d}H(z) \, . \end{aligned}$$

Note that

$$
\begin{aligned}
Q_{nk} &= n^{-3/2} \sum_{i=1}^{n} \sum_{j=1}^{n} \Bigg((I(X_j \le X_i) - H(X_i))\, \bar{C}_k(F(X_i) \mid \theta_0) \\
&\qquad - \int_{\mathbb{R}^d} (I(X_j \le x) - H(x))\, \bar{C}_k(F(x) \mid \theta_0)\, \mathrm{d}H(x) \Bigg) \\
&= \bar{Q}_{nk} + O(n^{-1/2}) \text{ a.s.},
\end{aligned}
$$

where

$$
\bar{Q}_{nk} = n^{-3/2} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left(\kappa(X_i, X_j) + \kappa(X_j, X_i) \right).
$$

Further

$$
\begin{aligned}
\mathrm{E}\, (\kappa(x, X) + \kappa(X, x)) &= \int_{\mathbb{R}^d} (\mathbf{1}(y \le x) - H(x))\, \mathrm{d}H(y) \bar{C}_k(F(x) \mid \theta_0) \\
&\quad - \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} (\mathbf{1}(y \le z) - H(z))\, \bar{C}_k(F(z) \mid \theta_0)\, \mathrm{d}H(z)\ \mathrm{d}H(y) \\
&= 0.
\end{aligned}
$$

Using Lemma 5.2.1A of Serfling [19], we obtain

$$
\operatorname{var}(\bar{Q}_{nk}) = O(n^{-1}).
$$

This completes the proof. □

Lemma 5.4. We have $T_{nk} \xrightarrow{\mathrm{P}} 0$ for $k = 1, \ldots, q$, where

$$
T_{nk} := \sqrt{n} \int_{\mathbb{R}^d} (C(F_n(x) \mid \theta_0) - C(F(x) \mid \theta_0))\, \bar{C}_k(F(x) \mid \theta_0)\, \mathrm{d} \left(\hat{H}_n(x) - H(x) \right).
$$

Proof. Observe that

$$
T_{nk} = T_{n1k} + T_{n2k},
$$

where

$$
T_{n1k} = \sqrt{n} \sum_{l=1}^{d} \int_{\mathbb{R}^d} \tilde{C}_l(F(x) \mid \theta_0)\, (F_{nl}(x_l) - F_l(x_l))\, \bar{C}_k(F(x) \mid \theta_0)\, \mathrm{d} \left(\hat{H}_n(x) - H(x) \right).
$$

We introduce $\delta_n := \max_{l=1,\ldots,d} \sup_{z \in \mathbb{R}} |F_{nl}(z) - F(z)|$. Further

$$
|T_{n2k}| \le \sqrt{n} \sum_{l=1}^{d} \sup_{v,w \in \mathbb{R}^d : \|v-w\| \le d\delta_n} \left| \tilde{C}_l(v \mid \theta_0) - \tilde{C}_l(w \mid \theta_0) \right| \cdot \delta_n = o_{\mathrm{P}}(1).
$$

Let

$$\begin{aligned}\kappa(x,y) &= \sum_{l=1}^{d} \left(\mathbf{1}(y_l \le x_l) - F_l(x_l)\right) \bar{C}_k(F(x) \mid \theta_0)\tilde{C}_l(F(x) \mid \theta_0) \\ &\quad - \sum_{l=1}^{d} \int_{\mathbb{R}^d} \left(\mathbf{1}(y_l \le z_l) - F_l(z_l)\right) \bar{C}_k(F(z) \mid \theta_0)\tilde{C}_l(F(z) \mid \theta_0)\, \mathrm{d}H(z)\,.\end{aligned}$$

Hence

$$T_{n1} = n^{-3/2} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left(k(X_i, X_j) + k(X_j, X_i)\right).$$

Further

$$\begin{aligned}\mathrm{E}\,(\kappa(x,X) + \kappa(X,x)) &= \int_{\mathbb{R}^d} \sum_{l=1}^{d} \left(\mathbf{1}(y_l \le x_l) - F_l(x_l)\right) \mathrm{d}H(y)\bar{C}_k(F(x) \mid \theta_0)\tilde{C}_l(F(x) \mid \theta_0) \\ &\quad - \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \sum_{l=1}^{d} \left(\mathbf{1}(y_l \le z_l) - F_l(z_l)\right) \\ &\qquad \bar{C}_k(F(z) \mid \theta_0)\tilde{C}_l(F(z) \mid \theta_0)\, \mathrm{d}H(z)\, \mathrm{d}H(y) \\ &= 0.\end{aligned}$$

By virtue of Lemma 5.2.1A of Serfling [19], we obtain

$$\mathrm{var}\,(T_{n1k}) = O(n^{-1}),$$

which proves the lemma. □

Analogously to the preceding lemma, one proves the following lemma.

Lemma 5.5. We have $U_n \xrightarrow{\mathrm{P}} 0$ where

$$\begin{aligned}U_n := \sqrt{n} &\int_{\mathbb{R}^d} (H(x) - C(F(x) \mid \theta_0)) \\ &\times \left(\bar{C}_k(F_n(x) \mid \theta_0) - \bar{C}_k(F(x) \mid \theta_0)\right) \mathrm{d}(\hat{H}_n(x) - H(x)).\end{aligned}$$

Lemma 5.6. We have

$$\sqrt{n}\nabla_\theta f_n(\theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma_2),$$

where Σ_2 as in Theorem 3.2.

P r o o f. Note that

$$\int_{\mathbb{R}^d} (H(x) - C(F(x) \mid \theta_0))\, \bar{C}_k(F(x) \mid \theta_0)\, \mathrm{d}H(x) = 0 \quad \text{for } k = 1, \dots, q.$$

Observe that

$$\nabla_\theta f_n(\theta_0) = \sum_{i=1}^{4} A_{ni} + B_n,$$

where

$$\begin{aligned}
A_{n1} &:= \int_{\mathbb{R}^d} \left(\hat{H}_n(x) - H(x) - C(V_n(x) \mid \theta_0) + C(F(x) \mid \theta_0) \right) \\
&\qquad (\nabla_\theta C(F_n(x) \mid \theta_0) - \nabla_\theta C(F(x) \mid \theta_0)) \, \mathrm{d}\hat{H}_n(x) \\
A_{n2} &:= \int_{\mathbb{R}^d} \left(\hat{H}_n(x) - H(x) \right) \nabla_\theta C(F(x) \mid \theta_0) \, \mathrm{d}(\hat{H}_n(x) - H(x)), \\
A_{n3} &:= \int_{\mathbb{R}^d} (C(F(x) \mid \theta_0) - C(F_n(x) \mid \theta_0)) \, \nabla_\theta C(F(x) \mid \theta_0) \, \mathrm{d}(\hat{H}_n(x) - H(x)), \\
A_{n4} &:= \int_{\mathbb{R}^d} (H(x) - C(F(x) \mid \theta_0)) \, (\nabla_\theta C(F_n(x) \mid \theta_0) - \nabla_\theta C(F(x) \mid \theta_0)) \, \mathrm{d}(\hat{H}_n(x) - H(x)), \\
B_n &:= \int_{\mathbb{R}^d} (H(x) - C(F(x) \mid \theta_0)) \, (\nabla_\theta C(F_n(x) \mid \theta_0) - \nabla_\theta C(F(x) \mid \theta_0)) \, \mathrm{d}H(x) \\
&\quad + \int_{\mathbb{R}^d} \left(\hat{H}_n(x) - C(F_n(x) \mid \theta_0) \right) \nabla_\theta C(F(x) \mid \theta_0) \, \mathrm{d}H(x) \\
&\quad + \int_{\mathbb{R}^d} (H(x) - C(F(x) \mid \theta_0)) \, \nabla_\theta C(F(x) \mid \theta_0) \, \mathrm{d}(\hat{H}_n(x) - H(x)).
\end{aligned}$$

By the law of iterated logarithm, we obtain

$$\begin{aligned}
|A_{n1}| &\leq \int_{\mathbb{R}^d} \left(\left| \hat{H}_n(x) - H(x) \right| + \sum_{j=1}^{d} \sup_{z \in \mathbb{R}} |F_{ni}(z) - F_i(z)| \right) \\
&\qquad \|\nabla_\theta C(F_n(x) \mid \theta_0) - \nabla_\theta C(F(x) \mid \theta_0)\| \, \mathrm{d}\hat{H}_n(x) \\
&= O(\ln \ln n / n) \text{ a.s.}
\end{aligned}$$

Using Lemmas 5.3 to 5.5, we obtain

$$|A_{n2}| = o_{\mathrm{P}}(n^{-1/2}), \ |A_{n3}| = o_{\mathrm{P}}(n^{-1/2}), \ |A_{n4}| = o_{\mathrm{P}}(n^{-1/2}).$$

Further we deduce

$$\begin{aligned}
t^T B_n &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{k=1}^{q} t_k \left(\sum_{j=1}^{d} \int_{\mathbb{R}^d} \left(I\left(X_i^{(j)} \leq x_j \right) - F_j(x_j) \right) \gamma_{kj}(x \mid \theta_0) \, \mathrm{d}H(x) \right. \\
&\quad + \int_{\mathbb{R}^d} \left(I\left(X_i \leq x \right) - H(x) \right) \bar{C}_k(F(x) \mid \theta_0) \, \mathrm{d}H(x) \\
&\quad + (H(X_i) - C(F(X_i) \mid \theta_0)) \, \bar{C}_k(F(X_i) \mid \theta_0) \\
&\qquad \left. - \int_{\mathbb{R}^d} (H(x) - C(F(x) \mid \theta_0)) \, \bar{C}_k(F(x) \mid \theta_0) \, \mathrm{d}H(x) \right) + \bar{B}_n
\end{aligned}$$

for $t = (t_1, \ldots, t_q)^T \in \mathbb{R}^q$. Further

$$\begin{aligned} \left|\bar{B}_n\right| &\leq \sqrt{n} \sum_{k=1}^{q} \sum_{j=1}^{d} |t_k| \Bigg(\sup_{v,w \in \mathbb{R}^d : \|v-w\| \leq d\delta_n} \left|\tilde{C}_{jk}(v \mid \theta_0) - \tilde{C}_{jk}(w \mid \theta_0)\right| \cdot \delta_n \\ &\qquad \sup_{v,w \in \mathbb{R}^d : \|v-w\| \leq d\delta_n} \left|\tilde{C}_j(v \mid \theta_0) - \tilde{C}_j(w \mid \theta_0)\right| \left|\bar{C}_k(F(x) \mid \theta_0)\right| \Bigg) \\ &= o_{\mathrm{P}}(1) \end{aligned}$$

with δ_n as in Lemma 5.4. The lemma follows by applying the central limit theorem and the Cramér–Wold device. □

Lemma 5.7. We obtain

$$\mathcal{H}_{nij}(t^*) \xrightarrow{\mathrm{P}} \mathcal{H}_{ij}(\theta_0)$$

with $\mathcal{H}_{ij}$ as in Section 3.

Proof. We deduce

$$\begin{aligned} \mathcal{H}_{nij}(t^*) &= \int_{\mathbb{R}^d} (H(x) - C(F(x) \mid \theta_0)) \frac{\partial^2}{\partial\theta_i \partial\theta_j} C(F_n(x) \mid t^*) \, \mathrm{d}\hat{H}_n(x) \\ &\quad + \int_{\mathbb{R}^d} \bar{C}_i(F(x) \mid \theta_0) \bar{C}_j(F(x) \mid \theta_0) \, \mathrm{d}\hat{H}_n(x) + o_{\mathrm{P}}(1) \\ &= \int_{\mathbb{R}^d} (H(x) - C(F(x) \mid \theta_0)) \frac{\partial^2}{\partial\theta_i \partial\theta_j} C(F(x) \mid \theta_0) \, \mathrm{d}\hat{H}_n(x) \\ &\quad + \int_{\mathbb{R}^d} \bar{C}_i(F(x) \mid \theta_0) \bar{C}_j(F(x) \mid \theta_0) \, \mathrm{d}\hat{H}_n(x) + o_{\mathrm{P}}(1). \end{aligned}$$

The lemma follows by applying the law of large numbers. □

Proof of Theorem 3.2. Note that $\tilde{\theta}_n - \hat{\theta}_n = o_{\mathrm{P}}(n^{-1/2})$. By (15), an application of Lemmas 5.6 and 5.7 leads to Theorem 3.2. □

ACKNOWLEDGEMENT

The author is grateful to the referees for their helpful suggestions.

(Received May 14, 2008.)

REFERENCES

[1] X. Chen and Y. Fan: Pseudo-likelihood ratio tests for semiparametric multivariate copula model selection. Canad. J. Statist. *33* (2005), 389–414.

[2] X. Chen and Y. Fan: Estimation of copula-based semiparametric time series models. J. Econometrics *130* (2006), 307–335.

[3] X. Chen and Y. Fan, and V. Tsyrennikov: Efficient estimation of semiparametric multivariate copula models. J. Amer. Statist. Assoc. *101* (2006), 1228–1240.

[4] J.-D. Fermanian: Goodness-of-fit tests for copulas. J. Multivariate Anal. *95* (2005), 119–152.

[5] C. Genest, K. Ghoudi, and L.-P. Rivest: A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. Biometrika *82* (1995), 543–552.

[6] C. Genest, J.-F. Quessy, and B. Rémillard: Goodness-of-fit procedures for copula models based on the probability integral transformation. Scand. J. Statist. *33* (2006), 337–366.

[7] C. Genest and L.-P. Rivest: Statistical inference procedures for bivariate Archimedian copulas. J. Amer. Statist. Assoc. *88* (1993), 1034–1043.

[8] C. Genest and B. J. M. Werker: Conditions for the asymptotic efficiency of an omnibus estimator of dependence estimators in copula models. In: Distributions with Given Marginals and Statistical Modelling (C. M. Cuadras et al., eds.), Kluwer, Dordrecht 2000, pp. 103–112.

[9] I. Gijbels and J. Mielniczuk: Estimating the density of a copula function. Comm. Statist. – Theory Methods *19* (1990), 445–464.

[10] P. Hall and N. Neumayer: Estimating a bivariate density when there are extra data on one or both components. Working Paper, 2006.

[11] C. Hess: Epi-convergence of sequences of normal integrands and strong consistency of the maximum likelihood estimator. Ann. Statist. *24* (1996), 1298–1315.

[12] H. Joe: Multivariate Models and Dependence Concepts. Chapman and Hall, London 1997.

[13] H. Joe: Asymptotic efficiency of the two-stage estimation method for copula-based models. J. Multivariate Anal. *94* (2005), 401–419.

[14] P. Lachout, E. Liebscher, and S. Vogel: Strong convergence of estimators as ε_n-minimisers of optimisation problems. Ann. Inst. Statist. Math. *57* (2005), 291–313.

[15] E. Liebscher: Semiparametric density estimators using copulas. Comm. Statist. – Theory Methods *34* (2005), 59–71.

[16] E. Liebscher: Construction of asymmetric multivariate copulas. J. Multivariate Anal. *99* (2008), 2234–2250.

[17] R. B. Nelsen: An Introduction to Copulas. (Lecture Notes in Statistics 139.) Springer, New York 1999.

[18] D. Oakes: Multivariate Survival distributions. J. Nonparametric Statist. *3* (1994), 343–354.

[19] R. Serfling: Approximation Theorems of Mathematical Statistics. Wiley, New York 1980.

[20] J. H. Shih and T. A. Louis: Inferences on the association parameter in copula models for bivariate survival data. Biometrics *55* (1995), 1384–1399.

[21] A. Suzukawa, H. Imai, and Y. Sato: Kullback–Leibler information consistent estimation for censored data. Ann. Inst. Statist. Math. *53* (2001), 262–276.

[22] H. Tsukahara: Semiparametric estimation in copula models. Canad. J. Statist. *33* (2005), 357–375.

[23] W. Wang and M. T. Wells: Model selection and semiparametric inference for bivariate failure-time data. J. Amer. Statist. Assoc. *95* (2000), 62–76.

Eckhard Liebscher, University of Applied Sciences Merseburg, Department of Computer Sciences and Communication Systems, Geusaer Straße 06217, Merseburg. Germany.

e-mail: eckhard.liebscher@hs-merseburg.de