

Liebscher, Eckhard

Article — Published Version

Fitting copulas in the case of missing data

Statistical Papers

Provided in Cooperation with:

Springer Nature

Suggested Citation: Liebscher, Eckhard (2024) : Fitting copulas in the case of missing data, Statistical Papers, ISSN 1613-9798, Springer, Berlin, Heidelberg, Vol. 65, Iss. 6, pp. 3681-3711, <https://doi.org/10.1007/s00362-024-01535-3>

This Version is available at:

<https://hdl.handle.net/10419/314990>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



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



Fitting copulas in the case of missing data

Eckhard Liebscher¹

Received: 21 February 2023 / Revised: 6 December 2023 / Published online: 27 March 2024

© The Author(s) 2024

Abstract

In this paper we deal with parametric estimation of the copula in the case of missing data. The data items with the same pattern of complete and missing data are combined into a subset. This approach corresponds to the MCAR model for missing data. We construct a specific Cramér–von Mises statistic as a sum of such statistics for the several missing data patterns. The minimization of the statistic gives the estimators for the parameters. We prove asymptotic normality of the parameter estimators and of the Cramér–von Mises statistic.

Keywords Copula · Cramér–von Mises statistic · Minimum distance estimators · Missing data

1 Introduction

When dealing with data from applications, especially from ecology or reliability, one can make the observation that missing values occur rather often. Unfortunately, there is no simple rule for handling missing data in the context of multivariate distributions. General strategies for statistical inference in the presence of missing data are discussed in the popular book by Little and Rubin (2019), see also Graham (2009). In the maximum-likelihood methodology, the EM-algorithm is the method of choice and provides good results in many cases. Many approaches dealing with missing data work by replacing missing data values with plausible values. These imputation methods are discussed in a survey paper by Van Buuren et al. (2006). Imputation is typically based on the idea of calculating the missing components by using its conditional distribution given the observed components.

As is well known, the copula model together with the (one-dimensional) marginal distribution functions define the multivariate distribution functions. A vast number of papers deal with the estimation of one-dimensional distributions. In this paper,

✉ Eckhard Liebscher
eckhard.liebscher@hs-merseburg.de

¹ Department of Engineering and Natural Sciences, University of Applied Sciences Merseburg, Eberhard-Leibnitz-Straße 2, 06217 Merseburg, Germany

the focus is on the parametric estimation of the copula. Here we do not consider imputation methods. We regard the data as given in a special structure, where the set of multivariate data is divided into a certain number of subsets of data with the same pattern of missingness. This approach corresponds to the MCAR model (missing completely at random model) in which the missingness of data items is independent of the data values. The several models for datasets including missing data are explained in Van Buuren et al. (2006) and Little and Rubin (2019), among others. For the data structure under consideration, an adapted Cramér–von Mises statistic is constructed which serves as an approximation measure and describes the discrepancy between the model and the data. Each pattern of missing data leads to a certain Cramér–von Mises statistic. The final statistic is then established as a linear combination of these partial statistics. The Cramér–von Mises statistic has numerical advantages over the Kullback–Leibler statistic (maximum-likelihood estimation) since the computation of the density is not required. The aim of the paper is to prove theorems on the almost sure convergence and on the asymptotic normality of the minimum distance estimators for the copula.

Because of the complexity of the multivariate distribution, we cannot expect that the underlying distribution of the sample vectors coincides with the hypothesis distribution. Thus, it makes sense to assume that the underlying copula does not belong to the parametric family under consideration. The reader finds an extensive discussion about goodness-of-approximation in the one-dimensional case in Liebscher (2014). Considering approximate estimators is another aspect of this paper. Since as a rule there is no explicit formula for the estimators, we have to evaluate the estimator for the copula parameters by a numerical algorithm and receive the estimator as a solution of an optimization problem only at a certain (small) error. These approximate estimators are the subject of the considerations in Sect. 5.

Concerning the estimation of the parameters of the copula, two types of estimators are studied mostly in the literature: maximum pseudo-likelihood estimators and minimum distance estimators. In our approach, minimum distance estimators on the basis of Cramér–von Mises divergence are the appropriate choice. In the case of complete data, minimum distance estimators for the parameters of copulas were examined in the papers by Tsukahara (2005) and by the author (2009). The asymptotic behaviour of likelihood estimators was investigated in papers by Genest and Rivest (1993), Genest et al. (1995), Chen and Fan (2005), and Hofert et al. (2012), among others. Joe (2005) published results on the asymptotic behaviour of two-stage estimation procedures. An application of the EM-algorithm to fitting Gaussian copulas in the presence of missing data can be found in the paper by Kertel and Pauly (2022). Rather few papers deal with estimation of copula parameters in the context of missing data. In the paper by Di Lascio et al. (2015), the authors studied the imputation method for parametric copula estimation. Hamori et al. (2019) considered the estimation of copula parameters in the missing at random model using only complete cases to estimate the actual parameter. In Wang et al. (2014), the estimation of the parameter of Gaussian copulas was examined under the missing completely at random assumption applying a special method tailored only for Gaussian copulas.

The paper is organized as follows: In Sect. 2 we introduce the data structure and the distribution functions of the subsets. The following Sect. 3 introduces the empirical

marginal distribution functions appropriate for the data structure and provides a law of iterated logarithm for them. The Cramér–von Mises divergence and its estimator are considered in Sect. 4. Section 5 provides the definition of approximate minimum distance estimators in our context. Moreover, we give the results on almost sure convergence and on asymptotic normality of the estimators of the copula parameters. The problem of goodness of approximation is discussed there, too. Section 6 contains a small simulation study. Section 7 provides the computational results of a data example. The asymptotic normality result of the Cramér–von Mises divergence can be found in Sect. 8. The proofs of the results are located in Sect. 9.

2 Data structure

Let $\mathbf{X} = (X^{(1)}, \dots, X^{(d)})^T$ be a d -dimensional random vector representing the data without missing values. In the case of a complete observation, we denote the joint distribution function by H and the marginal distribution functions of $X^{(j)}$ by F_1, \dots, F_d . Assume that F_j is continuous ($j = 1, \dots, d$). According to Sklar’s theorem (Sklar 1959), we have

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

where $C : [0, 1]^d \rightarrow [0, 1]$ is the uniquely determined d -dimensional copula. The reader can find the theory of copulas in the popular monographs by Joe (1997) and by Nelsen (2006).

Next we describe the structure of the data, including missing values. The sample breaks down into m subsets of data items with the same pattern of missing data, where m does not depend on the sample size n . Every pattern is modeled as a binary nonrandom vector $\mathbf{b} = (b_1, \dots, b_d)^T \in \{0, 1\}^d$, called the missing indicator vector, which has at least two components equal to 1. $b_j = 1$ means that the j -th component is observed whereas $b_j = 0$ means that the j -th component is missing. Now let $\mathbf{b}^{(1)}, \dots, \mathbf{b}^{(m)} \in \{0, 1\}^d$ be the pattern vectors of the data subsets. $J_\mu = \{l : b_l^{(\mu)} = 1\}$ denotes the set of numbers of observed non-missing components. To give an example, the pattern $\mathbf{b}^{(\mu)} = (0, 1, 0, 1)^T$ of data subset μ means that the data items of this subset have a non-missing second component and a non-missing fourth one, whereas components 1 and 3 are missing ($J_\mu = \{2, 4\}$). Let $\mathbf{1} = (1, \dots, 1)^T \in \mathbb{R}^d$.

$n_\mu = n_\mu(n)$ is the non-random number of sample items in the subset μ . In this paper the crucial assumption is that for all data subsets, the distribution function of the data items coincides with the corresponding multivariate marginal distribution functions resulting from H . More precisely, we assume that H is the underlying distribution function of the data, and therefore, the distribution function of data subset μ is given by

$$H_\mu(y_j, j \in J_\mu) = H(\bar{\mathbf{y}}) \quad \text{for } \bar{\mathbf{y}} \in \mathbb{R}^d, \tag{1}$$

where $\bar{y}_l = y_l$ for $l \in J_\mu$, and $\bar{y}_l = \infty$ for $l \notin J_\mu$. $(y_j, j \in J_\mu)$ denotes a vector of components of $\mathbf{y} \in \mathbb{R}^d$ with ascending indices from J_μ . The data structure and the subset distribution functions are then as follows:

Table 1 Structure of the data

Subset	Data	Pattern	Distribution fcn.
1	$Y_{11}, \dots, Y_{1n_1} \in \mathbb{R}^d$ complete data	$\mathbf{b}^{(1)} = \mathbf{1}$	$H_1 = H$
2	$Y_{21}, \dots, Y_{2n_2} \in \mathbb{R}^{d_2}$	$\mathbf{b}^{(2)}$	H_2
...			
m	$Y_{m1}, \dots, Y_{mn_m} \in \mathbb{R}^{d_m}$	$\mathbf{b}^{(m)}$	H_m

d_μ is the dimension of the data in the subset μ . Let C be the copula of distribution function H . The copulas of the subsets are determined by

$$C_\mu(u_j, j \in J_\mu) = C(\mathbf{u} \odot \mathbf{b}^{(\mu)} + \mathbf{1} - \mathbf{b}^{(\mu)}) \quad (\mu = 1, \dots, m)$$

for $\mathbf{u} = (u_1, \dots, u_d)^T \in [0, 1]^d$, where $\mathbf{a} \odot \mathbf{b} = \text{diag}(\mathbf{a}) \mathbf{b}$ is the Hadamard product of vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$. Define $\bar{F}(\mathbf{x}) = (F_1(x_1), \dots, F_d(x_d))^T$ for $\mathbf{x} = (x_1, \dots, x_d)^T$. Then we have

$$\begin{aligned} H_\mu(y_j, j \in J_\mu) &= C(\bar{F}(\mathbf{y}) \odot \mathbf{b}^{(\mu)} + \mathbf{1} - \mathbf{b}^{(\mu)}) \\ &= C_\mu(F_j(y_j), j \in J_\mu) \quad (\mu = 1, \dots, m). \end{aligned} \tag{2}$$

In the case $\mathbf{b}^{(\mu)} \geq \mathbf{b}^{(v)}$, i.e. $b_j^{(\mu)} \geq b_j^{(v)}$ for $j = 1, \dots, d$, the function $\psi_{v\mu}$ selects the components of subset μ , which are also present in subset v : $\psi_{v\mu}(y_j, j \in J_\mu) = (y_j, j \in J_v)$. The function $\bar{\psi}_{l\mu}$ selects the component l of the corresponding full data vector from data of subset μ : $\bar{\psi}_{l\mu}(y_j, j \in J_\mu) = y_l$.

The usage of the missing indicators and the sets J_μ is shown in an example.

Example We consider $d = 5$ and $\mathbf{b}^{(\mu)} = (1, 0, 1, 1, 0)^T \Rightarrow$

$$C_\mu(u_1, u_3, u_4) = C(u_1, 1, u_3, u_4, 1).$$

Further let $v = 2, \mu = 3, J_2 = \{2, 4\}, J_3 = \{2, 3, 4\}$. Then $\psi_{23}((y_2, y_3, y_4)^T) = (y_2, y_4)^T, \bar{\psi}_{43}((y_2, y_3, y_4)^T) = y_4$.

The requirements on the data following the MCAR model for missing data are summarized in Assumption $\mathcal{A}_{\text{MCAR}}$:

Assumption $\mathcal{A}_{\text{MCAR}}$: The data structure of Table 1 is given. Moreover (1) is valid where $J_\mu = \{l : b_l^{(\mu)} = 1\}$.

This data structure is present also in the situation, where m samples having the same underlying distribution are given. These samples may originate from several sources.

3 Empirical distribution functions

In this section we consider the empirical marginal distribution functions and their convergence properties. Let \tilde{n}_j be the number of data items where the j -th component is present:

$$\tilde{n}_j = \sum_{\mu: 1 \leq \mu \leq m, j \in J_\mu} n_\mu.$$

We introduce \bar{n}_ν to be the number of data items where at least the non-missing components of data subset ν are present

$$\bar{n}_\nu = \sum_{\mu: 1 \leq \mu \leq m, \mathbf{b}^{(\mu)} \geq \mathbf{b}^{(\nu)}} n_\mu$$

The inequality $\mathbf{b}^{(\mu)} \geq \mathbf{b}^{(\nu)}$ means that data subset μ has at least the non-missing components of data subset ν .

Notice that $\tilde{\psi}_{j\mu}(\mathbf{Y}_{\mu i})$ and $\psi_{\nu\mu}(\mathbf{Y}_{\mu i})$ have the distribution functions F_j and H_ν , respectively. Next we consider estimators for the marginal distributions and the joint distribution functions:

$$\hat{F}_{nj}(z) = \frac{1}{\tilde{n}_j} \sum_{\mu: 1 \leq \mu \leq m, j \in J_\mu} \sum_{i=1}^{n_\mu} \mathbf{1}\{\tilde{\psi}_{j\mu}(\mathbf{Y}_{\mu i}) \leq z\},$$

$$\hat{H}_{\nu\nu}(\mathbf{y}) = \frac{1}{\bar{n}_\nu} \sum_{\mu: 1 \leq \mu \leq m, \mathbf{b}^{(\mu)} \geq \mathbf{b}^{(\nu)}} \sum_{i=1}^{n_\mu} \mathbf{1}\{\psi_{\nu\mu}(\mathbf{Y}_{\mu i}) \leq \mathbf{y}\}$$

for $z \in \mathbb{R}, \mathbf{y} \in \mathbb{R}^{d_\nu}, \nu = 1, \dots, m$. We pose the following assumption on n_μ .

Assumption \mathcal{A}_n : For $\mu = 1, \dots, m$,

$$\lim_{n \rightarrow \infty} \frac{n_\mu(n)}{n} = \gamma_\mu$$

with constants $\gamma_\mu \in (0, 1]$. □

For empirical distribution functions, the following law of iterated logarithm holds true (cf. Kiefer 1961, for example).

Proposition 3.1 *Suppose that Assumptions \mathcal{A}_{MCAR} and \mathcal{A}_n are fulfilled.*

(a) *Then we have*

$$\max_{j=1, \dots, d} \sup_{t \in \mathbb{R}} |F_{nj}(t) - F_j(t)| = O\left(\sqrt{\frac{\ln \ln n}{n}}\right) \text{ a.s.}$$

(b) *Moreover,*

$$\max_{\mu=1, \dots, m} \sup_{\mathbf{y} \in \mathbb{R}^{d_\mu}} \left| \hat{H}_{n\mu}(\mathbf{y}) - H_\mu(\mathbf{y}) \right| = O \left(\sqrt{\frac{\ln \ln n}{n}} \right) \text{ a.s.}$$

for $n \rightarrow \infty$.

4 Cramér–von Mises divergence

Let $\mathcal{F} = \{C(\cdot \mid \theta)\}_{\theta \in \Theta}$ be a parametric family of copulas. $\Theta \subset \mathbb{R}^q$ is the parameter space. In this paper we want to approximate the sample copula C by the family \mathcal{F} . For this purpose, we consider the *Cramér–von Mises divergence* as a measure for the discrepancy between the copula C and \mathcal{F} . Define the model copula for subset μ :

$$C_\mu(u_j, j \in J_\mu \mid \theta) = C(\mathbf{u} \odot \mathbf{b}^{(\mu)} + \mathbf{1} - \mathbf{b}^{(\mu)} \mid \theta)$$

for $u \in [0, 1]^d, \theta \in \Theta, \mu = 1, \dots, m$. Let $\bar{F}_\mu^*(y_j, j \in J_\mu) = (F_j(y_j))_{j \in J_\mu}$, and $\check{F}_{n\mu}^*(y_j, j \in J_\mu) = (\hat{F}_{nj}(y_j))_{j \in J_\mu}$. \bar{F}_μ^* is the vector of the marginal distribution functions in subset μ , $\check{F}_{n\mu}^*$ is its empirical counter-part. We introduce the population version of the divergence as

$$\begin{aligned} \mathcal{D}(C, C(\cdot \mid \theta)) &= \sum_{\mu=1}^m \int_{[0,1]^{d_\mu}} (C_\mu(\bar{\mathbf{u}}) - C_\mu(\bar{\mathbf{u}} \mid \theta))^2 w_\mu(\bar{\mathbf{u}}) dC_\mu(\bar{\mathbf{u}}) \\ &= \sum_{\mu=1}^m \int_{\mathbb{R}^{d_\mu}} (H_\mu(\mathbf{y}) - C_\mu(\bar{F}_\mu^*(\mathbf{y}) \mid \theta))^2 w_\mu(\bar{F}_\mu^*(\mathbf{y})) dH_\mu(\mathbf{y}), \end{aligned} \tag{3}$$

where H_μ is as in (2). We pose the following assumption on w_μ :

Assumption \mathcal{A}_W : Assume that $w_\mu : [0, 1]^{d_\mu} \rightarrow [0, +\infty), \mu = 1, \dots, m$ are Lipschitz-continuous weight functions for data subsets μ . □

By assumption \mathcal{A}_W , the functions w_μ are bounded. An example of such a weight function is given by

$$w_\mu(\mathbf{u}) = \frac{\bar{w}_\mu}{a + \prod_{j=1}^{d_\mu} u_j(1 - u_j)} \text{ for } \mathbf{u} \in \mathbb{R}^{d_\mu}, \tag{4}$$

where $a, \bar{w}_\mu > 0$ are constants. The divergence \mathcal{D} is the weighted sum of the squared discrepancies between the sample copula and the parametric model copula within the concerning data subsets. In general, smaller values of the divergence \mathcal{D} show a better approximation by \mathcal{F} . Observe that $C_\mu(\bar{\mathbf{u}}) - C_\mu(\bar{\mathbf{u}} \mid \theta) = 0$ for $\bar{\mathbf{u}} \in \mathcal{B} := \{\mathbf{u} : u_j = 0 \text{ for at least one } j, \text{ or } u_j = 1 \text{ for all } j \text{ except one}\}$. To put more emphasis on the fit in

the boundary regions in the neighbourhood of \mathcal{B} , the weight functions can be defined in a suitable way similarly to (4).

The concept of a weighted divergence has already been applied by some authors. For instance, Rodriguez and Viollaz (1995) studied the asymptotic distribution of the weighted Cramér–von Mises divergence in the one-dimensional case. Medovikov (2016) examined weighted Cramér–von Mises tests for independence employing the weighted Cramér–von Mises statistic with independence copula as the model copula. We refer to the thorough discussion about weights in Medovikov’s paper where also further references can be found. The L^p -distance and the Kolmogorov–Smirnov distance are alternatives to \mathcal{D} , see Liebscher (2015), for example.

Next, we construct an estimator for $\mathcal{D}(C, \mathcal{C}(\cdot | \theta))$ in the situation where the data structure is as introduced in Sect. 2:

$$\widehat{\mathcal{D}}_n(\theta) = \sum_{\mu=1}^m \frac{1}{n_\mu} \sum_{i=1}^{n_\mu} \left(\widehat{H}_{n\mu}(\mathbf{Y}_{\mu i}) - C_\mu(\check{F}_{n\mu}^*(\mathbf{Y}_{\mu i}) | \theta) \right)^2 w_\mu(\check{F}_{n\mu}^*(\mathbf{Y}_{\mu i})) \quad (5)$$

for $\theta \in \Theta$, where $\widehat{H}_{n\mu}$ is defined in Sect. 3. This estimator has the advantage of being just a sum and does not require to compute an integral. Genest et al. (2009) found that the use of the Cramér–von Mises statistic leads to more powerful goodness-of-fit tests in comparison to other test statistics like the Kolmogorov–Smirnov one. The next section is devoted to the estimation of the parameter θ using the divergence (5).

5 Parameter estimation by the minimum distance method

Let the data structure be as in Sect. 2. Consider the family $\mathcal{F} = (\mathcal{C}(\cdot | \theta))_{\theta \in \Theta}$ of copulas with the parameter set $\Theta \subset \mathbb{R}^q$. Throughout the paper, we assume that $C \notin \mathcal{F}$. Many authors call this case the misspecification one. In our opinion, this term is not appropriate for the situation here. If we consider multivariate data, then typically, C does not belong to any parametric family. In this section the aim is to estimate the parameter θ_0 which gives the best approximation for the copula in the case of a unique minimizer of \mathcal{D} defined in (3):

$$\theta_0 = \arg \min_{\theta \in \Theta} \mathcal{D}(C, \mathcal{C}(\cdot | \theta)),$$

$\mathcal{D}(C, \mathcal{C}(\cdot | \theta))$ as in the previous section. It should be highlighted that in general, θ_0 depends on the choice of the discrepancy measure. There is no “true parameter”. Considering $\mathcal{D}(C, \mathcal{C}(\cdot | \theta))$, θ_0 depends on the weight functions w_μ , and these functions should be chosen prior to the analysis. In the case of constant weight functions, it seems to be reasonable to choose the weights w_μ such that the summands in (3) for the estimator of θ_0 are roughly equal.

The estimator $\widehat{\theta}_n$ is referred to as an *approximate minimum distance estimator* (AMDE) if

$$\widehat{\mathcal{D}}_n(\widehat{\theta}_n) \leq \min_{\theta \in \Theta} \widehat{\mathcal{D}}_n(\theta) + \varepsilon_n$$

holds true (\widehat{D}_n as in the previous section), where $\{\varepsilon_n\}$ is a sequence of random variables with $\varepsilon_n \rightarrow 0$ a.s. Note that $\hat{\theta}_n$ is an approximate minimizer of $\theta \mapsto \widehat{D}_n(\theta)$. We refer to Liebscher (2009), where the estimator was introduced. In the case of unique θ_0 , $\hat{\theta}_n$ is an estimator for θ_0 . Tsukahara (2005) examined properties of a similar (non-approximate) minimum distance estimator.

Let $\|\cdot\|$ be the Euclidean norm, and $d(x, A) = \inf_{y \in A} \|x - y\|$ for $x \in \mathbb{R}^q$ and subsets $A \subset \mathbb{R}^q$. The following theorem provides the result about the consistency of the AMDE including the case of sets of minimizers of \mathcal{D} .

Theorem 5.1 *Assume that Assumptions \mathcal{A}_{MCAR} , \mathcal{A}_n and \mathcal{A}_W are satisfied. Let $\theta \mapsto C(\mathbf{u} \mid \theta)$ be continuous on Θ for every $\mathbf{u} \in [0, 1]^d$. Suppose that Θ is compact.*

(a) *Then*

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

where $\Psi = \arg \min_{\theta \in \Theta} \mathcal{D}(C, \mathcal{C}(\cdot \mid \theta)) \subset \mathbb{R}^q$.

(b) *If in addition, the condition*

$$\mathcal{D}(C, \mathcal{C}(\cdot \mid \theta)) > \mathcal{D}(C, \mathcal{C}(\cdot \mid \theta_0)) \quad \text{for all } \theta \in \Theta \setminus \{\theta_0\} \tag{6}$$

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

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

Part (a) of Theorem 5.1 gives sufficient conditions for the almost sure convergence of AMDE $\hat{\theta}_n$ to the set of minimizers of \mathcal{D} w.r.t. θ whereas part b) is the ordinary consistency result. The proof is based on a result from Lachout et al. (1994). The assumption that Θ is compact is not as problematic as it seems. In many cases with infinite Θ , a continuous bijective function can be used to transform the parameter onto a finite interval. Then it can be verified that the consistency holds for the transformed parameter, and hence for the original parameter on suitable intervals. The assumption on compactness of Θ is posed to reduce the technical efforts in the proofs.

The next Theorem 5.2 states that $\hat{\theta}_n$ is asymptotically normally distributed in the case $\Psi = \{\theta_0\}$ under appropriate assumptions. The following assumption on partial derivatives of the copula is needed in this theorem.

Assumption \mathcal{A}_C : $\bar{C}_k(\cdot \mid \theta)$, $\bar{C}_{kl}(\cdot \mid \theta)$, $\bar{C}_j(\mathbf{u} \mid \cdot)$, $\bar{C}_{jk}(\mathbf{u} \mid \theta)$ denote the partial derivatives $\frac{\partial}{\partial \theta_k} \mathcal{C}(\cdot \mid \theta)$, $\frac{\partial^2}{\partial \theta_k \partial \theta_l} \mathcal{C}(\cdot \mid \theta)$, $\frac{\partial}{\partial u_j} \mathcal{C}(\mathbf{u} \mid \cdot)$, $\frac{\partial^2}{\partial \theta_k \partial u_j} \mathcal{C}(\mathbf{u} \mid \theta)$, respectively. We assume that these derivatives exist, and for $k, l = 1, \dots, q$, $j = 1, \dots, d$, the functions $(\mathbf{u}, t) \mapsto \bar{C}_{kl}(\mathbf{u} \mid t)$, $(\mathbf{u}, t) \mapsto \bar{C}_{jk}(\mathbf{u} \mid t)$ 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 . Moreover, the partial derivatives of $w_\mu : [0, 1]^{d_\mu} \rightarrow [0, +\infty)$ are denoted by $w_{\mu l}$, $l \in J_\mu$, and assumed to exist and be continuous. □

If Assumption \mathcal{A}_C is satisfied, then we use the notations

$$\bar{C}_{\mu k}^\circ(u_\lambda, \lambda \in J_\mu \mid \cdot) = \bar{C}_k^\circ(\tilde{\mathbf{u}}_\mu \mid \cdot), \quad \bar{C}_{\mu kl}^\circ(u_\lambda, \lambda \in J_\mu \mid \cdot) = \bar{C}_{kl}^\circ(\tilde{\mathbf{u}}_\mu \mid \cdot),$$

$$\tilde{C}_{\mu j}^{\circ}(u_{\lambda}, \lambda \in J_{\mu} \mid \cdot) = \tilde{C}_j(\tilde{\mathbf{u}}_{\mu} \mid \cdot), \quad \tilde{C}_{\mu j k}^{\circ}(u_{\lambda}, \lambda \in J_{\mu} \mid \cdot) = \tilde{C}_{j k}(\tilde{\mathbf{u}}_{\mu} \mid \cdot)$$

for $k, l = 1, \dots, q, j \in J_{\mu}$, where $\tilde{\mathbf{u}}_{\mu} = \mathbf{u} \odot \mathbf{b}^{(\mu)} + \mathbf{1} - \mathbf{b}^{(\mu)}$, $\mathbf{u} \in [0, 1]^d$. Define $\mathcal{H} = (\mathcal{H}_{kl})_{k,l=1,\dots,q}$ as the Hessian matrix of $\theta \mapsto \mathcal{D}(C, C(\cdot \mid \theta))$ at $\theta = \theta_0$:

$$\begin{aligned} \mathcal{H}_{kl} = & -2 \sum_{\mu=1}^m \int_{\mathbb{R}^{d_{\mu}}} \left((H_{\mu}(\mathbf{y}) - C_{\mu}(\bar{F}_{\mu}^*(\mathbf{y}) \mid \theta_0)) \bar{C}_{\mu kl}^{\circ}(\bar{F}_{\mu}^*(\mathbf{y}) \mid \theta_0) \right. \\ & \left. - \bar{C}_{\mu k}^{\circ}(\bar{F}_{\mu}^*(\mathbf{y}) \mid \theta_0) \bar{C}_{\mu l}^{\circ}(\bar{F}_{\mu}^*(\mathbf{y}) \mid \theta_0) \right) w_{\mu}(\bar{F}_{\mu}^*(\mathbf{y})) \, dH_{\mu}(\mathbf{y}). \end{aligned}$$

Now we give the theorem:

Theorem 5.2 *Assume that $\varepsilon_n = o_{\mathbb{P}}(n^{-1})$, and the matrix \mathcal{H} is positive definite. Suppose that Assumptions \mathcal{A}_C and the assumptions of Theorem 5.1b) are satisfied. Then*

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

Here $\Sigma = \mathcal{H}^{-1} \Sigma_D \mathcal{H}^{-1}$,

$$\begin{aligned} \mathbf{Z}_{\mu v} = & (H_{\mu}(\mathbf{Y}_{\mu 1}) - C_{\mu}(\bar{F}_{\mu}^*(\mathbf{Y}_{\mu 1}) \mid \theta_0)) \nabla_{\theta} C_{\mu}(\bar{F}_{\mu}^*(\mathbf{Y}_{\mu 1}) \mid \theta_0) w_{\mu}(\bar{F}_{\mu}^*(\mathbf{Y}_{\mu 1})) \\ & + \frac{\mathbf{1}(\mathbf{b}^{(v)} \geq \mathbf{b}^{(\mu)})}{\bar{\gamma}_{\mu}} \int_{\mathbb{R}^{d_{\mu}}} \mathbf{1}\{\psi_{\mu v}(\mathbf{Y}_{\mu 1}) \leq \mathbf{z}\} \nabla_{\theta} C_{\mu}(\bar{F}_{\mu}^*(\mathbf{z}) \mid \theta_0) w_{\mu}(\bar{F}_{\mu}^*(\mathbf{z})) \, dH_{\mu}(\mathbf{z}) \\ & + \sum_{l=1}^d \frac{1}{\bar{\gamma}_l} b_l^{(\mu)} b_l^{(v)} \int_{\mathbb{R}^{d_{\mu}}} \left(-\bar{C}_{\mu l}^{\circ}(\bar{F}_{\mu}(\mathbf{z}) \mid \theta_0) \nabla_{\theta} C_{\mu}(\bar{F}_{\mu}^*(\mathbf{z}) \mid \theta_0) w_{\mu}(\bar{F}_{\mu}^*(\mathbf{z})) \right. \\ & \left. + (H_{\mu}(\mathbf{z}) - C_{\mu}(\bar{F}_{\mu}(\mathbf{z}) \mid \theta_0)) \cdot \right. \\ & \left. \cdot \left(\bar{C}_{\mu l k}^{\circ}(\bar{F}_{\mu}^*(\mathbf{z}) \mid \theta_0) w_{\mu}(\bar{F}_{\mu}^*(\mathbf{z})) + \bar{C}_{\mu k}^{\circ}(\bar{F}_{\mu}^*(\mathbf{z}) \mid \theta_0) w_{\mu l}(\bar{F}_{\mu}^*(\mathbf{z})) \right)_{k=1,\dots,q} \right) \\ & \mathbf{1}\{\bar{\psi}_{l v}(\mathbf{Y}_{\mu 1}) \leq \bar{\psi}_{l \mu}(\mathbf{z})\} \, dH_{\mu}(\mathbf{z}), \\ \Sigma_D = & 4 \sum_{v=1}^m \gamma_v \sum_{\mu=1}^m \sum_{\bar{\mu}=1}^m \text{cov}(\mathbf{Z}_{\mu v}, \mathbf{Z}_{\bar{\mu} v}). \end{aligned}$$

Here $\text{cov}(\cdot, \cdot)$ is the cross-covariance matrix.

In Liebscher (2009) this result was proved in the case $m = 1$ and for complete data. Theorem 5.2 corrects some typos in the formula for Σ in the author’s 2009 paper. Tsukahara (2005) proved consistency and asymptotic normality for his minimum distance estimator in the case where the copula C of X_i belongs to a small neighbourhood of a member of the parametric family. The covariance structure of the estimator $\hat{\theta}_n$ is rather complicated. One potential approach is to estimate Σ by substituting distribution functions with their empirical counterparts, and θ_0 by $\hat{\theta}_n$. In view of the sophisticated structure of this estimator, one may use alternative techniques like bootstrap to get approximate values for the covariances.

To compare the various fitting results, we introduce the approximation coefficient

$$\hat{\rho} = 1 - \frac{\widehat{\mathcal{D}}_n(\hat{\theta}_n)}{\widehat{\mathcal{D}}_n^0},$$

where

$$\widehat{\mathcal{D}}_n^0 = \sum_{\mu=1}^m \frac{1}{n_\mu} \sum_{i=1}^{n_\mu} \left(\hat{H}_{n\mu}(\mathbf{Y}_{\mu i}) - \Pi(\check{F}_{n\mu}^*(\mathbf{Y}_{\mu i})) \right)^2 w_\mu \check{F}_{n\mu}^*(\mathbf{Y}_{\mu i}).$$

$\widehat{\mathcal{D}}_n^0(\theta)$ is the Cramér–von Mises divergence when the independence copula Π is used for fitting. Π is the reference copula in the definition of $\hat{\rho}$. The approximation coefficient is defined in analogy to the regression coefficient of determination and describes the grade of improvement of the fit using model \mathcal{F} in comparison to the independence copula. Obviously, we have $\hat{\rho} \leq 1$. In the case where the independence copula is included in the model family \mathcal{F} , the inequality $0 \leq \hat{\rho} \leq 1$ is fulfilled. Values $\hat{\rho}$ close to 1 indicate that the approximation is good. In case of models with very small dependencies, the value of $\hat{\rho}$ could be close to zero. Then it is recommended to use another reference copula instead of Π .

6 A small simulation study

A simulation study should show that the estimation method studied in this paper lead to a reasonable performance of the estimators. Here we simulated the following data with 100,000 repetitions of it:

three-dimensional data vectors $\mathbf{Y}_{\mu j}$ having copula C and marginal normal distributions $N(1, 0.6)$, $N(2, 0.3)$, $N(3, 0.8)$, respectively. Copula C is determined by

$$C(\mathbf{u}) = 0.5 \cdot C^{(1)}(\mathbf{u}) + 0.5 \cdot C^{(2)}(\mathbf{u}) \quad (\mathbf{u} \in \mathbb{R}^3),$$

where $C^{(1)}$ is the Clayton (3) copula and $C^{(2)}$ is the Frank (2) one (for formulas, see Nelsen (2006)).

$$\mathbf{b}^{(1)} = (1, 1, 1)^T, \quad \mathbf{b}^{(2)} = (1, 1, 0), \quad n_1 = n_2 = n/2$$

Half of the data are complete and the remaining data vectors have a missing third component. In the place of C , we considered the Frank, Clayton, Joe and Gumbel–Hougaard copula. Since the second summand in $\widehat{\mathcal{D}}_n$ ($\mu = 2$) is expected to be smaller than the first one (smaller dimension!), we chose $w_1 = 0.3$ and $w_2 = 0.7$. The results are summarized in Table 2.

The values θ_0 were computed using the computer algebra system Mathematica. The results of Table 2 indicate that the optimization leads to a reasonable approximation of copula C in the case of the considered copula families \mathcal{F} . The approximation becomes more precise when n increases; i.e. the average lies closer to θ_0 and the

Table 2 Simulation results (sd...standard deviation, mse...mean square error)

family \mathcal{F}	n	θ_0	Average	SD	MSE
Clayton	600	1.33492	1.3469	0.16735	0.028149
	1000	1.33492	1.3420	0.12911	0.016719
	2000	1.33492	1.3382	0.090843	0.0082630
Frank	600	4.08106	4.1019	0.39340	0.15520
	1000	4.08106	4.0924	0.30114	0.090814
	2000	4.08106	4.0878	0.21244	0.045175
Joe	600	2.20757	2.2227	0.17972	0.032527
	1000	2.20757	2.2168	0.13787	0.019095
	2000	2.20757	2.2119	0.097225	0.0094712
Gumbel-Hougaard	600	1.69526	1.7013	0.089772	0.0080960
	1000	1.69526	1.6987	0.069151	0.0047938
	2000	1.69526	1.6972	0.048827	0.0023879

Table 3 Used variables

Trait	Description
RCNC	Carbon–nitrogen ratio of roots
LCNC	Carbon–nitrogen ratio of leaves
HW	Height–width ratio of plant

standard deviation is smaller. Unfortunately, comparisons with results for other data structures are not very useful, since the divergence is constructed in accordance to the data scheme of Sect. 2. Note that θ_0 depends on the choice of the divergence. Further computations have revealed that the summands in \hat{D}_n for the several subsets differ only slightly when the weights w_j selected as above.

7 A data example

Here we consider a dataset from the TRY plant trait database, see Kattge et al. (2020). This dataset was already used for modelling and fitting in Liebscher et al. (2022). See this paper for a detailed description of the dataset. Here we restrict the considerations to three variables according to Table 2 and to 9 herb species: ‘Ac.mi’, ‘Be.pe’, ‘Ce.ja’, ‘Ga.mo’, ‘Ga.ve’, ‘Pl.la’, ‘Ra.ac’, ‘Ra.bu’, ‘Ve.ch’ (Table 3).

A first analysis of the dataset shows the various frequencies of the missing data patterns provided in Table 4.

Two patterns are sorted out because of too few data items. Therefore, we consider only two missing data patterns ($m = 2$) with equal weights $w_\mu = 0.5$. Now we want to fit product copulas to the ecological dataset (variables RCNC, LCNC, HW). Let $C^{(1)}, C^{(2)}, C^{(3)}$ be given copulas taken from parametric Archimedean copula families like the Frank, Clayton, Gumbel families (for formulas, see Nelsen 2006). The product copulas are defined by

Table 4 Missing data pattern

$b^{(j)}$	Count	
(1,0,0)	52	Not usable
(1,1,0)	4	Sorted out
(0,0,1)	14	Not usable
(1,0,1)	353	
(0,1,1)	6	Sorted out
(1,1,1)	197	

Table 5 Fitting results (abbreviations for copula families: C...Clayton, F...Frank, N...Nelsen #13)

Model	Estimated parameters	$\hat{\rho}$
$C^{(1)} * C^{(2)} = C * C$	(4.22483, -0.250466, 0.941397, 0.640162, 0)	0.8263
$C^{(1)} * C^{(2)} = C * N$	(4.34463, 0.559406, 0.964540, 0.635338, 0)	0.8262
$C^{(1)} * C^{(2)} = C * F$	(4.33694, -0.693147, 0.959506, 0.610715, 0)	0.8240
$C^{(1)} * C^{(2)} * C^{(3)} = C * C * N$	(3.76189, 69.4387, 0.0176261, 0.857769, 0.617536, 0, 0.25155, 0.27837, 0.36183)	0.8453
$C^{(1)} * C^{(2)} * C^{(3)} = C * N * N$	(4.03917, 0.01, 28.1875, 0.847152, 0.622940, 0, 0.762127, 0.736797, 0.584249)	0.8418

$$C(\mathbf{u}) = C^{(1)}(u_1^{\alpha_1}, \dots, u_d^{\alpha_d} | t_1)C^{(2)}(u_1^{1-\alpha_1}, \dots, u_d^{1-\alpha_d} | t_2),$$

in short $C^{(1)} * C^{(2)}$ with parameter vector $(t_1, t_2, \alpha_1, \dots, \alpha_d)^T$,

$$C(\mathbf{u}) = C^{(1)}(u_1^{\alpha_1}, \dots, u_d^{\alpha_d} | t_1)C^{(2)}(u_1^{(1-\alpha_1)\beta_1}, \dots, u_d^{(1-\alpha_d)\beta_d} | t_2)$$

$$C^{(3)}(u_1^{(1-\alpha_1)(1-\beta_1)}, \dots, u_d^{(1-\alpha_d)(1-\beta_d)} | t_3),$$

in short $C^{(1)} * C^{(2)} * C^{(3)}$ with parameter vector $(t_1, t_2, t_3, \alpha_1, \dots, \alpha_d, \beta_1, \dots, \beta_d)^T$.

In these formulas t_j is the parameter of $C^{(j)}$, and $\alpha_j, \beta_j \in [0, 1]$. Product copulas were studied in Liebscher (2008) in detail. Table 5 summarizes the fitting results.

From Table 5 and further computational results, we see that the approximation is fairly good for model C*F in the case of the product of 2 copulas, and for model C*N*N in the case of the product of 3 copulas. In the latter case, a slightly better approximation coefficient is obtained in comparison to the product of 2 copulas.

8 Convergence of the CvM-divergence

In this section we give the theorem on asymptotic normality of the CvM-divergence $\widehat{D}_n(\hat{\theta}_n)$.

Theorem 8.1 *Assume that the assumptions of Theorem 5.2 are satisfied. Then*

$$\sqrt{n}(\widehat{D}_n(\hat{\theta}_n) - \mathcal{D}(C, \mathcal{C}(\cdot | \theta_0))) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma_0).$$

The formula for Σ_0 is given in the proof.

In the case $m = 1$ and for complete data, this theorem was already proven in Liebscher (2015). Theorem 8.1 can be used to construct tests about the divergence $\mathcal{D}(C, \mathcal{C}(\cdot | \theta))$. We refer to the discussion in the author’s paper (2015).

9 Proofs

9.1 Proof of Theorem 5.1

Let $\Phi_n : \Theta \rightarrow \mathbb{R}$ be a random function, and $\hat{\theta}_n$ be an estimator satisfying

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

Here $\{\varepsilon_n\}$ is a sequence of random variables with $\varepsilon_n \rightarrow 0$ a.s. Theorem 2.2 of the paper Lachout et al. (2005) leads to the following proposition.

Proposition 9.1 *Assume that Θ is compact, and $\lim_{n \rightarrow \infty} \sup_{t \in \Theta} |\Phi_n(t) - \Phi(t)| = 0$ a.s. holds for a continuous function Φ .*

(a) *Then*

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

where $\Psi = \operatorname{argmin}_{t \in \Theta} \Phi(t) \subset \mathbb{R}^q$, $d(\cdot, \cdot)$ as above.

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

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

Let $\Phi_n(\theta) = \widehat{\mathcal{D}}_n(\theta)$ and $\Phi(\theta) = \mathcal{D}(C, \mathcal{C}(\cdot | \theta))$. In this section, the aim is to apply Proposition 9.1 in order to prove the strong consistency result for $\hat{\theta}_n$. The following lemma justifies the strong uniform consistency assumption in Proposition 9.1.

Lemma 9.2 *Assume that assumptions of Theorem 5.1 are fulfilled. Then*

$$\lim_{n \rightarrow \infty} \sup_{\theta \in \Theta} |\widehat{\mathcal{D}}_n(\theta) - \mathcal{D}(C, \mathcal{C}(\cdot | \theta))| = 0 \text{ a.s.}$$

Proof Notice that

$$|a^2 - b^2| \leq 2|a - b| \text{ for } a, b \in [0, 1].$$

Utilizing the Lipschitz continuity of copulas with Lipschitz constant 1 and the triangle inequality, we obtain

$$\sup_{\theta \in \Theta} |\widehat{\mathcal{D}}_n(\theta) - \mathcal{D}(C, \mathcal{C}(\cdot | \theta))|$$

$$\begin{aligned}
 &= \sup_{\theta \in \Theta} \left| \sum_{\mu=1}^m \frac{1}{n_\mu} \sum_{i=1}^{n_\mu} \left(\hat{H}_{n\mu}(\mathbf{Y}_{\mu i}) - C_\mu(\check{F}_{n\mu}^*(\mathbf{Y}_{\mu i}) \mid \theta) \right)^2 w_\mu(\check{F}_{n\mu}^*(\mathbf{Y}_{\mu i})) \right. \\
 &\quad \left. - \int_{\mathbb{R}^{d_\mu}} (H_\mu(y) - C_\mu(\bar{F}_\mu^*(y) \mid \theta))^2 w_\mu(\bar{F}_\mu^*(y)) \, dH_\mu(y) \right| \\
 &\leq 2m \max_{\mu: 1 \leq \mu \leq m} \left(\left(\sup_{\mathbf{y} \in \mathbb{R}^{d_\mu}} \left| \hat{H}_{n\mu}(\mathbf{y}) - H_\mu(\mathbf{y}) \right| + \sup_{\theta \in \Theta} \sup_{\mathbf{x} \in \mathbb{R}^{d_\mu}} \left| C_\mu(\check{F}_{n\mu}^*(\mathbf{x}) \mid \theta) - C_\mu(\bar{F}_\mu^*(\mathbf{x}) \mid \theta) \right| \right) \right. \\
 &\quad \left. \cdot \sup_{\mathbf{x} \in [0,1]^{d_\mu}} w_\mu(\mathbf{x}) \right) + Q_n + 2R_n \\
 &\leq 2m \max_{\mu: 1 \leq \mu \leq m} \left(\left(\sup_{\mathbf{y} \in \mathbb{R}^{d_\mu}} \left| \hat{H}_{n\mu}(\mathbf{y}) - H_\mu(\mathbf{y}) \right| + \sum_{j=1}^d \sup_{x \in \mathbb{R}} |F_{nj}(x) - F_j(x)| \right) \cdot \sup_{\mathbf{x} \in [0,1]^{d_\mu}} w_\mu(\mathbf{x}) \right) \\
 &\quad + Q_n + 2R_n, \tag{7}
 \end{aligned}$$

where

$$\begin{aligned}
 Q_n &= \sum_{\mu=1}^m \sup_{\theta \in \Theta} \frac{1}{n_\mu} \left| \sum_{i=1}^{n_\mu} \left(\hat{H}_{n\mu}(\mathbf{Y}_{\mu i}) - C_\mu(\check{F}_{n\mu}^*(\mathbf{Y}_{\mu i}) \mid \theta) \right)^2 \left(w_\mu(\check{F}_{n\mu}^*(\mathbf{Y}_{\mu i})) - w_\mu(\bar{F}_\mu^*(\mathbf{Y}_{\mu i})) \right) \right|, \\
 R_n &= \sum_{\mu=1}^m \sup_{\theta \in \Theta} \left| \frac{1}{n_\mu} \sum_{i=1}^{n_\mu} (H_\mu(\mathbf{Y}_{\mu i}) - C_\mu(\bar{F}_\mu^*(\mathbf{Y}_{\mu i}) \mid \theta))^2 w_\mu(\bar{F}_\mu^*(\mathbf{Y}_{\mu i})) \right. \\
 &\quad \left. - \int_{\mathbb{R}^{d_\mu}} (H_\mu(\mathbf{y}) - C_\mu(\bar{F}_\mu^*(\mathbf{y}) \mid \theta))^2 w_\mu(\bar{F}_\mu^*(\mathbf{y})) \, dH_\mu(\mathbf{y}) \right|.
 \end{aligned}$$

Since $\theta \mapsto (H_\mu(\mathbf{y}) - C_\mu(\bar{F}_\mu^*(\mathbf{y}) \mid \theta))^2 w_\mu(\bar{F}_\mu^*(\mathbf{y}))$ is continuous for $\mathbf{y} \in \mathbb{R}^{d_\mu}$ by assumptions, and the envelope function $w_\mu(\bar{F}_\mu^*(\cdot))$ is integrable, the strong Glivenko–Cantelli theorem (see Van der Vaart (1998), Theorem 19.4 and Example 19.8) implies

$$R_n \rightarrow 0 \text{ a.s.}$$

as $n \rightarrow \infty$. Further, by Assumption \mathcal{A}_W (L is the Lipschitz-constant of w_μ),

$$\begin{aligned}
 Q_n &\leq \sum_{\mu=1}^m \frac{1}{n_\mu} \sum_{i=1}^{n_\mu} \left| w_\mu(\check{F}_{n\mu}^*(\mathbf{Y}_{\mu i})) - w_\mu(\bar{F}_\mu^*(\mathbf{Y}_{\mu i})) \right| \\
 &\leq L \sum_{\mu=1}^m \frac{1}{n_\mu} \sum_{i=1}^{n_\mu} \left\| \check{F}_{n\mu}^*(\mathbf{Y}_{\mu i}) - \bar{F}_\mu^*(\mathbf{Y}_{\mu i}) \right\| \\
 &\leq Lm \sum_{j=1}^d \sup_{x \in \mathbb{R}} |F_{nj}(x) - F_j(x)| \rightarrow 0 \text{ a.s.}
 \end{aligned}$$

An application of Proposition 3.1 leads to the lemma. □

Proof of Theorem 5.1 Theorem 5.1 is a direct consequence of Proposition 9.1 and Lemma 9.2. □

9.2 Auxiliary statements

The following lemma about the convergence of the marginal empirical distribution functions can be stated.

Lemma 9.3

$$\max_{j=1, \dots, d} \sup_{z \in \mathbb{R}} \left| \hat{F}_{jn}(z) - F_j(z) \right| = O_{\mathbb{P}} \left(\frac{1}{\sqrt{n}} \right)$$

Proof The assertion follows from the Dvoretzky–Kiefer–Wolfowitz inequality, see van der Vaart (1998, p. 268).

In the following we derive central limit theorems. First we consider

$$W_n = \sum_{\mu=1}^m \sqrt{\frac{n_{\mu}}{n}} W_n^{(\mu)}, \quad W_n^{(\mu)} = \frac{1}{\sqrt{n_{\mu}}} \sum_{i=1}^{n_{\mu}} (g_{\mu}(\mathbf{Y}_{\mu i}) - \mathbb{E}g_{\mu}(\mathbf{Y}_{\mu i}))$$

with functions $g_{\mu} : \mathbb{R}^{d_{\mu}} \rightarrow \mathbb{R}^{\kappa}$ and provide a central limit theorem for W_n . □

Proposition 9.4 *Suppose that assumption \mathcal{A}_n is fulfilled, and $\mathbb{E} \|g_{\mu}(\mathbf{Y}_{\mu 1})\|^2 < \infty$ for $\mu = 1, \dots, m$. Then we have*

$$W_n \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma_W),$$

where $\Sigma_W = \sum_{\mu=1}^m \gamma_{\mu} \text{cov}(g_{\mu}(\mathbf{Y}_{\mu 1}))$, and $\text{cov}(Z)$ is covariance matrix of random vector Z .

Proof Applying the multivariate central limit theorem (see Serfling 1980, Theorem 1.9.1B), we obtain that $W_n^{(1)}, \dots, W_n^{(m)}$ are asymptotically normally distributed. Since these summands $W_n^{(\mu)}$ of W_n are independent, we can conclude the asymptotic normality of W_n . For the covariance matrix of W_n , we obtain

$$\frac{1}{n} \sum_{\mu=1}^m \sum_{i=1}^{n_{\mu}} \text{cov}(g_{\mu}(\mathbf{Y}_{\mu i})) = \frac{1}{n} \sum_{\mu=1}^m n_{\mu} \text{cov}(g_{\mu}(\mathbf{Y}_{\mu 1})) \rightarrow \sum_{\mu=1}^m \gamma_{\mu} \text{cov}(g_{\mu}(\mathbf{Y}_{\mu 1}))$$

as $n \rightarrow \infty$. □

Let $\Lambda_{\mu\nu} : \mathbb{R}^{d_{\mu}} \times \mathbb{R}^{d_{\nu}} \rightarrow \mathbb{R}^{\kappa}$ be measurable functions for $\mu, \nu = 1, \dots, m$, $\Lambda_{\mu\nu} = (\Lambda_{\mu\nu}^{(1)}, \dots, \Lambda_{\mu\nu}^{(\kappa)})^T$. Next we derive a central limit theorem for the U -statistic

$$U_n = \frac{1}{n\sqrt{n}} \sum_{\mu=1}^m \sum_{i=1}^{n_{\mu}} \sum_{\nu=1}^m \sum_{j=1}^{n_{\nu}} (\Lambda_{\mu\nu}(\mathbf{Y}_{\mu i}, \mathbf{Y}_{\nu j}) - \mathbb{E}\Lambda_{\mu\nu}(\mathbf{Y}_{\mu i}, \mathbf{Y}_{\nu j})).$$

Let $\theta_{\mu\nu} = \mathbb{E}\Lambda_{\mu\nu}(\mathbf{Y}_{\mu 1}, \mathbf{Y}_{\nu 2})$ for all μ, ν . We introduce

$$\tilde{h}_{\mu\nu}(\mathbf{y}) = \mathbb{E}\Lambda_{\mu\nu}(\mathbf{Y}_{\mu 1}, \mathbf{y}) + \mathbb{E}\Lambda_{\nu\mu}(\mathbf{y}, \mathbf{Y}_{\mu 1}) - \theta_{\mu\nu} - \theta_{\nu\mu} \text{ for } \mathbf{y} \in \mathbb{R}^{d_\nu},$$

$\tilde{h}_{\mu\nu} = (\tilde{h}_{\mu\nu}^{(1)}, \dots, \tilde{h}_{\mu\nu}^{(\kappa)})^T$. Note that $\mathbb{E}\tilde{h}_{\mu\nu}(\mathbf{Y}_{\nu 1}) = \mathbb{E}\tilde{h}_{\nu\mu}(\mathbf{Y}_{\mu 1}) = 0$ for all μ, ν . Proposition 9.5 provides the central limit theorem for U_n . In the proof we use Hájek’s projection principle.

Proposition 9.5 *Suppose that $\mathbb{E}\Lambda_{\mu\nu}^{(L)}(\mathbf{Y}_{\mu 1}, \mathbf{Y}_{\nu j})^2 < +\infty$ for all $\mu, \nu, L = 1, \dots, \kappa, j = 1, 2$. We have*

$$U_n \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma_U),$$

where $\Sigma_U = \sum_{\nu=1}^m \gamma_\nu \sum_{\mu=1}^m \sum_{\bar{\mu}=1}^m \gamma_\mu \gamma_{\bar{\mu}} \mathbb{E}\tilde{h}_{\mu\nu}(\mathbf{Y}_{\nu 1})\tilde{h}_{\bar{\mu}\nu}^T(\mathbf{Y}_{\nu 1})$.

Proof Define

$$U_n^\circ = \frac{1}{n\sqrt{n}} \sum_{\mu=1}^m \sum_{i=1}^{n_\mu} (\Lambda_{\mu\mu}(\mathbf{Y}_{\mu i}, \mathbf{Y}_{\mu i}) - \mathbb{E}\Lambda_{\mu\mu}(\mathbf{Y}_{\mu i}, \mathbf{Y}_{\mu i})).$$

We obtain (\mathbb{V} denotes the variance)

$$\begin{aligned} \mathbb{E}\|U_n^\circ\|^2 &= n^{-3} \sum_{l=1}^\kappa \mathbb{E} \left(\sum_{\mu=1}^m \sum_{i=1}^{n_\mu} (\Lambda_{\mu\mu}^{(l)}(\mathbf{Y}_{\mu i}, \mathbf{Y}_{\mu i}) - \mathbb{E}\Lambda_{\mu\mu}^{(l)}(\mathbf{Y}_{\mu i}, \mathbf{Y}_{\mu i})) \right)^2 \\ &\leq n^{-3} \sum_{l=1}^\kappa \sum_{\mu=1}^m \sum_{i=1}^{n_\mu} \mathbb{V}(\Lambda_{\mu\mu}^{(l)}(\mathbf{Y}_{\mu i}, \mathbf{Y}_{\mu i})) \\ &= O(n^{-2}). \end{aligned}$$

Hence $U_n^\circ \xrightarrow{\mathbb{P}} 0$ holds, and

$$\begin{aligned} U_n &= n^{-3/2} \sum_{\mu=1}^m \sum_{i=1}^{n_\mu} \sum_{\nu=\mu}^m \sum_{j=1+\delta_{\nu\mu}i}^{n_\nu} (\Lambda_{\mu\nu}(\mathbf{Y}_{\mu i}, \mathbf{Y}_{\nu j}) + \Lambda_{\nu\mu}(\mathbf{Y}_{\nu j}, \mathbf{Y}_{\mu i}) - \theta_{\mu\nu} - \theta_{\nu\mu}) \\ &\quad + o_{\mathbb{P}}(1), \end{aligned} \tag{8}$$

where $\delta_{\mu\mu} = 1$ and $\delta_{\nu\mu} = 0$ for $\nu \neq \mu$. Now we split the sum in (8) into two parts. Define

$$\begin{aligned} \tilde{U}_{n\mu\nu} &= n^{-3/2} \sum_{i=1}^{n_\mu} \sum_{j=1+\delta_{\nu\mu}i}^{n_\nu} \tilde{\Lambda}_{\mu\nu}(\mathbf{Y}_{\mu i}, \mathbf{Y}_{\nu j}), \text{ and} \\ \tilde{\Lambda}_{\mu\nu}(\mathbf{Y}_{\mu i}, \mathbf{Y}_{\nu j}) &= \Lambda_{\mu\nu}(\mathbf{Y}_{\mu i}, \mathbf{Y}_{\nu j}) + \Lambda_{\nu\mu}(\mathbf{Y}_{\nu j}, \mathbf{Y}_{\mu i}) - \tilde{h}_{\mu\nu}(\mathbf{Y}_{\nu j}) \end{aligned}$$

$$-\tilde{h}_{v\mu}(\mathbf{Y}_{\mu i}) - \theta_{\mu v} - \theta_{v\mu}.$$

Notice that $\mathbb{E}\tilde{\Lambda}_{\mu v}(\mathbf{Y}_{\mu i}, \mathbf{Y}_{v j}) = 0$. Then we have

$$U_n = \tilde{U}_n + \bar{U}_n + o_{\mathbb{P}}(1), \text{ where} \tag{9}$$

$$\tilde{U}_n = \sum_{\mu=1}^m \sum_{v=\mu}^m \tilde{U}_{n\mu v}, \quad \bar{U}_n = n^{-3/2} \sum_{\mu=1}^m \sum_{i=1}^{n_{\mu}} \sum_{v=\mu}^m \sum_{j=1+\delta_{v\mu i}}^{n_v} \left(\tilde{h}_{\mu v}(\mathbf{Y}_{v j}) + \tilde{h}_{v\mu}(\mathbf{Y}_{\mu i}) \right).$$

The next step is to show $\tilde{U}_n = o_{\mathbb{P}}(1)$. Later we prove asymptotic normality of \bar{U}_n . Observe that

$$\mathbb{E} \left(\tilde{\Lambda}_{\mu v}(\mathbf{Y}_{\mu i}, \mathbf{Y}_{v j}) \mid \mathbf{Y}_{\mu i} \right) = \tilde{h}_{v\mu}(\mathbf{Y}_{\mu i}) - \tilde{h}_{v\mu}(\mathbf{Y}_{\mu i}) = 0, \quad \mathbb{E} \left(\tilde{\Lambda}_{\mu v}(\mathbf{Y}_{\mu i}, \mathbf{Y}_{v j}) \right) = 0,$$

$$\mathbb{E} \left(\tilde{\Lambda}_{\mu v}(\mathbf{Y}_{\mu i}, \mathbf{Y}_{v j}) \mid \mathbf{Y}_{v j} \right) = \tilde{h}_{\mu v}(\mathbf{Y}_{v j}) - \tilde{h}_{\mu v}(\mathbf{Y}_{v j}) = 0, \quad \mathbb{E} \left(\tilde{\Lambda}_{\mu v}(\mathbf{Y}_{\mu i}, \mathbf{Y}_{v j}) \right) = 0$$

for $i \neq j$ or $\mu \neq v$. Therefore, identity

$$\mathbb{E}\tilde{\Lambda}_{\mu v}^{(L)}(\mathbf{Y}_{\mu i}, \mathbf{Y}_{v j})\tilde{\Lambda}_{\mu v}^{(L)}(\mathbf{Y}_{\mu i}, \mathbf{Y}_{v l})$$

$$= \mathbb{E} \left(\mathbb{E} \left(\tilde{\Lambda}_{\mu v}^{(L)}(\mathbf{Y}_{\mu i}, \mathbf{Y}_{v j}) \mid \mathbf{Y}_{\mu i} \right) \mathbb{E} \left(\tilde{\Lambda}_{\mu v}^{(L)}(\mathbf{Y}_{\mu i}, \mathbf{Y}_{v l}) \mid \mathbf{Y}_{\mu i} \right) \right) = 0$$

holds for $l \neq j$. Thus, by this equation and similar identities, we have

$$\mathbb{E}\tilde{\Lambda}_{\mu v}^{(L)}(\mathbf{Y}_{\mu i}, \mathbf{Y}_{v j})\tilde{\Lambda}_{\mu v}^{(L)}(\mathbf{Y}_{\mu k}, \mathbf{Y}_{v l}) = 0$$

for $(i = k, j \neq l) \vee (j = l, i \neq k) \vee (\mu = v, i = l, j \neq k) \vee (\mu = v, j = k, i \neq l)$. Obviously, this equation holds for different indices i, j, k, l . On the other hand, we have

$$\mathbb{E}\tilde{h}_{\mu v}^{(L)}(\mathbf{Y}_{v 1})^2 = \mathbb{E} \left(\mathbb{E} \left(\Lambda_{\mu v}^{(L)}(\mathbf{Y}_{\mu 1}, \mathbf{Y}_{v 1}) + \Lambda_{v\mu}^{(L)}(\mathbf{Y}_{v 1}, \mathbf{Y}_{\mu 1}) - \theta_{\mu v} - \theta_{v\mu} \mid \mathbf{Y}_{v 1} \right)^2 \right)$$

$$\leq \mathbb{E} \left(\mathbb{E} \left(\left(\Lambda_{\mu v}^{(L)}(\mathbf{Y}_{\mu 1}, \mathbf{Y}_{v 1}) + \Lambda_{v\mu}^{(L)}(\mathbf{Y}_{v 1}, \mathbf{Y}_{\mu 1}) - \theta_{\mu v} - \theta_{v\mu} \right)^2 \mid \mathbf{Y}_{v 1} \right) \right)$$

$$\leq 2 \left(\mathbb{E}\Lambda_{\mu v}^{(L)}(\mathbf{Y}_{\mu 1}, \mathbf{Y}_{v 1})^2 + \mathbb{E}\Lambda_{v\mu}^{(L)}(\mathbf{Y}_{v 1}, \mathbf{Y}_{\mu 1})^2 \right).$$

Consequently, for $\mu \neq v$, it can be derived

$$\mathbb{E} \left\| \tilde{U}_{n\mu v} \right\|^2 = n^{-3} \sum_{L=1}^{\kappa} \sum_{i=1}^{n_{\mu}} \sum_{j=1}^{n_v} \sum_{k=1}^{n_{\mu}} \sum_{l=1}^{n_v} \mathbb{E}\tilde{\Lambda}_{\mu v}^{(L)}(\mathbf{Y}_{\mu i}, \mathbf{Y}_{v j})\tilde{\Lambda}_{\mu v}^{(L)}(\mathbf{Y}_{\mu k}, \mathbf{Y}_{v l})$$

$$= n^{-3} \sum_{L=1}^{\kappa} \sum_{i=1}^{n_{\mu}} \sum_{j=1}^{n_v} \mathbb{E}\tilde{\Lambda}_{\mu v}^{(L)}(\mathbf{Y}_{\mu i}, \mathbf{Y}_{v j})^2$$

$$\begin{aligned} &\leq 4n^{-1} \sum_{L=1}^{\kappa} \left(\mathbb{E}\Lambda_{\mu\nu}^{(L)}(\mathbf{Y}_{\mu 1}, \mathbf{Y}_{\nu 1})^2 + \mathbb{E}\Lambda_{\nu\mu}^{(L)}(\mathbf{Y}_{\nu 1}, \mathbf{Y}_{\mu 1})^2 + \mathbb{E}\tilde{h}_{\mu\nu}^{(L)}(\mathbf{Y}_{\nu 1})^2 \right. \\ &\quad \left. + \mathbb{E}\tilde{h}_{\nu\mu}^{(L)}(\mathbf{Y}_{\mu 1})^2 \right) \\ &= O(n^{-1}). \end{aligned}$$

In a similar way, we obtain

$$\begin{aligned} \mathbb{E} \left\| \tilde{U}_{n\mu\mu} \right\|^2 &= n^{-3} \sum_{L=1}^{\kappa} \left(\sum_{i=1}^{n_{\mu}} \sum_{j=1+i}^{n_{\nu}} \mathbb{E}\tilde{\Lambda}_{\mu\mu}^{(L)}(\mathbf{Y}_{\mu i}, \mathbf{Y}_{\mu j})^2 \right. \\ &\quad \left. + \sum_{i=1}^{n_{\mu}} \sum_{j=1+i}^{n_{\nu}} \mathbb{E}\tilde{\Lambda}_{\mu\mu}^{(L)}(\mathbf{Y}_{\mu i}, \mathbf{Y}_{\mu j}) \tilde{\Lambda}_{\mu\mu}^{(L)}(\mathbf{Y}_{\mu j}, \mathbf{Y}_{\mu i}) \right) \\ &\leq 8n^{-1} \sum_{L=1}^{\kappa} \left(\mathbb{E}\Lambda_{\mu\mu}^{(L)}(\mathbf{Y}_{\mu 1}, \mathbf{Y}_{\mu 2})^2 + \mathbb{E}\tilde{h}_{\mu\mu}^{(L)}(\mathbf{Y}_{\mu 1})^2 \right) \\ &= O(n^{-1}). \end{aligned}$$

Hence $\tilde{U}_n \xrightarrow{\mathbb{P}} 0$ holds true. Notice that

$$\begin{aligned} \mathbb{E} \left\| n^{-3/2} \sum_{\mu=1}^m \sum_{i=1}^{n_{\mu}} \tilde{h}_{\mu\mu}(\mathbf{Y}_{\mu i}) \right\|^2 &= n^{-3} \sum_{L=1}^{\kappa} \mathbb{E} \left(\sum_{\mu=1}^m \sum_{i=1}^{n_{\mu}} \tilde{h}_{\mu\mu}^{(L)}(\mathbf{Y}_{\mu i}) \right)^2 \\ &= n^{-3} \sum_{L=1}^{\kappa} \sum_{\mu=1}^m \sum_{i=1}^{n_{\mu}} \mathbb{E}\tilde{h}_{\mu\mu}^{(L)}(\mathbf{Y}_{\mu i})^2 \\ &\leq 4n^{-2} \sum_{L=1}^{\kappa} \sum_{\mu=1}^m \mathbb{E}\Lambda_{\mu\mu}^{(L)}(\mathbf{Y}_{\mu 1}, \mathbf{Y}_{\mu 2})^2 \\ &= O(n^{-1}). \end{aligned}$$

On the other hand, it follows that

$$\begin{aligned} \tilde{U}_n &= n^{-3/2} \sum_{\mu=1}^m \sum_{i=1}^{n_{\mu}} \sum_{v=1}^m \sum_{j=1}^{n_{\nu}} \tilde{h}_{\mu\nu}(\mathbf{Y}_{\nu j}) \mathbf{1}(v \neq \mu \vee i \neq j) \\ &= n^{-3/2} \left(\sum_{\mu=1}^m n_{\mu} \sum_{v=1}^m \sum_{j=1}^{n_{\nu}} \tilde{h}_{\mu\nu}(\mathbf{Y}_{\nu j}) - \sum_{\mu=1}^m \sum_{i=1}^{n_{\mu}} \tilde{h}_{\mu\mu}(\mathbf{Y}_{\mu i}) \right) \\ &= U_n^{(1)} + U_n^{(2)} + o_{\mathbb{P}}(1), \end{aligned}$$

where

$$U_n^{(1)} = n^{-1/2} \sum_{v=1}^m \sum_{j=1}^{n_v} g_v(\mathbf{Y}_{vj}), \quad g_v(\mathbf{y}) = \sum_{\mu=1}^m \gamma_\mu \tilde{h}_{\mu v}(\mathbf{y}),$$

$$U_n^{(2)} = n^{-1/2} \sum_{\mu=1}^m \left(\frac{n_\mu}{n} - \gamma_\mu \right) \sum_{v=1}^m \sum_{j=1}^{n_v} \tilde{h}_{\mu v}(\mathbf{Y}_{vj}).$$

An application of Proposition 9.4 to the sum $U_n^{(1)}$ leads to $U_n^{(1)} \xrightarrow{d} \mathcal{N}(0, \Sigma_U)$, where

$$\begin{aligned} \Sigma_U &= \sum_{v=1}^m \gamma_v \mathbb{E} g_v(\mathbf{Y}_{v1}) g_v(\mathbf{Y}_{v1})^T \\ &= \sum_{v=1}^m \gamma_v \sum_{\mu=1}^m \sum_{\bar{\mu}=1}^m \gamma_\mu \gamma_{\bar{\mu}} \mathbb{E} \tilde{h}_{\mu v}(\mathbf{Y}_{v1}) \tilde{h}_{\bar{\mu} v}^T(\mathbf{Y}_{v1}). \end{aligned}$$

Analogously, one shows that

$$\sum_{v=1}^m \sum_{j=1}^{n_v} \tilde{h}_{\mu v}(\mathbf{Y}_{vj}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma_0)$$

for $\mu = 1, \dots, m$ with an appropriate Σ_0 , which implies $U_n^{(2)} \xrightarrow{\mathbb{P}} 0$. Therefore, we have $\tilde{U}_n \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma_U)$. By (9) and $\tilde{U}_n \xrightarrow{\mathbb{P}} 0$, the proof is complete. \square

9.3 Proof of Theorem 5.2

Throughout this section we suppose that the assumptions of Theorem 5.1b), and Assumption \mathcal{A}_C are satisfied. Here θ_0 is the unique minimizer of \mathcal{D} defined in (3). We introduce $\mathcal{H}_n(\theta) = (\mathcal{H}_{nkl}(\theta))_{k,l=1,\dots,q}$ as the Hessian of $\widehat{\mathcal{D}}_n(\theta)$:

$$\begin{aligned} \mathcal{H}_{nkl}(\theta) &= \sum_{\mu=1}^m \frac{2}{n_\mu} \sum_{i=1}^{n_\mu} \left(- \left(\widehat{H}_{n\mu}(\mathbf{Y}_{\mu i}) - \mathcal{C}_\mu(\check{F}_{n\mu}^*(\mathbf{Y}_{\mu i}) \mid \theta) \right) \bar{C}_{\mu kl}^\circ(\check{F}_{n\mu}^*(\mathbf{Y}_{\mu i}) \mid \theta) \right. \\ &\quad \left. + \bar{C}_{\mu k}^\circ(\check{F}_{n\mu}^*(\mathbf{Y}_{\mu i}) \mid \theta) \bar{C}_{\mu l}^\circ(\check{F}_{n\mu}^*(\mathbf{Y}_{\mu i}) \mid \theta) \right) w_\mu(\check{F}_{n\mu}^*(\mathbf{Y}_{\mu i})). \end{aligned}$$

$\nabla_\theta \psi(\theta)$ denotes the gradient of function ψ w.r.t. θ , and $\nabla_\theta \psi(\theta_0)$ is the abbreviation for $\nabla_\theta \psi(\theta)|_{\theta=\theta_0}$. Observe that

$$\begin{aligned} \nabla_\theta \widehat{\mathcal{D}}_n(\theta) &= - \sum_{\mu=1}^m \frac{2}{n_\mu} \sum_{i=1}^{n_\mu} \left(\widehat{H}_{n\mu}(\mathbf{Y}_{\mu i}) - \mathcal{C}_\mu(\check{F}_{n\mu}^*(\mathbf{Y}_{\mu i}) \mid \theta) \right) \\ &\quad \nabla_\theta \mathcal{C}_\mu(\check{F}_{n\mu}^*(\mathbf{Y}_{\mu i}) \mid \theta) w_\mu(\check{F}_{n\mu}^*(\mathbf{Y}_{\mu i})). \end{aligned}$$

Let $\tilde{\theta}_n$ be a minimizer of $\widehat{\mathcal{D}}_n(\cdot)$. Since $\nabla_{\theta} \widehat{\mathcal{D}}_n(\tilde{\theta}_n) = 0$, we can use the Taylor formula to derive (note that θ_0 is an interior point of Θ)

$$\nabla_{\theta} \widehat{\mathcal{D}}_n(\theta_0) = \nabla_{\theta} \widehat{\mathcal{D}}_n(\theta_0) - \nabla_{\theta} \widehat{\mathcal{D}}_n(\tilde{\theta}_n) = -\mathcal{H}_n^* \left(\tilde{\theta}_n - \theta_0 \right), \tag{10}$$

where $t_{nk}^* = \theta_0 + \eta_{nk} \left(\tilde{\theta}_n - \theta_0 \right)$ and $\mathcal{H}_n^* = (\mathcal{H}_{nkl}(t_{nk}^*))_{k,l=1,\dots,q}$. Here $\eta_{nk} \in (0, 1)$ is a random variable, $k = 1, \dots, q$.

Taking identity (10) into account, Theorem 5.2 is proven in three steps: we show the asymptotic normality of $\nabla_{\theta} \widehat{\mathcal{D}}_n(\theta_0)$, we prove that \mathcal{H}_n^* converges in probability to a certain matrix, and we show that $\tilde{\theta}_n - \hat{\theta}_n$ is $o_{\mathbb{P}}(n^{-1/2})$. The following lemma includes the first step.

Lemma 9.6 *We have*

$$\sqrt{n} \nabla_{\theta} \widehat{\mathcal{D}}_n(\theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma_D),$$

with Σ_D as in Theorem 5.2.

Proof We decompose $-\nabla_{\theta} \widehat{\mathcal{D}}_n(\theta_0)$ and obtain

$$-\nabla_{\theta} \widehat{\mathcal{D}}_n(\theta_0) = A_{1n} + A_{2n} + A_{3n} + A_{4n}, \tag{11}$$

$$\begin{aligned} A_{1n} &= \sum_{\mu=1}^m \frac{2}{n_{\mu}} \sum_{i=1}^{n_{\mu}} \left(H_{\mu}(\mathbf{Y}_{\mu i}) - C_{\mu}(\bar{F}_{\mu}^*(\mathbf{Y}_{\mu i}) \mid \theta_0) \right) \nabla_{\theta} C_{\mu}(\bar{F}_{\mu}^*(\mathbf{Y}_{\mu i}) \mid \theta_0) w_{\mu}(\bar{F}_{\mu}^*(\mathbf{Y}_{\mu i})), \\ A_{2n} &= \sum_{\mu=1}^m \frac{2}{n_{\mu}} \sum_{i=1}^{n_{\mu}} \left(\hat{H}_{n\mu}(\mathbf{Y}_{\mu i}) - C_{\mu}(\check{F}_{n\mu}^*(\mathbf{Y}_{\mu i}) \mid \theta_0) - H_{\mu}(\mathbf{Y}_{\mu i}) + C_{\mu}(\bar{F}_{\mu}^*(\mathbf{Y}_{\mu i}) \mid \theta_0) \right) \\ &\quad \left(\nabla_{\theta} C_{\mu}(\check{F}_{n\mu}^*(\mathbf{Y}_{\mu i}) \mid \theta_0) w_{\mu}(\check{F}_{n\mu}^*(\mathbf{Y}_{\mu i})) - \nabla_{\theta} C_{\mu}(\bar{F}_{\mu}^*(\mathbf{Y}_{\mu i}) \mid \theta_0) w_{\mu}(\bar{F}_{\mu}^*(\mathbf{Y}_{\mu i})) \right), \\ A_{3n} &= \sum_{\mu=1}^m \frac{2}{n_{\mu}} \sum_{i=1}^{n_{\mu}} \left(\hat{H}_{n\mu}(\mathbf{Y}_{\mu i}) - C_{\mu}(\check{F}_{n\mu}^*(\mathbf{Y}_{\mu i}) \mid \theta_0) - H_{\mu}(\mathbf{Y}_{\mu i}) + C_{\mu}(\bar{F}_{\mu}^*(\mathbf{Y}_{\mu i}) \mid \theta_0) \right) \\ &\quad \nabla_{\theta} C_{\mu}(\bar{F}_{\mu}^*(\mathbf{Y}_{\mu i}) \mid \theta_0) w_{\mu}(\bar{F}_{\mu}^*(\mathbf{Y}_{\mu i})), \\ A_{4n} &= \sum_{\mu=1}^m \frac{2}{n_{\mu}} \sum_{i=1}^{n_{\mu}} \left(H_{\mu}(\mathbf{Y}_{\mu i}) - C_{\mu}(\bar{F}_{\mu}^*(\mathbf{Y}_{\mu i}) \mid \theta_0) \right) \\ &\quad \left(\nabla_{\theta} C_{\mu}(\check{F}_{n\mu}^*(\mathbf{Y}_{\mu i}) \mid \theta_0) w_{\mu}(\check{F}_{n\mu}^*(\mathbf{Y}_{\mu i})) - \nabla_{\theta} C_{\mu}(\bar{F}_{\mu}^*(\mathbf{Y}_{\mu i}) \mid \theta_0) w_{\mu}(\bar{F}_{\mu}^*(\mathbf{Y}_{\mu i})) \right). \end{aligned}$$

Further we define

$$A_{3n}^* = \sum_{\mu=1}^m \frac{2}{n_{\mu}} \sum_{i=1}^{n_{\mu}} \nabla_{\theta} C_{\mu}(\bar{F}_{\mu}^*(\mathbf{Y}_{\mu i}) \mid \theta_0) w_{\mu}(\bar{F}_{\mu}^*(\mathbf{Y}_{\mu i}))$$

$$\begin{aligned}
 & \left(\hat{H}_{n\mu}(\mathbf{Y}_{\mu i}) - H_{\mu}(\mathbf{Y}_{\mu i}) - \sum_{l \in J_{\mu}} \tilde{C}_{\mu l}^{\circ}(\bar{F}_{\mu}^*(\mathbf{Y}_{\mu i}) \mid \theta_0) \right. \\
 & \quad \left. \left(\hat{F}_{nl}(\bar{\psi}_{l\mu}(\mathbf{Y}_{\mu i})) - F_l(\bar{\psi}_{l\mu}(\mathbf{Y}_{\mu i})) \right) \right), \\
 A_{4nk}^* &= \sum_{\mu=1}^m \frac{2}{n_{\mu}} \sum_{i=1}^{n_{\mu}} \left(H_{\mu}(\mathbf{Y}_{\mu i}) - C_{\mu}(\bar{F}_{\mu}^*(\mathbf{Y}_{\mu i}) \mid \theta_0) \right) \\
 & \quad \sum_{l \in J_{\mu}} \left(\tilde{C}_{\mu lk}^{\circ}(\bar{F}_{\mu}^*(\mathbf{Y}_{\mu i}) \mid \theta_0) w_{\mu}(\bar{F}_{\mu}^*(\mathbf{Y}_{\mu i})) + \tilde{C}_{\mu k}^{\circ}(\bar{F}_{\mu}^*(\mathbf{Y}_{\mu i}) \mid \theta_0) w_{\mu}(\bar{F}_{\mu}^*(\mathbf{Y}_{\mu i})) \right) \\
 & \quad \left(\hat{F}_l(\bar{\psi}_{l\mu}(\mathbf{Y}_{\mu i})) - F_l(\bar{\psi}_{l\mu}(\mathbf{Y}_{\mu i})) \right),
 \end{aligned}$$

and $A_{4n}^* = (A_{4nk}^*)_{k=1, \dots, q}$, $A_{4n} = (A_{4nk})_{k=1, \dots, q}$. Obviously, $\nabla_{\theta} \mathcal{D}(C, \mathcal{C}(\cdot \mid \theta_0)) = 0$ such that $\mathbb{E}A_{1n} = 0$. The next step is to show that $A_{2n} + (A_{3n} - A_{3n}^*) + (A_{4n} - A_{4n}^*) = o_{\mathbb{P}}(n^{-1/2})$. Note that copulas are Lipschitz continuous. Since the partial derivatives $\tilde{C}_{\mu l}^{\circ}(\cdot \mid \theta_0)$ are Lipschitz continuous by assumption \mathcal{A}_C and the weight functions w_{μ} are Lipschitz continuous by assumption \mathcal{A}_W , we obtain

$$\begin{aligned}
 \|A_{2n}\| &\leq O(1) \max_{\mu=1, \dots, m} \sup_{\mathbf{y} \in \mathbb{R}^{d_{\mu}}} \left(\left| \hat{H}_{n\mu}(\mathbf{y}) - H_{\mu}(\mathbf{y}) \right| + \left\| \check{F}_{n\mu}^*(\mathbf{y}) - \bar{F}_{\mu}^*(\mathbf{y}) \right\| \right) \\
 & \quad \max_{\mu=1, \dots, m} \sup_{\mathbf{y} \in \mathbb{R}^{d_{\mu}}} \left\| \check{F}_{n\mu}^*(\mathbf{y}) - \bar{F}_{\mu}^*(\mathbf{y}) \right\| \\
 &= \left(O\left(\sqrt{\frac{\ln \ln n}{n}}\right) + \sum_{j=1}^d \sup_{z \in \mathbb{R}} \left| \hat{F}_{jn}(z) - F_j(z) \right| \right) \sum_{j=1}^d \sup_{z \in \mathbb{R}} \left| \hat{F}_{jn}(z) - F_j(z) \right| \\
 &= O\left(\frac{\ln \ln n}{n}\right) \text{ a.s.} \tag{12}
 \end{aligned}$$

by applying Proposition 3.1. Let $\tau_n := \max_{j=1, \dots, d} \sup_{z \in \mathbb{R}} \left| \hat{F}_{jn}(z) - F_j(z) \right|$. Observe that $\tilde{C}_{\mu l}^{\circ}(\cdot \mid \theta)$ is uniformly continuous on $[0, 1]^{d_{\mu}}$ for $\theta \in U(\theta_0)$ in view of the Heine-Cantor theorem. Further by Lemma 9.3 and the mean value theorem, we obtain

$$\begin{aligned}
 \|A_{3n} - A_{3n}^*\| &\leq \sum_{\mu=1}^m \frac{2}{n_{\mu}} \sum_{i=1}^{n_{\mu}} \sum_{l \in J_{\mu}} \left| \hat{F}_{nl}(\bar{\psi}_{l\mu}(\mathbf{Y}_{\mu i})) - F_l(\bar{\psi}_{l\mu}(\mathbf{Y}_{\mu i})) \right| \\
 & \quad \sup_{0 \leq \eta \leq 1} \left| \tilde{C}_{\mu l}^{\circ}(\bar{F}_{\mu}^*(\mathbf{Y}_{\mu i}) \mid \theta_0) - \tilde{C}_{\mu l}^{\circ}(\bar{F}_{\mu}^*(\mathbf{Y}_{\mu i})) \right. \\
 & \quad \left. + \eta \left(\check{F}_{n\mu}^*(\mathbf{Y}_{\mu i}) - \bar{F}_{\mu}^*(\mathbf{Y}_{\mu i}) \right) \mid \theta_0 \right| \\
 & \quad \sup_{\mathbf{u} \in [0, 1]^{d_{\mu}}} \left(\left\| \nabla_{\theta} C_{\mu}(\mathbf{u} \mid \theta_0) \right\| \mid w_{\mu}(\mathbf{u}) \right)
 \end{aligned}$$

$$\begin{aligned} &\leq O(\tau_n) \cdot \sum_{\mu=1}^m \sum_{l \in J_\mu} \sup_{\mathbf{u} \in [0,1]^{d_\mu}} \sup_{\bar{\eta}: \|\bar{\eta}\| \leq d\tau_n} \left| \tilde{C}_{\mu l}^\circ(\mathbf{u} \mid \theta_0) - \tilde{C}_{\mu l}^\circ(\mathbf{u} + \bar{\eta} \mid \theta_0) \right| \\ &= o_{\mathbb{P}}(n^{-1/2}). \end{aligned} \tag{13}$$

On the other hand by Lemma 9.3 and the mean value theorem, we derive $(\check{F}_{n\mu\eta}^{**}(y) := \bar{F}_\mu^*(y) + \eta (\check{F}_{n\mu}^*(y) - \bar{F}_\mu^*(y)))$ for $y \in \mathbb{R}^{d_\mu}$

$$\begin{aligned} &|A_{4nk} - A_{4nk}^*| \\ &\leq \sum_{\mu=1}^m \frac{2}{n_\mu} \sum_{i=1}^{n_\mu} \left(\sup_{\eta: 0 \leq \eta \leq 1} \left| \tilde{C}_{\mu lk}^\circ(\bar{F}_\mu^*(\mathbf{Y}_{\mu i}) \mid \theta_0) w_\mu(\bar{F}_\mu^*(\mathbf{Y}_{\mu i})) - \tilde{C}_{\mu lk}^\circ(\check{F}_{n\mu\eta}^{**}(\mathbf{Y}_{\mu i}) \mid \theta_0) w_\mu(\check{F}_{n\mu\eta}^{**}(\mathbf{Y}_{\mu i})) \right| \right. \\ &\quad \left. + \sup_{\eta: 0 \leq \eta \leq 1} \left| \tilde{C}_{\mu k}^\circ(\bar{F}_\mu^*(\mathbf{Y}_{\mu i}) \mid \theta_0) w_{\mu l}(\bar{F}_\mu^*(\mathbf{Y}_{\mu i})) - \tilde{C}_{\mu k}^\circ(\check{F}_{n\mu\eta}^{**}(\mathbf{Y}_{\mu i}) \mid \theta_0) w_{\mu l}(\check{F}_{n\mu\eta}^{**}(\mathbf{Y}_{\mu i})) \right| \right) \\ &\quad \left| \hat{F}_l(\check{\psi}_{l\mu}(\mathbf{Y}_{\mu i})) - F_l(\bar{\psi}_{l\mu}(\mathbf{Y}_{\mu i})) \right| \\ &\leq O(\tau_n) \sum_{\mu=1}^m \max_{l \in J_\mu} \left(\sup_{\mathbf{u} \in [0,1]^{d_\mu}} \sup_{\bar{\eta}: \|\bar{\eta}\| \leq d\tau_n} \left(\left| \tilde{C}_{\mu lk}^\circ(\mathbf{u} \mid \theta_0) - \tilde{C}_{\mu lk}^\circ(\mathbf{u} + \bar{\eta} \mid \theta_0) \right| + |w_\mu(\mathbf{u}) - w_\mu(\mathbf{u} + \bar{\eta})| \right) \right. \\ &\quad \left. + \sup_{\mathbf{u} \in [0,1]^{d_\mu}} \sup_{\bar{\eta}: \|\bar{\eta}\| \leq d\tau_n} \left(\left| \tilde{C}_{\mu k}^\circ(\mathbf{u} \mid \theta_0) - \tilde{C}_{\mu k}^\circ(\mathbf{u} + \bar{\eta} \mid \theta_0) \right| + |w_{\mu l}(\mathbf{u}) - w_{\mu l}(\mathbf{u} + \bar{\eta})| \right) \right) \\ &= o_{\mathbb{P}}(n^{-1/2}) \quad (k = 1, \dots, q) \end{aligned} \tag{14}$$

since $w_\mu, w_{\mu l}, \tilde{C}_{\mu lk}^\circ(\cdot \mid \theta_0)$ and $\tilde{C}_{\mu k}^\circ(\cdot \mid \theta_0)$ are uniformly continuous on $[0, 1]^{d_\mu}$. Identities (12)-(14) imply

$$A_{2n} + (A_{3n} - A_{3n}^*) + (A_{4n} - A_{4n}^*) = o_{\mathbb{P}}(n^{-1/2}). \tag{15}$$

In the remaining part of the proof, we show the asymptotic normality of $\sqrt{n}A_n$, where $A_n = A_{1n} + A_{3n}^* + A_{4n}^*$. We have

$$\begin{aligned} A_n &= A_{1n} \\ &\quad + \sum_{\mu=1}^m \frac{2}{n_\mu} \sum_{i=1}^{n_\mu} \nabla_{\theta} C_\mu(\bar{F}_\mu^*(\mathbf{Y}_{\mu i}) \mid \theta_0) w_\mu(\bar{F}_\mu^*(\mathbf{Y}_{\mu i})) \\ &\quad \left(\frac{1}{\bar{n}_\mu} \sum_{v: 1 \leq v \leq m, \mathbf{b}^{(v)} \geq \mathbf{b}^{(\mu)}} \sum_{j=1}^{n_\mu} (\mathbf{1}\{\psi_{\mu v}(\mathbf{Y}_{v j}) \leq \mathbf{Y}_{\mu i}\}) - H_\mu(\mathbf{Y}_{\mu i}) \right) \\ &\quad - \sum_{l \in J_\mu} \tilde{C}_{\mu l}^\circ(\bar{F}_\mu^*(\mathbf{Y}_{\mu i}) \mid \theta_0) \end{aligned}$$

$$\begin{aligned} & \left. \frac{1}{\tilde{n}_l} \sum_{v:1 \leq v \leq m, l \in J_v} \sum_{j=1}^{n_v} (\mathbf{1}\{\tilde{\psi}_{lv}(\mathbf{Y}_{vj}) \leq \tilde{\psi}_{l\mu}(\mathbf{Y}_{\mu i})\} - F_l(\tilde{\psi}_{l\mu}(\mathbf{Y}_{\mu i}))) \right) \\ & + \sum_{\mu=1}^m \frac{2}{n_\mu} \sum_{i=1}^{n_\mu} (H_\mu(\mathbf{Y}_{\mu i}) - C_\mu(\bar{F}_\mu^*(\mathbf{Y}_{\mu i}) \mid \theta_0)) \\ & \sum_{l \in J_\mu} (\tilde{C}_{\mu l k}^\circ(\bar{F}_\mu^*(\mathbf{Y}_{\mu i}) \mid \theta_0) w_\mu(\bar{F}_\mu^*(\mathbf{Y}_{\mu i})) + \bar{C}_{\mu k}^\circ(\bar{F}_\mu^*(\mathbf{Y}_{\mu i}) \mid \theta_0) w_{\mu l}(\bar{F}_\mu^*(\mathbf{Y}_{\mu i})))_{k=1, \dots, q} \\ & \frac{1}{\tilde{n}_l} \sum_{v:1 \leq v \leq m, l \in J_v} \sum_{j=1}^{n_v} (\mathbf{1}\{\tilde{\psi}_{lv}(\mathbf{Y}_{vj}) \leq \tilde{\psi}_{l\mu}(\mathbf{Y}_{\mu i})\} - F_l(\tilde{\psi}_{l\mu}(\mathbf{Y}_{\mu i}))) \\ & = 2 \sum_{\mu=1}^m \sum_{i=1}^{n_\mu} \sum_{v=1}^m \sum_{j=1}^{n_v} \\ & \left(\frac{1}{n} \frac{1}{n_\mu} \Lambda_\mu^{(1)}(\mathbf{Y}_{\mu i}) + \frac{1}{n_\mu} \frac{1}{\tilde{n}_\mu} \Lambda_{\mu v}^{(2)}(\mathbf{Y}_{\mu i}, \mathbf{Y}_{vj}) + \sum_{l \in J_\mu \cap J_v} \frac{1}{n_\mu} \frac{1}{\tilde{n}_l} \Lambda_{\mu vl}^{(3)}(\mathbf{Y}_{\mu i}, \mathbf{Y}_{vj}) \right), \end{aligned}$$

where

$$\begin{aligned} \Lambda_\mu^{(1)}(\mathbf{y}) &= (H_\mu(\mathbf{y}) - C_\mu(\bar{F}_\mu^*(\mathbf{y}) \mid \theta_0)) \nabla_\theta C_\mu(\bar{F}_\mu^*(\mathbf{y}) \mid \theta_0) w_\mu(\bar{F}_\mu^*(\mathbf{y})), \\ \Lambda_{\mu v}^{(2)}(\mathbf{y}, \mathbf{z}) &= \begin{cases} (\mathbf{1}\{\psi_{\mu v}(\mathbf{z}) \leq \mathbf{y}\} - H_\mu(\mathbf{y})) \nabla_\theta C_\mu(\bar{F}_\mu^*(\mathbf{y}) \mid \theta_0) w_\mu(\bar{F}_\mu^*(\mathbf{y})) & \text{for } v : \mathbf{b}^{(v)} \geq \mathbf{b}^{(\mu)}, \\ 0 & \text{otherwise,} \end{cases} \\ \Lambda_{\mu vl}^{(3)}(\mathbf{y}, \mathbf{z}) &= (-\tilde{C}_{\mu l}^\circ(\bar{F}_\mu^*(\mathbf{y}) \mid \theta_0) \nabla_\theta C_\mu(\bar{F}_\mu^*(\mathbf{y}) \mid \theta_0) w_\mu(\bar{F}_\mu^*(\mathbf{y})) \\ & + (H_\mu(\mathbf{y}) - C_\mu(\bar{F}_\mu^*(\mathbf{y}) \mid \theta_0)) \\ & (\tilde{C}_{\mu l k}^\circ(\bar{F}_\mu^*(\mathbf{y}) \mid \theta_0) w_\mu(\bar{F}_\mu^*(\mathbf{y})) + \bar{C}_{\mu k}^\circ(\bar{F}_\mu^*(\mathbf{y}) \mid \theta_0) w_{\mu l}(\bar{F}_\mu^*(\mathbf{y})))_{k=1, \dots, q} \\ & (\mathbf{1}\{\tilde{\psi}_{lv}(\mathbf{z}) \leq \tilde{\psi}_{l\mu}(\mathbf{y})\} - F_l(\tilde{\psi}_{l\mu}(\mathbf{y}))). \end{aligned}$$

Now we decompose $\sqrt{n}A_n$ such that we obtain

$$\sqrt{n}A_n = n^{-3/2} \sum_{\mu=1}^m \sum_{i=1}^{n_\mu} \sum_{v=1}^m \sum_{j=1}^{n_v} \Lambda_{\mu v}(\mathbf{Y}_{\mu i}, \mathbf{Y}_{vj}) + \sqrt{n}\bar{A}_n, \tag{16}$$

where

$$\Lambda_{\mu v}(\mathbf{y}, \mathbf{z}) = \frac{2}{\gamma_\mu} \left(\Lambda_{\mu v}^{(1)}(\mathbf{y}) + \frac{1}{\tilde{\gamma}_\mu} \Lambda_{\mu v}^{(2)}(\mathbf{y}, \mathbf{z}) + \sum_{l=1}^d \frac{b_l^{(\mu)} b_l^{(v)}}{\tilde{\gamma}_l} \Lambda_{\mu vl}^{(3)}(\mathbf{y}, \mathbf{z}) \right),$$

and

$$\bar{A}_n = n^{-3/2} \sum_{\mu=1}^m \sum_{i=1}^{n_\mu} \sum_{v=1}^m \sum_{j=1}^{n_v} \left(\left(\frac{n}{n_\mu} - \frac{1}{\gamma_\mu} \right) \Lambda_\mu^{(1)}(\mathbf{Y}_{\mu i}) \right)$$

$$\begin{aligned}
 &+ \left(\frac{n^2}{n_\mu \bar{n}_\mu} - \frac{1}{\gamma_\mu \bar{\gamma}_\mu} \right) \Lambda_{\mu\nu}^{(2)}(\mathbf{Y}_{\mu i}, \mathbf{Y}_{\nu j}) \\
 &+ \sum_{l \in J_\mu \cap J_\nu} \left(\frac{n^2}{n_\mu \bar{n}_l} - \frac{1}{\gamma_\mu \bar{\gamma}_l} \right) \Lambda_{\mu\nu l}^{(3)}(\mathbf{Y}_{\mu i}, \mathbf{Y}_{\nu j}) \Big).
 \end{aligned}$$

In the following we prove that $\bar{A}_n = o_{\mathbb{P}}(n^{-1/2})$. Note that $\mathbb{E}\Lambda_\mu^{(1)}(\mathbf{Y}_{\mu i}) = 0$. We have

$$\mathbb{V} \left(\frac{1}{n^{1/2}} \sum_{i=1}^{n_\mu} \left(\Lambda_\mu^{(1)}(\mathbf{Y}_{\mu i}) \right)_k \right) = \frac{n_\mu}{n} \mathbb{V} \left(\left(\Lambda_\mu^{(1)}(\mathbf{Y}_{\mu i}) \right)_k \right) = O(1) \quad (k = 1, \dots, q).$$

Moreover, in view of Proposition 9.5,

$$\begin{aligned}
 &\frac{1}{n^{3/2}} \sum_{i=1}^{n_\mu} \sum_{j=1}^{n_\nu} \left(\Lambda_{\mu\nu}^{(2)}(\mathbf{Y}_{\mu i}, \mathbf{Y}_{\nu j}) \right)_k \xrightarrow{d} \mathcal{N}(0, \Sigma_{\mu\nu k}), \text{ and} \\
 &\frac{1}{n^{3/2}} \sum_{i=1}^{n_\mu} \sum_{j=1}^{n_\nu} \left(\Lambda_{\mu\nu l}^{(3)}(\mathbf{Y}_{\mu i}, \mathbf{Y}_{\nu j}) \right)_k \xrightarrow{d} \mathcal{N}(0, \Sigma_{\mu\nu k l})
 \end{aligned}$$

with certain (finite) covariance matrices $\Sigma_{\mu\nu k}, \Sigma_{\mu\nu k l}$ ($k = 1, \dots, q, l \in J_\mu \cap J_\nu, \mu, \nu = 1, \dots, m$). Hence by Assumption \mathcal{A}_n ,

$$\bar{A}_n = o_{\mathbb{P}}(n^{-1/2}),$$

and by equations (11), (15), (16),

$$\sqrt{n} \nabla_\theta \widehat{D}_n(\theta_0) = -n^{-3/2} \sum_{\mu=1}^m \sum_{i=1}^{n_\mu} \sum_{\nu=1}^m \sum_{j=1}^{n_\nu} \Lambda_{\mu\nu}(\mathbf{Y}_{\mu i}, \mathbf{Y}_{\nu j}) + o_{\mathbb{P}}(1). \tag{17}$$

An application of the central limit theorem in Proposition 9.5 gives the asymptotic normality of $\sqrt{n} \nabla_\theta \widehat{D}_n(\theta_0)$. To derive a formula for the covariance matrix, we consider

$$\begin{aligned}
 \tilde{h}_{\mu\nu}(\mathbf{y}) &:= \mathbb{E}\Lambda_{\mu\nu}(\mathbf{Y}_{\mu 1}, \mathbf{y}) + \mathbb{E}\Lambda_{\nu\mu}(\mathbf{y}, \mathbf{Y}_{\mu 1}) - \mathbb{E}\Lambda_{\mu\nu}(\mathbf{Y}_{\mu 1}, \mathbf{Y}_{\nu 2}) - \mathbb{E}\Lambda_{\nu\mu}(\mathbf{Y}_{\nu 1}, \mathbf{Y}_{\mu 2}) \\
 &= \frac{2}{\gamma_\mu} \left((H_\mu(\mathbf{y}) - \mathcal{C}_\mu(\bar{F}_\mu^*(\mathbf{y}) \mid \theta_0)) \nabla_\theta \mathcal{C}_\mu(\bar{F}_\mu^*(\mathbf{y}) \mid \theta_0) w_\mu(\bar{F}_\mu^*(\mathbf{y})) - \mathbf{e}_\mu \right) \\
 &\quad + \frac{1}{\gamma_\mu} \mathbf{1}(\mathbf{b}^{(\nu)} \geq \mathbf{b}^{(\mu)}) \\
 &\quad + \int_{\mathbb{R}^{d_\mu}} \left(\mathbf{1}(\psi_{\mu\nu}(\mathbf{y}) \leq \mathbf{z}) - H_\mu(\mathbf{z}) \right) \nabla_\theta \mathcal{C}_\mu(\bar{F}_\mu^*(\mathbf{z}) \mid \theta_0) w_\mu(\bar{F}_\mu^*(\mathbf{z})) \, dH_\mu(\mathbf{z}) \\
 &\quad + \sum_{l=1}^d \frac{1}{\tilde{\gamma}_l} b_l^{(\mu)} b_l^{(\nu)} \int_{\mathbb{R}^{d_\mu}} \left(-\tilde{\mathcal{C}}_{\mu l}^\circ(\bar{F}_\mu^*(\mathbf{z}) \mid \theta_0) \nabla_\theta \mathcal{C}_\mu(\bar{F}_\mu^*(\mathbf{z}) \mid \theta_0) w_\mu(\bar{F}_\mu^*(\mathbf{z})) \right)
 \end{aligned}$$

$$\begin{aligned}
 &+ (H_\mu(\mathbf{z}) - C_\mu(\bar{F}_\mu^*(\mathbf{z}) \mid \theta_0)) \cdot \\
 &\cdot \left(\bar{C}_{\mu lk}^\circ(\bar{F}_\mu^*(\mathbf{z}) \mid \theta_0) w_{\mu k}(\bar{F}_\mu^*(\mathbf{z})) + \bar{C}_{\mu k}^\circ(\bar{F}_\mu^*(\mathbf{z}) \mid \theta_0) w_{\mu l}(\bar{F}_\mu^*(\mathbf{z})) \right)_{k=1, \dots, q} \\
 &(\mathbf{1}\{\bar{\psi}_{lv}(\mathbf{y}) \leq \bar{\psi}_{l\mu}(\mathbf{z})\} - F_l(\bar{\psi}_{l\mu}(\mathbf{z}))) \, dH_\mu(\mathbf{z}),
 \end{aligned}$$

where

$$\mathbf{e}_\mu = \frac{2}{\gamma_\mu} \int_{\mathbb{R}^{d_\mu}} (H_\mu(\mathbf{y}) - C_\mu(\bar{F}_\mu^*(\mathbf{y}) \mid \theta_0)) \nabla_\theta C_\mu(\bar{F}_\mu^*(\mathbf{y}) \mid \theta_0) w_\mu(\bar{F}_\mu^*(\mathbf{y})) dH_\mu(\mathbf{y}).$$

By Proposition 9.5 we obtain the formula for the covariance matrix Σ_D . □

Next we deal with the convergence of \mathcal{H}_n^* and prove the following lemma.

Lemma 9.7 *Suppose that $t_{nk}^* \rightarrow \theta_0$ for $k = 1, \dots, q$. Then we have*

$$\mathcal{H}_{nkl}(t_{nk}^*) \longrightarrow \mathcal{H}_{kl} \text{ a.s.}$$

for $k, l = 1, \dots, q$.

Proof Notice that $\bar{\tau}_n := \sup_{y \in \mathbb{R}^{d_\mu}} \|\check{F}_{n\mu}^*(y) - \bar{F}_\mu^*(y)\| \rightarrow 0$ a.s. Moreover, $C_\mu, \bar{C}_{\mu k}$, and $\bar{C}_{\mu kl}$ are uniformly continuous on $[0, 1]^{d_\mu} \times U(\theta_0)$ in view of the Heine-Cantor theorem, and we obtain

$$\begin{aligned}
 \zeta_{nk} &:= \max_{\mu=1, \dots, m} \sup_{\mathbf{y}_1, \mathbf{y}_2: \|\mathbf{y}_1 - \mathbf{y}_2\| \leq \bar{\tau}_n} |C_\mu(\mathbf{y}_1 \mid t_{nk}^*) - C_\mu(\mathbf{y}_2 \mid \theta_0)| \rightarrow 0 \text{ a.s.}, \\
 \bar{\zeta}_{nk} &:= \max_{\mu=1, \dots, m} \sup_{\mathbf{y}_1, \mathbf{y}_2: \|\mathbf{y}_1 - \mathbf{y}_2\| \leq \bar{\tau}_n} \left| \bar{C}_{\mu k}^\circ(\mathbf{y}_1 \mid t_{nk}^*) - \bar{C}_{\mu k}^\circ(\mathbf{y}_2 \mid \theta_0) \right| \rightarrow 0 \text{ a.s.}, \\
 \bar{\zeta}_{nkl} &:= \max_{\mu=1, \dots, m} \sup_{\mathbf{y}_1, \mathbf{y}_2: \|\mathbf{y}_1 - \mathbf{y}_2\| \leq \bar{\tau}_n} \left| \bar{C}_{\mu kl}^\circ(\mathbf{y}_1 \mid t_{nk}^*) - \bar{C}_{\mu kl}^\circ(\mathbf{y}_2 \mid \theta_0) \right| \rightarrow 0 \text{ a.s.}, \\
 \bar{\zeta}_n &:= \max_{\mu=1, \dots, m} \sup_{\mathbf{y}_1, \mathbf{y}_2: \|\mathbf{y}_1 - \mathbf{y}_2\| \leq \bar{\tau}_n} |w_\mu(\mathbf{y}_1) - w_\mu(\mathbf{y}_2)| \rightarrow 0 \text{ a.s.}
 \end{aligned}$$

for $k, l = 1, \dots, q$. We define

$$\begin{aligned}
 \check{\mathcal{H}}_{nkl} &= \sum_{\mu=1}^m \frac{2}{n_\mu} \sum_{i=1}^{n_\mu} \left(- (H_\mu(\mathbf{Y}_{\mu i}) - C_\mu(\bar{F}_\mu^*(\mathbf{Y}_{\mu i}) \mid \theta_0)) \bar{C}_{\mu kl}^\circ(\bar{F}_\mu^*(\mathbf{Y}_{\mu i}) \mid \theta_0) \right. \\
 &\quad \left. + \bar{C}_{\mu k}^\circ(\bar{F}_\mu^*(\mathbf{Y}_{\mu i}) \mid \theta_0) \bar{C}_{\mu l}^\circ(\bar{F}_\mu^*(\mathbf{Y}_{\mu i}) \mid \theta_0) \right) w_\mu(\bar{F}_\mu^*(\mathbf{Y}_{\mu i})).
 \end{aligned}$$

Therefore, and by Theorem 3.1, we have

$$\max_{k, l=1, \dots, q} \left| \mathcal{H}_{nkl}(t_{nk}^*) - \check{\mathcal{H}}_{nkl} \right|$$

$$\begin{aligned}
 &\leq O(1) \sum_{\mu=1}^m \frac{1}{n_\mu} \sum_{i=1}^{n_\mu} \max_{k,l=1,\dots,q} \left(\left| \hat{H}_{n\mu}(\mathbf{Y}_{\mu i}) - \mathcal{C}_\mu(\check{F}_{n\mu}^*(\mathbf{Y}_{\mu i}) \mid t_{nk}^*) \right. \right. \\
 &\quad \left. \left. - H_\mu(\mathbf{Y}_{\mu i}) + \mathcal{C}_\mu(\bar{F}_\mu^*(\mathbf{Y}_{\mu i}) \mid \theta_0) \right| \right. \\
 &\quad \left. + \left| \bar{\mathcal{C}}_{\mu kl}^\circ(\check{F}_{n\mu}^*(\mathbf{Y}_{\mu i}) \mid t_{nk}^*) - \bar{\mathcal{C}}_{\mu kl}^\circ(\bar{F}_\mu^*(\mathbf{Y}_{\mu i}) \mid \theta_0) \right| \right. \\
 &\quad \left. + \left| w_\mu(\check{F}_{n\mu}^*(\mathbf{Y}_{\mu i})) - w_\mu(\bar{F}_\mu^*(\mathbf{Y}_{\mu i})) \right| \right. \\
 &\quad \left. + \left| \bar{\mathcal{C}}_{\mu k}^\circ(\check{F}_{n\mu}^*(\mathbf{Y}_{\mu i}) \mid t_{nk}^*) - \bar{\mathcal{C}}_{\mu k}^\circ(\bar{F}_\mu^*(\mathbf{Y}_{\mu i}) \mid \theta_0) \right| \right. \\
 &\quad \left. + \left| \bar{\mathcal{C}}_{\mu l}^\circ(\check{F}_{n\mu}^*(\mathbf{Y}_{\mu i}) \mid t_{nk}^*) - \bar{\mathcal{C}}_{\mu l}^\circ(\bar{F}_\mu^*(\mathbf{Y}_{\mu i}) \mid \theta_0) \right| \right) \\
 &\leq O(1) \left(O\left(\sqrt{\frac{\ln \ln n}{n}}\right) + \max_{k,l=1,\dots,q} (\zeta_{nk} + \bar{\zeta}_{nkl} + \bar{\zeta}_{nk} + \bar{\zeta}_{nl}) + \bar{\zeta}_n \right) \\
 &= o(1) \text{ a.s.}
 \end{aligned}$$

In view of the law of large numbers, we have

$$\begin{aligned}
 \check{\mathcal{H}}_{nkl} &\longrightarrow \mathcal{H}_{kl} := 2 \sum_{\mu=1}^m \int_{\mathbb{R}^{d_\mu}} \left(- (H_\mu(\mathbf{y}) - \mathcal{C}_\mu(\bar{F}_\mu^*(\mathbf{y}) \mid \theta_0)) \bar{\mathcal{C}}_{\mu kl}^\circ(\bar{F}_\mu^*(\mathbf{y}) \mid \theta_0) \right. \\
 &\quad \left. + \bar{\mathcal{C}}_{\mu k}^\circ(\bar{F}_\mu^*(\mathbf{y}) \mid \theta_0) \bar{\mathcal{C}}_{\mu l}^\circ(\bar{F}_\mu^*(\mathbf{y}) \mid \theta_0) \right) w_\mu(\bar{F}_\mu^*(\mathbf{y})) \, dH_\mu(\mathbf{y})
 \end{aligned}$$

a.s. for $k, l = 1, \dots, q$. This completes the proof. □

Now we are in a position to prove Theorem 5.2.

Proof of Theorem 5.2 First we derive the convergence rate of $\tilde{\theta}_n - \hat{\theta}_n$. Note that $\nabla_{\tilde{\theta}} \widehat{\mathcal{D}}_n(\tilde{\theta}_n) = 0$, and θ_0 is an interior point of Θ . Then the Taylor formula leads to

$$\widehat{\mathcal{D}}_n(\tilde{\theta}_n) - \widehat{\mathcal{D}}_n(\hat{\theta}_n) = (\tilde{\theta}_n - \hat{\theta}_n)^T \mathcal{H}_n^{**} (\tilde{\theta}_n - \hat{\theta}_n),$$

where $t_{nk}^{**} = \tilde{\theta}_n + \eta_{nk} (\hat{\theta}_n - \tilde{\theta}_n)$, $0 \leq \eta_{nk} \leq 1$ and $\mathcal{H}_n^{**} = (\mathcal{H}_{nkl}(t_{nk}^{**}))_{k,l=1,\dots,q}$.

In view of Theorem 5.1, the estimators $\tilde{\theta}_n$ and $\hat{\theta}_n$ are strongly consistent such that $t_{nk}^{**} \rightarrow \theta_0$ a.s. Using Lemma 9.7, we obtain $\mathcal{H}_n^{**} \rightarrow \mathcal{H}$ a.s.. Hence

$$\begin{aligned}
 \varepsilon_n &\geq \left| (\tilde{\theta}_n - \hat{\theta}_n)^T \mathcal{H}_n^{**} (\tilde{\theta}_n - \hat{\theta}_n) \right| \geq \lambda_{\min}(\mathcal{H}_n^{**}) \left\| \tilde{\theta}_n - \hat{\theta}_n \right\|^2 \\
 &= (\lambda_{\min}(\mathcal{H}) + o_{\mathbb{P}}(1)) \left\| \tilde{\theta}_n - \hat{\theta}_n \right\|^2,
 \end{aligned}$$

where $\lambda_{\min}(A)$ is the smallest absolute eigenvalue of the matrix A . $\lambda_{\min}(\mathcal{H})$ is positive by assumption. Therefore, it follows that

$$\tilde{\theta}_n - \hat{\theta}_n = o_{\mathbb{P}}(n^{-1/2}). \tag{18}$$

Using (10) and Lemmas 9.6, 9.7, an application of Slutsky’s theorem leads to

$$\sqrt{n} \left(\hat{\theta}_n - \theta_0 \right) = -\sqrt{n} \mathcal{H}_n^{*-1} \nabla_{\theta} \widehat{\mathcal{D}}_n(\theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mathcal{H}^{-1} \Sigma_D \mathcal{H}^{-1}).$$

In view of (18), the proof of Theorem 5.2 is complete. □

9.4 Proof of Theorem 8.1

First we prove a lemma on asymptotic normality of $\sqrt{n} \left(\widehat{\mathcal{D}}_n(\theta_0) - \mathcal{D}(C, \mathcal{C}(\cdot | \theta_0)) \right)$, which is crucial for the proof of asymptotic normality of $\sqrt{n} \left(\widehat{\mathcal{D}}_n(\hat{\theta}_n) - \mathcal{D}(C, \mathcal{C}(\cdot | \theta_0)) \right)$ in Theorem 8.1.

Lemma 9.8 *Let the assumptions of Theorem 5.2 be satisfied. Then*

$$\sqrt{n} \left(\widehat{\mathcal{D}}_n(\theta_0) - \mathcal{D}(C, \mathcal{C}(\cdot | \theta_0)) \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma_0). \tag{19}$$

Proof Define

$$\bar{e}_{\mu} := \int_{\mathbb{R}^{d_{\mu}}} \left(H_{\mu}(\mathbf{y}) - C_{\mu}(\bar{F}_{\mu}^*(\mathbf{y}) | \theta_0) \right)^2 w_{\mu}(\bar{F}_{\mu}^*(\mathbf{y})) \, dH_{\mu}(\mathbf{y}).$$

We obtain

$$\begin{aligned} & \widehat{\mathcal{D}}_n(\theta_0) - \mathcal{D}(C, \mathcal{C}(\cdot | \theta_0)) \\ &= \sum_{\mu=1}^m \left(\frac{1}{n_{\mu}} \sum_{i=1}^{n_{\mu}} \left(\hat{H}_{n\mu}(\mathbf{Y}_{\mu i}) - C_{\mu}(\check{F}_{n\mu}^*(\mathbf{Y}_{\mu i}) | \theta_0) \right)^2 w_{\mu}(\check{F}_{n\mu}^*(\mathbf{Y}_{\mu i})) - \bar{e}_{\mu} \right) \\ &= B_{1n} + B_{2n} + B_{3n} + \bar{B}_{3n} + B_{4n}, \end{aligned}$$

where

$$\begin{aligned} B_{1n} &= \sum_{\mu=1}^m \frac{1}{n_{\mu}} \sum_{i=1}^{n_{\mu}} \left(\left(H_{\mu}(\mathbf{Y}_{\mu i}) - C_{\mu}(\bar{F}_{\mu}^*(\mathbf{Y}_{\mu i}) | \theta_0) \right)^2 w_{\mu}(\bar{F}_{\mu}^*(\mathbf{Y}_{\mu i})) - \bar{e}_{\mu} \right), \\ B_{2n} &= \sum_{\mu=1}^m \frac{1}{n_{\mu}} \sum_{i=1}^{n_{\mu}} \left(\left(\hat{H}_{n\mu}(\mathbf{Y}_{\mu i}) - C_{\mu}(\check{F}_{n\mu}^*(\mathbf{Y}_{\mu i}) | \theta_0) \right)^2 \right. \\ & \quad \left. - \left(H_{\mu}(\mathbf{Y}_{\mu i}) - C_{\mu}(\bar{F}_{\mu}^*(\mathbf{Y}_{\mu i}) | \theta_0) \right)^2 \right) \\ & \quad \left(w_{\mu}(\check{F}_{n\mu}^*(\mathbf{Y}_{\mu i})) - w_{\mu}(\bar{F}_{\mu}^*(\mathbf{Y}_{\mu i})) \right), \\ B_{3n} &= \sum_{\mu=1}^m \frac{1}{n_{\mu}} \sum_{i=1}^{n_{\mu}} \left(\hat{H}_{n\mu}(\mathbf{Y}_{\mu i}) - C_{\mu}(\check{F}_{n\mu}^*(\mathbf{Y}_{\mu i}) | \theta_0) - H_{\mu}(\mathbf{Y}_{\mu i}) + C_{\mu}(\bar{F}_{\mu}^*(\mathbf{Y}_{\mu i}) | \theta_0) \right)^2 \\ & \quad w_{\mu}(\bar{F}_{\mu}^*(\mathbf{Y}_{\mu i})), \end{aligned}$$

$$\begin{aligned} \bar{B}_{3n} &= \sum_{\mu=1}^m \frac{2}{n_\mu} \sum_{i=1}^{n_\mu} \left(\hat{H}_{n\mu}(\mathbf{Y}_{\mu i}) - C_\mu(\check{F}_{n\mu}^*(\mathbf{Y}_{\mu i}) \mid \theta_0) - H_\mu(\mathbf{Y}_{\mu i}) + C_\mu(\bar{F}_\mu^*(\mathbf{Y}_{\mu i}) \mid \theta_0) \right. \\ &\quad \left. (H_\mu(\mathbf{Y}_{\mu i}) - C_\mu(\bar{F}_\mu^*(\mathbf{Y}_{\mu i}) \mid \theta_0)) w_\mu(\bar{F}_\mu^*(\mathbf{Y}_{\mu i})), \right. \\ B_{4n} &= \sum_{\mu=1}^m \frac{1}{n_\mu} \sum_{i=1}^{n_\mu} (H_\mu(\mathbf{Y}_{\mu i}) - C_\mu(\bar{F}_\mu^*(\mathbf{Y}_{\mu i}) \mid \theta_0))^2 \left(w_\mu(\check{F}_{n\mu}^*(\mathbf{Y}_{\mu i})) - w_\mu(\bar{F}_\mu^*(\mathbf{Y}_{\mu i})) \right). \end{aligned}$$

Analogously to the proof of Lemma 9.6, we can derive

$$\begin{aligned} B_{2n} &= o_{\mathbb{P}}(n^{-1/2}), & B_{3n} &= o_{\mathbb{P}}(n^{-1/2}), \\ \bar{B}_{3n} &= B_{3n}^* + o_{\mathbb{P}}(n^{-1/2}), & B_{4n} &= B_{4n}^* + o_{\mathbb{P}}(n^{-1/2}), \end{aligned}$$

where

$$\begin{aligned} B_{3n}^* &= \sum_{\mu=1}^m \frac{2}{n_\mu} \sum_{i=1}^{n_\mu} \left(\hat{H}_{n\mu}(\mathbf{Y}_{\mu i}) - H_\mu(\mathbf{Y}_{\mu i}) \right. \\ &\quad \left. - \sum_{l \in J_\mu} \check{C}_{\mu l}^\circ(\bar{F}_\mu^*(\mathbf{Y}_{\mu i}) \mid \theta_0) \left(\hat{F}_{nl}(\check{\psi}_{l\mu}(\mathbf{Y}_{\mu i})) - F_l(\check{\psi}_{l\mu}(\mathbf{Y}_{\mu i})) \right) \right) \\ &\quad \left(H_\mu(\mathbf{Y}_{\mu i}) - C_\mu(\bar{F}_\mu^*(\mathbf{Y}_{\mu i}) \mid \theta_0) \right) w_\mu(\bar{F}_\mu^*(\mathbf{Y}_{\mu i})), \\ B_{4n}^* &= \sum_{\mu=1}^m \frac{1}{n_\mu} \sum_{i=1}^{n_\mu} (H_\mu(\mathbf{Y}_{\mu i}) - C_\mu(\bar{F}_\mu^*(\mathbf{Y}_{\mu i}) \mid \theta_0))^2 \\ &\quad \sum_{l \in J_\mu} w_{\mu l}(\bar{F}_\mu^*(\mathbf{Y}_{\mu i})) \left(\hat{F}_l(\check{\psi}_{l\mu}(\mathbf{Y}_{\mu i})) - F_l(\check{\psi}_{l\mu}(\mathbf{Y}_{\mu i})) \right). \end{aligned}$$

Further,

$$\sqrt{n} \left(\widehat{\mathcal{D}}_n(\theta_0) - \mathcal{D}(C, \mathcal{C}(\cdot \mid \theta_0)) \right) = n^{-3/2} \sum_{\mu=1}^m \sum_{i=1}^{n_\mu} \sum_{v=1}^m \sum_{j=1}^{n_v} \Lambda_{\mu v}(\mathbf{Y}_{\mu i}, \mathbf{Y}_{v j}) + o_{\mathbb{P}}(1),$$

where

$$\begin{aligned} \Lambda_{\mu v}(\mathbf{y}, \mathbf{z}) &= \left(\frac{1}{\gamma_\mu} \Lambda_{\mu v}^{(1)}(\mathbf{y}) + \frac{2}{\gamma_\mu \bar{\gamma}_\mu} \Lambda_{\mu v}^{(2)}(\mathbf{y}, \mathbf{z}) + \sum_{l=1}^d \frac{b_l^{(\mu)} b_l^{(v)}}{\gamma_\mu \bar{\gamma}_l} \Lambda_{\mu v l}^{(3)}(\mathbf{y}, \mathbf{z}) \right), \\ \Lambda_\mu^{(1)}(\mathbf{y}) &= \left(H_\mu(\mathbf{y}) - C_\mu(\bar{F}_\mu^*(\mathbf{y}) \mid \theta_0) \right)^2 w_\mu(\bar{F}_\mu^*(\mathbf{y})) - \bar{e}_\mu, \\ \Lambda_{\mu v}^{(2)}(\mathbf{y}, \mathbf{z}) &= \begin{cases} \mathbf{1}(\psi_{\mu v}(\mathbf{z}) \leq \mathbf{y}) - H_\mu(\mathbf{y}) \left(H_\mu(\mathbf{Y}_{\mu i}) - C_\mu(\bar{F}_\mu^*(\mathbf{Y}_{\mu i}) \mid \theta_0) \right) w_\mu(\bar{F}_\mu^*(\mathbf{y})) & \text{for } v: \mathbf{b}^{(v)} \geq \mathbf{b}^{(\mu)}, \\ 0 & \text{otherwise,} \end{cases} \\ \Lambda_{\mu v l}^{(3)}(\mathbf{y}, \mathbf{z}) &= \left(-2\check{C}_{\mu l}^\circ(\bar{F}_\mu^*(\mathbf{Y}_{\mu i}) \mid \theta_0) \left(H_\mu(\mathbf{Y}_{\mu i}) - C_\mu(\bar{F}_\mu^*(\mathbf{Y}_{\mu i}) \mid \theta_0) \right) w_\mu(\bar{F}_\mu^*(\mathbf{y})) \right) \end{aligned}$$

$$\begin{aligned}
 &+ \left(H_\mu(\mathbf{Y}_{\mu i}) - C_\mu(\bar{F}_\mu^*(\mathbf{Y}_{\mu i}) \mid \theta_0) \right)^2 w_{\mu l}(\bar{F}_\mu^*(\mathbf{y})) \\
 &(\mathbf{1} \{ \bar{\psi}_{l\nu}(\mathbf{z}) \leq \bar{\psi}_{l\mu}(\mathbf{y}) \} - F_l(\bar{\psi}_{l\mu}(\mathbf{y}))).
 \end{aligned}$$

The convergence of the " $o_{\mathbb{P}}(1)$ "-term is again proven analogously to the proof of Lemma 9.6. An application of the central limit theorem in Proposition 9.5 gives the asymptotic normality of $\sqrt{n}\nabla_{\theta}\widehat{D}_n(\theta_0)$. Eventually, we derive the formula for the covariance matrix as follows

$$\begin{aligned}
 \tilde{h}_{\mu\nu}(\mathbf{y}) &:= \mathbb{E}\Lambda_{\mu\nu}(\mathbf{Y}_{\mu 1}, \mathbf{y}) + \mathbb{E}\Lambda_{\nu\mu}(\mathbf{y}, \mathbf{Y}_{\mu 1}) - \mathbb{E}\Lambda_{\mu\nu}(\mathbf{Y}_{\mu 1}, \mathbf{Y}_{\nu 2}) - \mathbb{E}\Lambda_{\nu\mu}(\mathbf{Y}_{\nu 1}, \mathbf{Y}_{\mu 2}) \\
 &= \frac{1}{\gamma_\mu} \left((H_\mu(\mathbf{y}) - C_\mu(\bar{F}_\mu^*(\mathbf{y}) \mid \theta_0))^2 w_\mu(\bar{F}_\mu^*(\mathbf{y})) - \bar{e}_\mu \right) \\
 &+ \frac{2}{\gamma_\mu \tilde{\gamma}_\mu} \mathbf{1}(\mathbf{b}^{(\nu)} \geq \mathbf{b}^{(\mu)}) \\
 &\int_{\mathbb{R}^{d_\mu}} (\mathbf{1}(\psi_{\mu\nu}(\mathbf{y}) \leq \mathbf{z}) - H_\mu(\mathbf{z})) (H_\mu(\mathbf{z}) - C_\mu(\bar{F}_\mu^*(\mathbf{z}) \mid \theta_0)) \\
 &w_\mu(\bar{F}_\mu^*(\mathbf{z})) dH_\mu(\mathbf{z}) \\
 &+ \sum_{l=1}^d \frac{1}{\gamma_\mu \tilde{\gamma}_l} b_l^{(\mu)} b_l^{(\nu)} \int_{\mathbb{R}^{d_\mu}} \left(-2\tilde{C}_{\mu l}^\circ(\bar{F}_\mu^*(\mathbf{z}) \mid \theta_0) (H_\mu(\mathbf{z}) - C_\mu(\bar{F}_\mu^*(\mathbf{z}) \mid \theta_0)) \right. \\
 &w_\mu(\bar{F}_\mu^*(\mathbf{z})) \\
 &+ \left. (H_\mu(\mathbf{z}) - C_\mu(\bar{F}_\mu^*(\mathbf{z}) \mid \theta_0))^2 w_{\mu l}(\bar{F}_\mu^*(\mathbf{z})) \right) \\
 &(\mathbf{1} \{ \bar{\psi}_{l\nu}(\mathbf{y}) \leq \bar{\psi}_{l\mu}(\mathbf{z}) \} - F_l(\bar{\psi}_{l\mu}(\mathbf{z}))) dH_\mu(\mathbf{z}).
 \end{aligned}$$

Moreover, we have

$$\Sigma_0 = \sum_{\nu=1}^m \gamma_\nu \sum_{\mu=1}^m \sum_{\tilde{\mu}=1}^m \gamma_\mu \gamma_{\tilde{\mu}} \mathbb{E}\tilde{h}_{\mu\nu}(\mathbf{Y}_{\nu 1}) \tilde{h}_{\tilde{\mu}\nu}^T(\mathbf{Y}_{\nu 1}).$$

□

Proof of Theorem 8.1 Analogously to (18), we have

$$\widehat{D}_n(\tilde{\theta}_n) - \widehat{D}_n(\theta_0) = (\tilde{\theta}_n - \theta_0)^T \mathcal{H}_n^\# (\tilde{\theta}_n - \theta_0),$$

where $t_{nk}^\# = \tilde{\theta}_n + \eta_{nk}(\theta_0 - \tilde{\theta}_n)$, $0 \leq \eta_{nk} \leq 1$ and $\mathcal{H}_n^\# = (\mathcal{H}_{nkl}(t_{nk}^\#))_{k,l=1,\dots,q}$. An application of Theorem 5.2 leads to

$$\widehat{D}_n(\tilde{\theta}_n) - \widehat{D}_n(\theta_0) = O_{\mathbb{P}}(n^{-1}).$$

An application of Lemma 9.8 completes the proof.

□

Acknowledgements The author would like to thank the two anonymous reviewers for their valuable hints leading to improvements of the paper.

Funding Open Access funding enabled and organized by Projekt DEAL.

Declarations

Conflict of interest The author has no relevant financial or non-financial interests to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Chen X, Fan Y (2005) Pseudo-likelihood ratio tests for semiparametric multivariate copula model selection. *Can J Stat* 33:389–414. <https://doi.org/10.1002/cjs.5540330306>
- Di Lascio FML, Giannerini S, Reale A (2015) Exploring copulas for the imputation of complex dependent data. *Stat Methods Appl* 24:159–175. <https://doi.org/10.1007/s10260-014-0287-2>
- Genest C, Rivest L-P (1993) Statistical inference procedures for bivariate Archimedean copulas. *J Am Stat Assoc* 88:1034–1043. <https://doi.org/10.1080/01621459.1993.10476372>
- Genest C, Ghoudi K, Rivest L-P (1995) A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika* 82:543–552. <https://doi.org/10.1093/biomet/82.3.543>
- Genest C, Rémillard B, Beaudoin D (2009) Goodness-of-fit tests for copulas: a review and a power study. *Insur Math Econ* 44:199–213. <https://doi.org/10.1016/j.insmatheco.2007.10.005>
- Graham JW (2009) Missing data analysis: making it work in the real world. *Annu Rev Psychol* 60:549–576. <https://doi.org/10.1146/annurev.psych.58.110405.085530>
- Hamori S, Motegi K, Zhang Z (2019) Calibration estimation of semiparametric copula models with data missing at random. *J Multivariate Anal* 173:85–109. <https://doi.org/10.1016/j.jmva.2019.02.003>
- Hofert M, Mächler M, McNeil AJ (2012) Likelihood inference for Archimedean copulas in high dimensions under known margins. *J Multivariate Anal* 110:133–150. <https://doi.org/10.1016/j.jmva.2012.02.019>
- Joe H (1997) *Multivariate models and dependence concepts*. Chapman & Hall, London
- Joe H (2005) Asymptotic efficiency of the two-stage estimation method for copula-based models. *J Multivariate Anal* 94:401–419. <https://doi.org/10.1016/j.jmva.2004.06.003>
- Kattge J, Bönnisch G, Díaz S, Lavorel S, Prentice IC, Leadley P, Tautenhahn S, Werner GDA, Aakala T, Abedi M, Acosta ATR, Adamidis GC, Adamson K, Aiba M, Albert CH, Alcántara JM, Alcázar CC, Aleixo I, Ali H, Wirth C (2020) TRY plant trait database-enhanced coverage and open access. *Glob Change Biol* 26:119–188. <https://doi.org/10.1111/gcb.14904>
- Kertel M, Pauly M (2022) Estimating Gaussian copulas with missing data with and without expert knowledge. *Entropy* 24:1849. <https://doi.org/10.3390/e24121849>
- Kiefer J (1961) On large deviations of the empiric D. F. of vector chance variables and a law of the iterated logarithm. *Pac J Math* 11:649–660
- Lachout P, Liebscher E, Vogel S (2005) Strong convergence of estimators as ε_n -minimisers of optimisation problems. *Ann Inst Stat Math* 57:291–313. <https://doi.org/10.1007/BF02507027>
- Liebscher E (2008) Construction of asymmetric multivariate copulas. *J Multivariate Anal* 99:2234–2250. <https://doi.org/10.1016/j.jmva.2008.02.025>
- Liebscher E (2009) Semiparametric estimation of the parameters of multivariate copulas. *Kybernetika* 6:972–991

- Liebscher E (2015) Goodness-of-approximation of copulas by a parametric family. In: Stochastic models, statistics and their applications. Springer Proceedings in Mathematics & Statistics, vol 122, pp 101–109. https://doi.org/10.1007/978-3-319-13881-7_12
- Liebscher E, Taubert F, Waltschew D, Hetzer J (2022) Modelling multivariate data using product copulas and minimum distance estimators: an exemplary application to ecological traits. *Environ Ecol Stat* 29:315–338. <https://doi.org/10.1007/s10651-021-00525-0>
- Little RJA, Rubin DB (2019) Statistical analysis with missing data, 3rd edn. Wiley, Hoboken
- Medovikov I (2016) Non-parametric weighted tests for independence based on empirical copula process. *J Stat Comput Simul* 86:105–121. <https://doi.org/10.1080/00949655.2014.995657>
- Nelsen RB (2006) An introduction to copulas. Lecture notes in statistics, vol 139, 2nd edn. Springer, Berlin
- Rodriguez JC, Viollaz AJ (1995) A Cramer–von Mises type goodness of fit test with asymmetric weight function. *Commun Stat* 24:1095–1120. <https://doi.org/10.1080/03610929508831542>
- Serfling RJ (1980) Approximation theorems of mathematical statistics. Wiley, New York
- Sklar A (1959) Fonctions de répartition à n dimensions et leurs marges. *Publ Inst Stat Univ Paris* 8:229–231
- Tsukahara H (2005) Semiparametric estimation in copula models. *Can J Stat* 33:357–375. <https://doi.org/10.1002/cjs.5540330304>
- Van Buuren S, Brand JPL, Groothuis-Oudshoorn CGM, Rubin DB (2006) Fully conditional specification in multivariate imputation. *J Stat Comput Simul* 76:1049–1064. <https://doi.org/10.1080/10629360600810434>
- Van der Vaart AW (1998) Asymptotic statistics. Cambridge University Press, Cambridge
- Wang H, Fazayeli F, Chatterjee S, Banerjee A (2014) Gaussian copula precision estimation with missing values. International Conference on Artificial Intelligence and Statistics 2014

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.