Dept. of Math. University of Oslo
Statistical Research Report No. 7
ISSN 0806–3842 June 2008

THE COPULA INFORMATION CRITERIA

STEFFEN GRØNNEBERG AND NILS LID HJORT

Abstract. When estimating parametric copula models by the semiparametric pseudo maximum likelihood procedure (MPLE), many practitioners have used the Akaike Information Criterion (AIC) for model selection in spite of the fact that the AIC formula has no theoretical basis in this setting. We adapt the arguments leading to the original AIC formula in the fully parametric case to the MPLE. This gives a significantly different formula than the AIC, which we name the Copula Information Criterion (CIC). However, we also show that such a model-selection procedure cannot exist for a large class of commonly used copula models.

We note that this research report is a revision of a research report dated June 2008. The current version encorporates corrections of the proof of Theorem 1. The conclusions of the previous manuscript are still valid, however.

1. Introduction and summary

Suppose given independent, identically distributed d-dimensional observations $X_1, X_2, \ldots, X_n$ with density $f^\circ(x)$ and distribution function

$$F^\circ(x) = P(X_{i,1} \le x_1, X_{i,2} \le x_2, \ldots X_{i,d} \le x_d) = C^\circ(F_\perp^\circ(x)).$$

Here, C° is the copula of F° and $F_\perp^\circ$ is the vector of marginal distributions of F°, that is,

$$F_\perp^\circ(x) := (F_1^\circ(x_1), \ldots, F_d^\circ(x_d)), \qquad F_i(x_j) = P(X_{i,j} \le x_j).$$

Given a parametric copula model expressed through a set of densities $c(u, \theta)$ for $\Theta \subseteq \mathbb{R}^p$ and $u \in [0,1]^d$, the maximum pseudo likelihood estimator $\hat\theta_n$, also called the MPLE, is defined as the minimizer of the pseudo likelihood

$$\ell_n(\theta) := \sum_{i=1}^n \log c(F_{n,\perp}(X_i), \theta).$$

The pseudo likelihood is expressed in terms of the so-called pseudo-observations $F_{n,\perp}(X_i) \in [0,1]^d$, in which $F_{n,\perp}$ is the vector of re-normalized marginal empirical distribution functions

$$F_{n,\perp}(x) := (F_{n,1}(x_1), \ldots, F_{n,d}(x_d)), \text{ where } F_{n,j}(x_j) := \frac{1}{n+1}\sum_{i=1}^n I\{X_{i,j} \le x_j\}.$$

The non-standard normalization constant $1/(n+1)$ – instead of the classical $1/n$ – is to avoid evaluating $u \mapsto \log c(u, \theta)$ at the boundary $u \in \partial[0,1]^d$ where most copula models of interest are infinite.

Many investigations, such as Chen & Fan (2005), use

$$\text{AIC}^\star = 2\ell_{n,\max} - 2\text{length}(\theta) \tag{1}$$

Date: Revised in January 2011.

Key words and phrases. AIC, CIC, copulae, model selection, MPLE, multivariate rank statistics.

as a model selection criterion for the MPLE, with $\ell_{n,\max} = \ell_n(\hat{\theta})$ being the maximum pseudo likelihood. This is inspired from the traditional Akaike information criterion AIC $= 2\ell^{\#}_{n,\max} - 2\mathrm{length}(\theta)$, where $\ell^{\#}_{n,\max}$ is the usual maximum likelihood for a fully parametric model. One computes this AIC* score for each candidate model and in the end chooses the model with highest score.

This cannot be quite correct, however, as the arguments underlying the derivations of the traditional AIC do not apply here – since $\ell_n(\cdot)$ at work here is not a proper log-likelihood function for a model, but a pseudo likelihood, based on the multivariate rank statistics $F_{n,\perp}$. In other words, the AIC* formula above ignores the noise inherent in the transformation step that takes X_i to $F_{n,\perp}(X_i)$. Such a formula would be appropriate only if we could use $F^\circ_k(X_{i,k})$ instead of the pseudo-observations, or if we would model the marginals by a parametric model $F_{k,\gamma(k)}$ that would lead to the classical AIC formula $2\ell_{n,\max^\#} - 2\mathrm{length}(\theta) - 2\sum_{k=1}^d \mathrm{length}(\gamma(k))$ where $\ell^{\#}_{n,\max}$ is the standard maximized likelihood.

This paper reconsider the steps leading to the original AIC-formula in the MPLE setting and derive the appropriate modifications. This leads to the Copula Information Criterion presented in Section 2. However, we will see that the formula yields infinite values when the copula model has extreme behaviour near the edge of the unit cube. Such copulae are overwhelmingly more popular than copulae which are smoother near the edge, making the Copula Information Criterion of limited applicability.

We find that the cause of the typical non-existence of the CIC is that the MPLE can be perceived as a two-stage estimator, where the marginals are estimated non-parametrically. This two-stage procedure introduces a certain bias, which becomes highly significant at the $O_P(n^{-1})$-scale that we will see is the scale defined as low-level noise in the classical AIC-formula.

We will consistently apply the perpendicular subscript to indicate vectors of marginal distributions, such as $F_{n,\perp}$. Note that we will sometimes use the multivariate empirical distribution function F_n, which is defined with the standard scaling $1/n$ in contrast to our marginal empirical distrubtions that are scaled according to $1/(n+1)$. We will also use the circle superscript to denote any size related to F°. Hats will denote estimators, generic elements of $[0,1]^d$ or $[0,1]$ will be denoted by u or v, while elements of $\mathbb{R}^d$ not constrained to $[0,1]^d$ will be denoted by x or y. For a general introduction to copula models, see Joe (1997), and for a general introduction to the model selection problem, see Claeskens & Hjort (2008).

2. The Copula Information Criterion

Like the AIC, the copula information criterion is based on asymptotic likelihood theory. The maximum pseudo likelihood estimator can be written as

$$\hat{\theta}_n = \underset{\theta\in\Theta}{\operatorname{argmax}} \frac{1}{n}\ell_n(\theta) = \underset{\theta\in\Theta}{\operatorname{argmax}} \int_{u\in[0,1]^d} \log c(u,\theta)\,\mathrm{d}C_n(u)$$

where C_n is the empirical copula

$$C_n(u) := \frac{1}{n}\sum_{i=1}^n I\{F_{n,\perp}(X_i) \le u\}.$$

We typically have

$$\hat{\theta}_n \xrightarrow[n\to\infty]{\mathscr{P}} \underset{\theta\in\Theta}{\operatorname{argmax}} \int_{u\in[0,1]^d} \log c(u,\theta)\,\mathrm{d}C^\circ(u) =: \theta^\circ,$$

in which θ° is the least false parameter according to the relative entropy – also known as the Kullback–Leibler divergence – between $c^\circ(u)$ and $\{c(u,\theta) : \theta \in \Theta\}$. That is,

$$\theta^\circ = \operatorname*{argmin}_{\theta\in\Theta} \mathrm{KL}[c^\circ(u), c(u,\theta)] = \operatorname*{argmin}_{\theta\in\Theta} \int_{u\in[0,1]^d} \log \frac{c^\circ(u)}{c(u,\theta)} c^\circ(u)\,\mathrm{d}u. \tag{2}$$

Central to our investigation is the behavior of the pseudo-log-likelihood normalized by sample size

$$A_n(\theta) := \frac{1}{n}\ell_n(\theta) = \int_{[0,1]^d} \log c(u;\theta)\,\mathrm{d}C_n(u)$$

for which we have

$$A_n(\theta) \xrightarrow[n\to\infty]{\mathscr{P}} A(\theta) := \int_{[0,1]^d} \log c(u;\theta)\,\mathrm{d}C^\circ = \int_{[0,1]^d} c^\circ(v) \log c(v;\theta)\,\mathrm{d}v$$

for each θ under regularity conditions.

The basic idea of model selection in the style of the AIC is to choose the model with the least attained Kullback–Leibler divergence to the true model c°. According to eq. (2), we only need to find the model with the largest value of $A(\hat\theta)$. As the function $\theta \mapsto A(\theta)$ is unknown, we will use $A_n(\hat\theta)$ to approximate $A(\hat\theta)$, and then study the difference $A_n(\hat\theta) - A(\hat\theta)$ to make small-sample corrections to the estimator $A_n(\hat\theta)$. We will follow the AIC formula in providing bias-correction terms specifically on the $o_P(n^{-1})$-level.

For simplicity and directness, we will follow the classical score-based likelihood theory of Genest et al. (1995). Let

$$U_n := \frac{\partial A_n(\theta_0)}{\partial\theta} = \frac{1}{n}\frac{\partial \ell_n(\theta_0)}{\partial\theta}$$

be the normalized pseudo-score function, evaluated at θ_0.

Lemma 1. *Given the regularity assumptions on $\{c_\theta : \theta \in \Theta\}$ of Genest et al. (1995), or if $v \mapsto \log c(v,\theta)$ is of bounded Hardy–Krause-variation, then*

$$\sqrt{n}U_n \xrightarrow[n\to\infty]{\mathscr{W}} U \sim N_p(0,\Sigma)$$

where $\Sigma := \mathcal{I} + \mathcal{W}$ in which $\mathcal{I}$ is the Information matrix

$$\mathcal{I} = \mathbb{E}\phi(\xi,\theta_0)\phi(\xi,\theta_0)^t$$

and $\mathcal{W} = \mathrm{Var}\, Z$ accounts for the fact that we are dealing with a pseudo-likelihood. Here

$$Z := \sum_{k=1}^{d} \int_{[0,1]^d} \frac{\partial\phi(v,\theta_0)}{\partial v_k}(I\{\xi_k \le v_k\} - v_k)\,\mathrm{d}C^\circ(v)$$

in which ξ is a random vector distributed according to C° and $\phi(u,\theta) := (\partial/\partial\theta)\log c(u;\theta)$.

Proof. Theorem 6 of Fermanian et al. (2004) proves the statement of bounded variation, but seems to omit that they require Hardy–Krause-variation (and not some other multivariate variational concept). □

We shall also need the symmetric matrix

$$J = -A''(\theta_0) = -\int_{[0,1]^d} c^\circ(v)\frac{\partial^2 \log c(v;\theta_0)}{\partial\theta\partial\theta^t}\,\mathrm{d}v,$$

assumed to be of full rank. A useful random process is now the localized and centred likelihood process

$$H_n(s) = n\{A_n(\theta_0 + s/\sqrt{n}) - A_n(\theta_0)\}.$$

It is defined for those $s \in \mathbb{R}^p$ for which $\theta_0 + s/\sqrt{n}$ is inside the parameter region Θ; in particular, for any $s \in \mathbb{R}^p$, $H_n(s)$ is defined for all large n.

A Taylor expansion demonstrate that for each s

$$H_n(s) = s^t \sqrt{n} U_n - \frac{1}{2} s^t J_n s + o_P(1),$$

where

$$J_n := -\int_{[0,1]^d} \frac{\partial^2 \log c(v;\theta_0)}{\partial\theta\partial\theta^t} \,\mathrm{d}C_n(u) \xrightarrow[n\to\infty]{\mathscr{P}} J.$$

This is close to showing the process convergence

$$H_n(s) \xrightarrow[n\to\infty]{\mathscr{W}} H(s) = s^t U - \frac{1}{2} s^t J s$$

in the Skorokhod spaces $D[-a,a]^p$ for each $a > 0$. The first consequence of note is the limiting distribution of the maximum pseudo-likelihood estimator. Under appropriate conditions (See e.g. van der Vaart & Wellner, 1996), we may use the continuity of the argmax functional to conclude that

$$M_n = \operatorname{argmax}(H_n) \xrightarrow[n\to\infty]{\mathscr{W}} M = \operatorname{argmax}(H),$$

but this is the same as

$$\sqrt{n}(\hat\theta - \theta_0) \xrightarrow[n\to\infty]{\mathscr{W}} J^{-1}U \sim N_p(0, J^{-1}\Sigma J^{-1}). \tag{3}$$

We will avoid making such an argmax-continuity argument mathematically rigorous, it would require some mathematical sophistication and we will only need the basic convergence of eq. (3) in the following. Such convergence is proved in Genest et al. (1995) under classical conditions on the parametrization of the model. We use this notation to show that our developments are completely parallel to the derivation of the classical AIC formula given in e.g. Claeskens & Hjort (2008).

Secondly, we investigate the actual Kullback–Leibler distance from the true model to that used for fitting the parametric family given by

$$\mathrm{KL}(c^\circ(u), c(u,\hat\theta)) = \int_{[0,1]^d} c^\circ(u) \log c^\circ(u)\,\mathrm{d}v - \int_{[0,1]^d} c^\circ(u) \log c(u,\hat\theta)\,\mathrm{d}u.$$

It is rather difficult (but possible) to estimate the first term from data, but we may ignore it, since it is common to all parametric families. For the purposes of model selection it therefore suffices to estimate the second term, which is $A(\hat\theta)$.

We now examine

$$\text{estimator } A_n(\hat\theta) = \frac{1}{n}\ell_{n,\max} \text{ vis-á-vis target } A(\hat\theta).$$

In the fully parametric ML case, the estimator $A_n(\hat\theta)$ (defined mutatis mutandis) always *overshoots* its target $A(\hat\theta)$ (again defined mutatis mutandis), and the AIC is simply a renormalization of $A_n(\hat\theta)$, minus a penalization for model complexity. This penalty term serves is roughly a first order bias-correction term. In the present, semiparametric case, we will shortly see that $A_n(\hat\theta)$ can both overshoot and undershoot its target. Let

$$Z_n = n\{A_n(\hat\theta) - A_n(\theta_0)\} - n\{A(\hat\theta) - A(\theta_0)\}.$$

Some re-arrangement shows that

$$A_n(\hat\theta) - A(\hat\theta) = \frac{1}{n} Z_n + A_n(\theta_0) - A(\theta_0). \tag{4}$$

Also,

$$Z_n = H_n(M_n) + \frac{1}{2} n(\hat\theta - \theta_0)^t J(\hat\theta - \theta_0) + o_P(1),$$

in which we define the stochastically significant part as p_n, giving rise to

$$p_n := H_n(M_n) + \frac{1}{2} n(\hat\theta - \theta_0)^t J(\hat\theta - \theta_0) \xrightarrow[n\to\infty]{\mathscr{W}} H(M) + \frac{1}{2} U^t J^{-1} U = U^t J^{-1} U =: P.$$

We have

$$p^* = \mathbb{E}P = \mathbb{E}U^t J^{-1} U = \mathrm{Tr}(J^{-1}\Sigma) = \mathrm{Tr}\left(J^{-1}\mathcal{I}\right) + \mathrm{Tr}\left[J^{-1}\mathcal{W}\right].$$

Note that similarly to the fully parametric case, we have $p^* \geq 0$ since all matrices involved are positive definite, and the trace of positive definite matrices are positive.

The standard argument leading to the AIC formula ends at this point. When working with a fully parametric model estimated through Maximum Likelihood, the only work left is providing estimators for p^*. However, as we are to provide bias-correction terms at the $o_P(n^{-1})$-level, careful examination of $A_n(\theta_0) - A(\theta_0)$ is required.

2.1. **The study of $A_n(\theta_0) - A(\theta_0)$.** Although $\sqrt{n}[A_n(\theta_0) - A(\theta_0)]$ is typically asymptotically mean zero normal, it does not have zero mean for finite n. This is in sharp contrast to the AIC-case, where the analogous term in its derivation leads to a difference of the form $\int_{[0,1]^d} \log c(x, \theta_0)\,\mathrm{d}[F_n - F^\circ](x)$. As $\mathbb{E}\int_{[0,1]^d} \log c(x,\theta_0)\,\mathrm{d}F_n(x) = \int_{[0,1]^d} \log c(x,\theta_0)\,\mathrm{d}F^\circ(x)$, this difference has precisely zero mean – and not merely *asymptotically* zero mean. If we are to derive a model selection formula in the vein of the AIC formula, further study of the difference

$$A_n(\theta_0) - A(\theta_0) = \frac{1}{n}\sum_{i=1}^n \log c(F_{n,\perp}(X_i);\theta_0) - \int c^\circ(u) \log c(u;\theta_0)\,\mathrm{d}u$$

is required. If $v \mapsto \log c(v;\theta_0)$ is two times continuously differentiable, a two-term Taylor-expansion of each term in $A_n(\theta_0)$ around $F_{n,\perp}(X_i) - F_\perp^\circ(X_i)$ gives the fundamental relation

$$A_n(\theta_0) - A(\theta_0) = \int \log c(F_\perp^\circ(x), \theta_0)\,\mathrm{d}[F_n - F^\circ] + Q_n + R_n + B_n \tag{5}$$

where

$$\begin{aligned} Q_n &= \frac{1}{n}\sum_{i=1}^n \zeta'(F_\perp^\circ(X_i),\theta_0)^t (F_{n,\perp}(X_i) - F_\perp^\circ(X_i)), \\ R_n &= \frac{1}{2n}\sum_{i=1}^n (F_{n,\perp}(X_i) - F_\perp^\circ(X_i))^t \zeta''(F_\perp^\circ(X_i),\theta_0)(F_{n,\perp}(X_i) - F_\perp^\circ(X_i)) \end{aligned}$$

in which

$$\zeta'(v,\theta) = \frac{\partial \log c(v,\theta)}{\partial v} \text{ and } \zeta''(v,\theta) = \frac{\partial^2 \log c(v,\theta)}{\partial v \partial v^t}$$

and

$$B_n = \frac{1}{2n}\sum_{i=1}^n (F_{n,\perp}(X_i) - F_\perp^\circ(X_i))^t \left[\zeta''(H_n(X_i),\theta_0) - \zeta''(F_\perp^\circ(X_i),\theta_0)\right] (F_{n,\perp}(X_i) - F_\perp^\circ(X_i))$$

where H_n is a vector function with entries $H_{n,i}(x) = F_i^\circ(x_i) + \tau_{n,i}(x)[F_{n,i}(x_i) - F_i(x_i)]$ for some stochastic vector $\tau_n(x) = (\tau_{n,1}, \dots, \tau_{n,d}) \in (0,1)^d$.

Theorem 1 will give conditions for when B_n is $o_p(n^{-1})$, and thus considered low-level noise. Clearly, the first term of eq. (5) has zero mean, and it remains to find the expectation of the stochastically significant parts of Q_n and R_n. This is described by following two lemmas, proved in the Appendix.

Lemma 2. *We have the decomposition $Q_n = \frac{1}{n}q_n + Z_{Q,n}$ where $\mathbb{E}Z_{Q,n} = 0$ and*

$$\begin{aligned} q_n &= \frac{n}{n+1}\int \zeta'(F_\perp^\circ(x),\theta_0)^t\,(\mathbf{1} - F_\perp^\circ(x))\,\mathrm{d}F_n = O_p(1), \\ \mathbb{E}q_n &= \frac{n}{n+1}\int_{[0,1]^d} \zeta'(v,\theta_0)^t\,(\mathbf{1} - v)\,\mathrm{d}C^\circ(v) \end{aligned}$$

Lemma 3. *Let $C_{a,b}$ be the cumulative copula of $(X_{1,a}, X_{1,b})$. We have $n\mathbb{E}R_n \to \mathbf{1}^t\Upsilon\mathbf{1}$ where $\Upsilon = (\Upsilon_{a,b})_{1\le a,b\le d}$ is the symmetric matrix with*

$$\Upsilon_{a,a} = \frac{1}{2}\int_{[0,1]^d} \zeta''_{a,a}(u;\theta_0)u_a(1-u_a)\,\mathrm{d}C^\circ,$$

$$\Upsilon_{a,b} = \frac{1}{2}\int_{[0,1]^d} \zeta''_{a,b}(u;\theta_0)\,[C_{a,b}(u_a,u_b) - u_a u_b]\;\mathrm{d}C^\circ \quad (\textit{when } a \neq b),$$

and $\mathbb{E}R_n$ is finite only if Υ is.

This leads to the following result.

Theorem 1. *If $v \mapsto \log c(v,\theta)$ is two times continuously differentiable on $(0,1)^d$ and if ζ'' and C° follow the conditions of Proposition 2 in the Appendix, then*

$$A_n(\theta_0) - A(\theta_0) = \frac{1}{n}(q_n + r_n) + \tilde{Z}_n + o_P(n^{-1}), \tag{6}$$

in which $\mathbb{E}\tilde{Z}_n = 0$ and

$$q^* := \lim_{n\to\infty} \mathbb{E}q_n = \int_{[0,1]^d} \zeta'(v,\theta_0)^t\,(\mathbf{1}-v)\;\mathrm{d}C^\circ(v)$$

$$r^* := \lim_{n\to\infty} \mathbb{E}r_n = \mathbf{1}^t\Upsilon\mathbf{1}$$

where $\mathbb{E}r_n$ and $\mathbb{E}q_n$ is infinite only if r^* and q^* respectively is infinite.

Proof. This is a direct consequence of Lemma 2, Lemma 3 and Proposition 2 in the Appendix. □

While $\mathbb{E}q_n$ is usually finite, Υ is not. To illustrate this problem, let $d = 2$ and assume that the model is correctly specified, so that $c^\circ(v) = c(v;\theta_0)$. We then have

$$\zeta''_{i,j}(u,\theta_0) = \partial_j\frac{\partial_i c^\circ(u)}{c^\circ(u)} = \frac{\partial_{i,j}c^\circ(u)}{c^\circ(u)} - \frac{\partial_i c^\circ(u)\partial_j c^\circ(u)}{c^\circ(u)^2},$$

yielding

$$\Upsilon_{1,2} = \int_{[0,1]^2}\left[c^\circ(u) - \frac{\partial_1 c^\circ(u)\partial_2 c^\circ(u)}{c^\circ(u)}\right][C^\circ(u,v) - uv]\;\mathrm{d}C^\circ(u,v),$$

$$\Upsilon_{1,1} = \int_{[0,1]^2}\left[c^\circ(u) - \frac{\partial_1 c^\circ(u)\partial_1 c^\circ(u)}{c^\circ(u)}\right]u(1-u)\,\mathrm{d}C^\circ(u,v),$$

$$\Upsilon_{2,2} = \int_{[0,1]^2}\left[c^\circ(u) - \frac{\partial_2 c^\circ(u)\partial_2 c^\circ(u)}{c^\circ(u)}\right]v(1-v)\,\mathrm{d}C^\circ(u,v).$$

Example 1. Consider the bivariate Kimeldorf & Sampson family of copulae with density

$$c(u,v;\delta) = \frac{1+\delta}{(uv)^{\delta+1}}\left(1/u^\delta + 1/v^\delta - 1\right)^{2+1/\delta},$$

which is copula B4 in Joe (1997). The B4 density is simply a rational polynomial when $\delta = 1$. This enables us to give closed form expressions for $\Upsilon_{a,b}$ with the help of a computer algebra system. This shows that

$$\Upsilon_{1,2} = \int_0^1\left[\frac{1}{5}v^{-1} - \frac{3}{10}v + \frac{1}{10}\right]\mathrm{d}v,$$

$$\Upsilon_{1,1} = \int_0^1\left[v^{-1} + \frac{1}{2}v^{-2} + \frac{3}{2}\right]v(1-v)\,\mathrm{d}v,$$

$$\Upsilon_{2,2} = \int_0^1 \frac{1}{2}v^{-1}\,\mathrm{d}v.$$

As $\int_0^1 v^{-1}\,\mathrm{d}v = \infty$, we get that Υ, and hence also $\mathbb{E}R_n$, is infinite.

In fact, the B4 copula is not a pathology. Although it is typical that $R_n = O_P(1)$, it is also typical that $\mathbb{E}R_n$ is infinite. Almost all of the copula models categorized in Joe (1997) has infinite Υ-values, i.e. the distribution of R_n has very heavy tails.

Although this infinitude is somewhat surprising, it is not a paradox and is another example of why expectation is not the same as a typical value of a random variable. The most basic example of this phenomenon is an *iid* sequence $\xi_1, \xi_2, \dots$ of Cauchy variables. The expectation $\mathbb{E}\bar{X}_n$ is infinite, while $\bar{X}_n$ is again Cauchy distributed for each n – and hence trivially $O_P(1)$. However, the implication and interpretation of the infinite expectation of the bias-correction term is that it is fundamentally impossible to complete the AIC programme as defined above for the MPLE, even under enough regularity to secure the validity of the above Taylor-expansions. This is a second order effect of estimating the marginals non-parametrically.

Grønneberg (2010) argues that the MPLE can be seen as a natural estimator of the copula parameter under symmetry conditions. Its well-known lack of semiparametric efficiency is not a crucial deficiency in the context of model selection as semiparametric efficiency in the sense of Bickel et al. (1993) assumes that the model is correctly specified. In this case, symmetry considerations provide natural motivation for using the MPLE compared to other semiparametrically efficient estimators. However, the lack of an AIC-like model selection formula is a more serious limitation of the MPLE.

2.2. **Empirical estimates.** The CIC formulae now follows when empirical estimates of the asymptotic expectation of q_n and r_n are found. Just as for the fully parametric case, significant simplifications can be made when the model is assumed correct. This leads to a CIC-formula that we call the AIC-like CIC formula, derived in Section 2.2.1. If the model is not assumed correct, nonparametric estimates are required and we get the so-called TIC-like CIC formula, given in Section 2.2.2.

2.2.1. *AIC-like formula.* This section works under the assumption of a correct super-model, as was the case for the original AIC formula. This assumption leads to several simplifications, as shown by the following result whose proof is deferred to the Appendix.

Proposition 1. *If the parametric model is correctly specified, we have* $q^* = 0$ *and* $p^* =$ *length*$(\theta) + \mathrm{Tr}[\mathcal{I}^{-1}\mathcal{W}]$.

This motivates the AIC-like Copula Information Criterion

$$\boxed{\mathrm{CIC} = 2\ell_{n,\max} - 2(\hat{p}^* + \hat{r}^*),} \tag{7}$$

where $\hat{p}^*$ and $\hat{r}^*$ estimates p^* and r^* respectively.

A natural estimator of r^* is $\hat{r}^* = \mathbf{1}^t \hat{\Upsilon} \mathbf{1}$, defined in terms of the plug-in estimators

$$\hat{\Upsilon}_{a,a} = \frac{1}{2} \int_{[0,1]^d} c(v;\hat{\theta}) \zeta''_{a,a}(v;\hat{\theta}) v_a(1-v_a)\,\mathrm{d}v,$$

$$\hat{\Upsilon}_{a,b} = \frac{1}{2} \int_{[0,1]^d} c(v;\hat{\theta}) \zeta''_{a,b}(v;\hat{\theta}) \left[C_{a,b}(v_a, v_b;\hat{\theta}) - v_a v_b \right] \mathrm{d}v$$

where $C_{a,b}(v_a, v_b;\theta)$ is the cumulative copula of (Y_a, Y_b) where $(Y_1, Y_2, \dots, Y_d) \sim C(v;\theta)$. A natural estimation procedure for p^* is to use

$$\hat{p}^* = \mathrm{length}(\theta) + \mathrm{Tr}\left(\hat{\mathcal{I}}^{-}\hat{W}\right)$$

denoting the generalized inverse of $\hat{\mathcal{I}}$ by $\hat{\mathcal{I}}^-$ and where $\hat{\mathcal{I}}$ is the pseudo-empirical information matrix

$$\hat{\mathcal{I}} = \mathbb{E}_{\hat{\theta}} \phi(\tilde{\xi}, \hat{\theta}) \phi(\tilde{\xi}, \hat{\theta})^t \tag{8}$$

and

$$\hat{W} = \mathrm{Var}_{\hat{\theta}} \left\{ \int_{[0,1]^d} \left(\frac{\partial^2}{\partial\theta\partial v^t} \log c(v, \hat{\theta}) \right)^t (I\{\xi \le v\}_\perp - v) \, \mathrm{d}C(v, \hat{\theta}) \right\} \tag{9}$$

where $\tilde{\xi} \sim C(v; \hat{\theta})$. These integrals can easily be evaluated through numerical integration routines such as Monte-Carlo simulation. Note, however, that in contrast to the classical AIC formula, which has exactly zero small-sample bias, the AIC-like CIC formula requires that both $\mathrm{Tr}\left(\mathcal{I}^{-1}W\right)$ and r^* is estimated from data.

We note that these estimators are somewhat different from the ones suggested by Genest et al. (1995), which are based on using the empirical copula as plug-in estimates of the expectation operator $\mathbb{E}_{\hat{\theta}}$. This would give

$$\hat{\mathcal{I}}^\star = \int_{u\in[0,1]^d} \phi(u, \hat{\theta})\phi(u, \hat{\theta})^t \, \mathrm{d}\hat{C}(u) = \frac{1}{n} \sum_{k=1}^n \phi(\hat{\xi}^{(k)}, \hat{\theta})\phi(\hat{\xi}^{(k)}, \hat{\theta})^t$$

and $\hat{W}^\star$ as the empirical variance of

$$\int_{[0,1]^d} \left(\frac{\partial^2}{\partial\theta\partial v^t} \log c(v, \hat{\theta}) \right)^t (I\{\hat{\xi}^{(k)} \le v\}_\perp - v) \, \mathrm{d}C_n(v)$$

for $\hat{\xi}^{(k)} = F_{n,\perp}(X_k)$ together with analogues for $\hat{r}^*$. These estimates are valid also when the copula model is incorrectly specified, and has the further advantage of being very simple to calculate by avoiding the need for numerical integration.

2.2.2. *TIC-like formula.* We now have to rely on nonparametric estimates. A natural estimator for q^* is the plug-in estimators

$$\hat{q}^* = \int_{[0,1]^d} \zeta'(v; \hat{\theta})^t \big(\mathbf{1} - v\big) \, \mathrm{d}\hat{C}(v)$$

while for r^* is $\hat{r}^* = \mathbf{1}^t \hat{\Upsilon} \mathbf{1}$ where now

$$\hat{\Upsilon}_{a,a} = \frac{1}{2} \int_{[0,1]^d} \zeta''_{a,a}(v; \hat{\theta}) v_a (1 - v_a) \, \mathrm{d}\hat{C}_n,$$
$$\hat{\Upsilon}_{a,b} = \frac{1}{2} \int_{[0,1]^d} \zeta''_{a,b}(v; \hat{\theta}) \left[\hat{C}_{n,a,b}(v_a, v_b) - v_a v_b \right] \mathrm{d}\hat{C}_n$$

where $C_{n,a,b}$ is the empirical copula based on $(X_{1,a}, X_{1,b}), (X_{2,a}, X_{2,b}), \ldots, (X_{n,a}, X_{n,b})$. As for the estimation of p^*, we use $\hat{p}^* = \mathrm{Tr}\, \hat{J}_n^{-1} \hat{\Sigma}$ where

$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^n \left\{ \phi(\hat{\xi}^{(i)}; \hat{\theta}) + \hat{Z}_i \right\} \left\{ \phi(\hat{\xi}^{(i)}; \hat{\theta}) + \hat{Z}_i \right\}^t$$

with

$$\hat{Z}_i = \sum_{j=1}^d \frac{1}{n} \sum_{s=1, s\neq i}^n \left. \frac{\partial \phi(v; \hat{\theta})}{\partial v_j} \right|_{v=\hat{\xi}^{(s)}} \left(I\left\{ \hat{\xi}_j^{(i)} \le \hat{\xi}_j^{(s)} \right\} - \hat{\xi}_j^{(s)} \right)$$

using $\hat{\xi}^{(k)} = F_{n,\perp}(X_k)$.

Appendix A. Technical proofs

This appendix gathers technical proofs needed for the above results. In addition to the already introduced notation, we will work with the empirical processes

$$\mathbb{G}_{n,k}(x_k) = \sqrt{n}\left[F_{n,k}(x_k) - F_k^\circ(x_k)\right], \qquad \mathbb{G}_{n,\perp}(x) = \sqrt{n}\left[F_{n,\perp}(x) - F_\perp^\circ(x)\right],$$
$$\mathbb{G}_n(x) = \sqrt{n}\left[F_n(x) - F^\circ(x)\right], \qquad \mathbb{C}_n(u) = \sqrt{n}\left[C_n(u) - C^\circ(u)\right].$$

A.1. **Proofs for Expectation Structure.**

Proof of Lemma 2. Define

$$\mathbb{G}_{n,\perp,-i} = \frac{\sqrt{n}}{n+1} \sum_{1\le k\le n, k\neq i} [I\{X_k \le x\} - F_\perp^\circ(x)]$$

so that $\mathbb{G}_{n,\perp}(x) = \mathbb{G}_{n,\perp,-i}(x) - \sqrt{n}/(n+1)\,[I\{X_i \le x\}_\perp - F_\perp(x)]$. This shows

$$Q_n = \frac{1}{\sqrt{n}} \int \zeta'(F_\perp^\circ(x);\theta_0)^t \mathbb{G}_{n,\perp}(x)\,\mathrm{d}F_n(x) = \frac{1}{n\sqrt{n}} \sum_{i=1}^n \zeta'(F_\perp^\circ(X_i);\theta_0)^t \mathbb{G}_{n,\perp,-i}(X_i)$$

$$\frac{1}{n^2}\frac{n}{n+1}\sum_{i=1}^n \zeta'(F_\perp^\circ(X_i);\theta_0)^t\,[I\{X_i \le X_i\}_\perp - F_\perp(X_i)]\,.$$

The second term is q_n/n. By independence, we have

$$\mathbb{E}\zeta'(F_\perp^\circ(X_i);\theta_0)^t\mathbb{G}_{n,\perp,-i,+1}(X_i) = \mathbb{E}\mathbb{E}\left[\zeta'(F_\perp^\circ(X_i);\theta_0)^t\mathbb{G}_{n,\perp,-i,+1}(X_i)\middle|X_i\right] = 0.$$

□

Proof of Lemma 3. Notice that

$$\begin{aligned}
R_n =& \frac{1}{2n^2}\sum_{i=1}^n \mathbb{G}_{n,\perp}(X_i)^t\zeta''(F_\perp^\circ(X_i),\theta_0)\mathbb{G}_{n,\perp}(X_i)\\
=& \frac{1}{2n^2}\sum_{i=1}^n \mathbb{G}_{n,\perp,-i}(X_i)^t\zeta''(F_\perp^\circ(X_i),\theta_0)\mathbb{G}_{n,\perp,-i}(X_i)\\
&+ \frac{1}{2n^2}\frac{\sqrt{n}}{n+1}\sum_{i=1}^n \mathbb{G}_{n,\perp,-i}(X_i)^t\zeta''(F_\perp^\circ(X_i),\theta_0)\,[I\{X_i \le X_i\}_\perp - F_\perp(X_i)]\\
&+ \frac{1}{2n^2}\frac{\sqrt{n}}{n+1}\sum_{i=1}^n [I\{X_i \le X_i\}_\perp - F_\perp(X_i)]^t\,\zeta''(F_\perp^\circ(X_i),\theta_0)\mathbb{G}_{n,\perp,-i}(X_i)\\
&+ \frac{1}{2n^2}\left(\frac{\sqrt{n}}{n+1}\right)^2\sum_{i=1}^n [I\{X_i \le X_i\}_\perp - F_\perp(X_i)]^t\,\zeta''(F_\perp^\circ(X_i),\theta_0)\,[I\{X_i \le X_i\}_\perp - F_\perp(X_i)]\,.
\end{aligned}$$

After multiplying with n, only the first term will have an effect on the expectation as $n \to \infty$. By independence, its expectation is given by

$$\begin{aligned}
&\frac{1}{2n}\mathbb{E}\int_{\mathbb{R}^d} \mathbb{G}_{n-1,\perp}(x)^t\zeta''(F_\perp^\circ(x);\theta_0)\mathbb{G}_{n-1,\perp}(x)\,\mathrm{d}F^\circ(x)\\
=&\frac{1}{n}\int_{\mathbb{R}^d}\mathbb{E}\left[\mathbb{G}_{n-1,\perp}(x)^t\zeta''(F_\perp^\circ(x);\theta_0)\mathbb{G}_{n-1,\perp}(x)\right]\,\mathrm{d}F^\circ(x)\\
=&\frac{1}{n}\sum_{1\le a,b\le d}\int_{\mathbb{R}^d}\zeta''_{a,b}(F_\perp^\circ(x);\theta_0)\mathbb{E}\left[\mathbb{G}^{(k)}_{n-1,a}(x_a)\mathbb{G}_{n-1,b}(x_b)\right]\,\mathrm{d}F^\circ(x).
\end{aligned}$$

Let $\rho_n = n^2/(n+1)^2$. We have

$$\begin{aligned}
\mathbb{E}\mathbb{G}_{n,a}(x_a)\mathbb{G}_{n,b}(x_b) &= \rho_n\frac{1}{n}\mathbb{E}\left[\sum_{i=1}^n I\{X_{i,k} \le x_k\} - F_k^\circ(x_k)\right]\left[\sum_{j=1}^n I\{X_{j,l} \le x_l\} - F_l^\circ(x_l)\right]\\
&= \rho_n\frac{1}{n}\sum_{i=1}^n \mathbb{E}\,[I\{X_{i,l} \le x_l\} - F_l^\circ(x_l)]\,[I\{X_{i,k} \le x_k\} - F_k^\circ(x_k)]\\
&\quad + \rho_n\frac{1}{n}\mathbb{E}\sum_{1\le i,j\le n, i\neq j}[I\{X_{i,k} \le x_k\} - F_k^\circ(x_k)]\,[I\{X_{j,l} \le x_l\} - F_l^\circ(x_l)]\,.
\end{aligned}$$

The second term vanishes by independence, yielding

$$\begin{aligned}\mathbb{E}\mathbb{G}_{n,a}(x_a)\mathbb{G}_{n,b}(x_b) &= \rho_n \frac{1}{n}\sum_{i=1}^n \Big\{ \mathbb{E}\left[I\{X_{i,l} \le x_l\} - F_l^\circ(x_l)\right] I\{X_{i,k} \le x_k\} \\ &\qquad + \mathbb{E}\left[I\{X_{i,l} \le x_l\} - F_l^\circ(x_l)\right] F_k^\circ(x_k) \Big\} \\ &= \rho_n \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[I\{X_{i,l} \le x_l\} I\{X_{i,k} \le x_k\} - F_k^\circ(x_k) F_l^\circ(x_l)\right],\end{aligned}$$

which is equal to $x_a(1-x_a)$ if $a=b$ and $P\{X_{1,l} \le x_l, X_{1,k} \le x_k\} - F_k^\circ(x_k)F_l^\circ(x_l)$ otherwise. Thus,

$$\begin{aligned}&\frac{1}{2n}\mathbb{E}\int_{\mathbb{R}^d} \mathbb{G}_{n-1,\perp}(x)^t \zeta''(F_\perp^\circ(x);\theta_0)\mathbb{G}_{n-1,\perp}(x)\,\mathrm{d}F^\circ(x) \\ =&\rho_n \frac{1}{2n}\sum_{1\le a,b\le d, a\ne b}\int_{\mathbb{R}^d} \zeta''_{a,b}(F_\perp^\circ(x);\theta_0)\left[P\{X_{1,a}\le x_a, X_{1,} \le x_b\} - F_a^\circ(x_a)F_b^\circ(x_b)\right]\,\mathrm{d}F^\circ(x) \\ &+ \rho_n \frac{1}{2n}\sum_{1\le a\le d}\int_{\mathbb{R}^d} \zeta''_{a,a}(F_\perp^\circ(x);\theta_0)x_a(1-x_a)\,\mathrm{d}F^\circ(x).\end{aligned}$$

A change of variables shows that this is equal to

$$\begin{aligned}&\rho_n \frac{1}{2n}\sum_{1\le a,b\le d, a\ne b}\int_{[0,1]^d} \zeta''_{a,b}(u;\theta_0)\left[C_{a,b}(u_a,u_b) - u_a u_b\right]\,\mathrm{d}C^\circ(u) \\ &\qquad + \rho_n \frac{1}{2n}\sum_{1\le a\le d}\int_{[0,1]^d} \zeta''_{a,a}(u;\theta_0)u_a(1-u_a)\,\mathrm{d}C^\circ(u),\end{aligned}$$

which approaches Υ once multiplied by n. □

Proof of Proposition 1. The assumption $c^\circ(u) = c(u,\theta_0)$ validates the information matrix equality $J = \mathcal{I}$, which gives the reduced formula for p^*. As for q^*, let us first notice that the fundamental theorem of calculus shows that

$$c(v;\theta_0)\Big|_{v_k = x} = \frac{\mathrm{d}}{\mathrm{d}x}\int_0^x c(v;\theta_0)\,\mathrm{d}v_k = \frac{\mathrm{d}}{\mathrm{d}x}\int_0^1 c(v;\theta_0) I\{0 \le v_k \le x\}\,\mathrm{d}v_k.$$

As $c(v;\theta_0)I\{0 \le v_k \le x\}$ is dominated by $c(v;\theta_0)$ which is integrable, dominated convergence allows us to move the differential sign in and out of integrals. As $c(v;\theta_0)$ has uniform marginals, this shows

$$\int_0^1\int_0^1\cdots\int_0^1 c(v;\theta_0)\Big|_{v_k=x}\prod_{i\ne k}\mathrm{d}v_i = \frac{\mathrm{d}}{\mathrm{d}x}\int_0^1\int_0^1\cdots\int_0^1\int_0^x c(v;\theta_0)\,\mathrm{d}v_k\prod_{i\ne k}\mathrm{d}v_i = \frac{\mathrm{d}}{\mathrm{d}x}x = 1. \tag{10}$$

We have

$$\begin{aligned}q^* &= \int_{[0,1]^d} \zeta'(v;\theta_0)^t\big(\mathbf{1} - v\big)\,\mathrm{d}C(v;\theta_0) \\ &= \sum_{k=1}^d \int_0^1\int_0^1\cdots\int_0^1 c(v;\theta_0)\frac{\partial \log c(v;\theta_0)}{\partial v_k}(1-v_k)\,\mathrm{d}v_k\prod_{i\ne k}\mathrm{d}v_i \\ &= \sum_{k=1}^d \int_0^1\int_0^1\cdots\int_0^1 \frac{\partial c(v;\theta_0)}{\partial v_k}(1-v_k)\,\mathrm{d}v_k\prod_{i\ne k}\mathrm{d}v_i.\end{aligned}$$

Let $\varepsilon > 0$, and write

$$\int_0^1 \frac{\partial c(v;\theta_0)}{\partial v_k}(1-v_k)\,\mathrm{d}v_k = \int_\varepsilon^{1-\varepsilon}\frac{\partial c(v;\theta_0)}{\partial v_k}(1-v_k)\,\mathrm{d}v_k + \int_{[0,1]\setminus(\varepsilon,1-\varepsilon)}\frac{\partial c(v;\theta_0)}{\partial v_k}(1-v_k)\,\mathrm{d}v_k$$

The first term can be written as

$$
\begin{aligned}
c(v;\theta_0)(1-v_k)\Big|_{v_k=\varepsilon}^{1-\varepsilon} + \int_\varepsilon^{1-\varepsilon} c(v;\theta_0)\,\mathrm{d}v_k &= c(v;\theta_0)\Big|_{v_k=1-\varepsilon}\varepsilon - c(v;\theta_0)\Big|_{v_k=\varepsilon}(1-\varepsilon) \\
&\quad + \int_\varepsilon^{1-\varepsilon} c(v;\theta_0)\,\mathrm{d}v_k \\
&= c(v;\theta_0)\Big|_{v_k=1-\varepsilon}\varepsilon + c(v;\theta_0)\Big|_{v_k=\varepsilon}\varepsilon - c(v;\theta_0)\Big|_{v_k=\varepsilon} \\
&\quad + \int_\varepsilon^{1-\varepsilon} c(v;\theta_0)\,\mathrm{d}v_k
\end{aligned}
$$

through partial integration. By eq. (10), we get

$$
\begin{aligned}
q^* = \sum_{k=1}^d \int_0^1\int_0^1\cdots\int_0^1 \int_{[0,1]\setminus(\varepsilon,1-\varepsilon)} \frac{\partial c(v;\theta_0)}{\partial v_k}(1-v_k)\,\mathrm{d}v_k \prod_{i\neq k}\mathrm{d}v_i \\
+ 2\varepsilon d - d + \sum_{k=1}^d \int_0^1\int_0^1\cdots\int_0^1\int_\varepsilon^{1-\varepsilon} c(v;\theta_0)\,\mathrm{d}v_k\prod_{i\neq k}\mathrm{d}v_i
\end{aligned}
$$

which can be made arbitrarily close to zero by choosing ε sufficiently small. Thus $q^*=0$. □

A.2. **Sufficient conditions for $B_n = o_P(n^{-1})$.** We follow Genest et al. (1995) and Tsukahara (2005) by applying the techniques of Ruymgaart et al. (1972) and Ruymgaart (1974).

Definition 1. (1) *Let $\mathcal{Q}$ be the set of continuous functions q on $[0,1]$, which are positive on $(0,1)$, symmetric about $1/2$, decreasing on $[0,1/2]$ and satisfy $\int_0^1\{q(t)\}^2\,\mathrm{d}t<\infty$.*

(2) *A function $r:(0,1)\mapsto(0,\infty)$ is called u-shaped if it is symmetric about $1/2$ and decreasing on $(0,1/2]$.*

(3) *For $0<\beta<1$ and a u-shaped function r, we define*

$$
r_\beta(t) = \begin{cases} r(\beta t), & \text{if } 0<t\le 1/2; \\ r\left(1-\beta[1-t]\right), & \text{if } 1/2<t\le 1 \end{cases}
$$

If for every $\beta>0$ in a neighbourhood of 0, there exists a constant M_β, such that $r_\beta\le M_\beta r$ on $(0,1)$, then r is called a reproducing u-shaped function. We denote by $\mathcal{R}$ the set of reproducing u-shaped functions.

The importance of $\mathcal{Q}$ and $\mathcal{R}$ comes from the following two Lemmas, proved in Pyke & Shorack (1968) and Ruymgaart (1974) respectively.

Lemma 4. *Suppose $q_k\in\mathcal{Q}$, then $\|\mathbb{G}_{n,k}/q_k\| = O_P(1)$ where $\mathbb{G}_{n,k}$ is the k'th univariate empirical process.*

Lemma 5. *Suppose $H_{n,k}$ satisfies*

$$
\min\left(F_k^\circ(x_k), \frac{1}{n+1}\sum_{i=1}^n I\{X_{i,k}\le x_k\}\right) \le H_{n,k}(x_k) \le \max\left(F_k^\circ(x_k), \frac{1}{n+1}\sum_{i=1}^n I\{X_{i,k}\le x_k\}\right)
$$

for all x_k and let $\Lambda_{n,k} = [\min_{1\le i\le n} X_{i,k}, \max_{1\le i\le n} X_{i,k}]\subset\mathbb{R}$. Let $r\in\mathcal{R}$. Then

$$
\sup_{x_k\in\Lambda_{n,k}} \frac{r(H_{n,k}(x_k))}{r(F_k^\circ(x_k))} = O_P(1)
$$

uniformly in n.

For simplicity, let us assume that $X_1, X_2, \ldots \sim C^\circ$ so that $F_\perp^\circ(x) = x$. By Lemma 1 of Fermanian et al. (2004) this does not entail any loss of generality.

Proposition 2. *Assume that $u \mapsto \zeta''(u, \theta_0)$ is continuous on $(0,1)^d$ and that for each $1 \le k \le d$ and $1 \le a, b \le d$ there exists functions $r_k, \tilde{r}_{k,l,1}, \tilde{r}_{k,l,2} \in \mathcal{R}$, and $q_k \in \mathcal{Q}$ such that*

$$|\zeta''_{a,b}(u, \theta_0)| \le \tilde{r}_{a,b,1}(u_a)\tilde{r}_{a,b,2}(u_b) \prod_{1 \le k \le d, k \ne a,b} r_k(u_k) \tag{11}$$

and

$$\int_{[0,1]^d} q_a(u_a) q_b(u_b) \tilde{r}_{a,b,1}(u_a)\tilde{r}_{a,b,2}(u_b) \prod_{1 \le k \le d, k \ne a,b} r_k(u_k) \,\mathrm{d}C^\circ(u) < \infty. \tag{12}$$

Then $B_n = o_P(n^{-1})$.

Proof. Note that

$$B_n = \frac{1}{2n^2} \sum_{i=1}^n \mathbb{G}_{n,\perp}(X_i)^t \left[\zeta''(H_n(X_i), \theta_0) - \zeta''(F_\perp^\circ(x), \theta_0)\right] \mathbb{G}_{n,\perp}(X_i).$$

For each $0 < \gamma < 1$, let $S_\gamma = [\gamma, 1-\gamma]^d$ and $S_\gamma^c = [0,1]^d \setminus S_\gamma$. Write

$$\begin{aligned} 2nB_n = &\int_{S_\gamma} \mathbb{G}_{n,\perp}(x)^t \left[\zeta''(H_n(X_i), \theta_0) - \zeta''(F_\perp^\circ(x), \theta_0)\right] \mathbb{G}_{n,\perp}(x) \,\mathrm{d}F_n(x) \\ &+ \int_{S_\gamma^c} \mathbb{G}_{n,\perp}(x)^t \left[\zeta''(H_n(X_i), \theta_0) - \zeta''(F_\perp^\circ(x), \theta_0)\right] \mathbb{G}_{n,\perp}(x) \,\mathrm{d}F_n(x), \end{aligned}$$

and denote these integrals by $D_{n,1,\gamma}$ and $D_{n,2,\gamma}$. The absolute value of $D_{n,1,\gamma}$ is bounded by

$$d \sup_{1 \le k,l \le d} \left[\|\mathbb{G}_{n,k}\|_{[\gamma,1-\gamma]}\right] \times \|\mathbb{G}_{n,l}\|_{[\gamma,1-\gamma]} \times \|\zeta''(H_n(X_i), \theta_0) - \zeta''(F_\perp^\circ(x), \theta_0)\|_{S_\gamma}$$

where $\|\cdot\|_E$ is the appropriate sup-norm constrained to the set E. As

$$\|H_n - F_\perp^\circ\| = \|\tau_n[F_{n,\perp} - F_\perp^\circ]\| \le \max_{1 \le k \le d} \|\tau_{n,k}\| \|F_{n,\perp} - F_\perp^\circ\| \le \|F_{n,\perp} - F_\perp^\circ\| = o_P(1)$$

by the Glivenko-Cantelli theorem, the assumed continuity of ζ'' on $(0,1)^d$ implies that ζ'' is uniformly continuous on S_γ. Hence, $\|\zeta''(H_n(X_i), \theta_0) - \zeta''(F_\perp^\circ(x), \theta_0)\| = o_P(1)$. As $\|\mathbb{G}_{n,k}\| = O_P(1)$, this shows $D_{n,1,\gamma} = o_P(1)$. As for $D_{n,2,\gamma}$, its absolute value is bounded by

$$\left\|\frac{\mathbb{G}_{n,a}}{q_a}\right\| \left\|\frac{\mathbb{G}_{n,b}}{q_b}\right\| \left[\int_{S_\gamma^c} \left|q_a(x_a)\zeta''_{a,b}(H_n(x), \theta_0) q_b(x_b)\right| \,\mathrm{d}F_n(x) + \int_{S_\gamma^c} \left|q_a(x_a)\zeta''_{a,b}(F_\perp^\circ(x), \theta_0) q_b(x_b)\right| \,\mathrm{d}F_n(x)\right],$$

which by eq. (11) is bounded by

$$\begin{aligned} \left\|\frac{\mathbb{G}_{n,a}}{q_a}\right\| \left\|\frac{\mathbb{G}_{n,b}}{q_b}\right\| \Bigg[&\int_{S_\gamma^c} q_a(x_a) q_b(x_b) \tilde{r}_{a,b,1}(\tilde{x}_a)\tilde{r}_{a,b,2}(\tilde{x}_b) \prod_{1 \le k \le d, k \ne a,b} r_k(\tilde{x}_k) \,\mathrm{d}F_n(x) \\ &- \int_{S_\gamma^c} q_a(x_a) q_b(x_b) \tilde{r}_{a,b,1}(x_a)\tilde{r}_{a,b,2}(x_b) \prod_{1 \le k \le d, k \ne a,b} r_k(x_k) \,\mathrm{d}F_n(x)\Bigg] \end{aligned}$$

where $\tilde{x}_k = F_{n,\perp}(1, \ldots, 1, x_k, 1, \ldots, 1)$. By Lemma 4, we have $\|\mathbb{G}_{n,a}/q_a\| \|\mathbb{G}_{n,b}/q_b\| = O_P(1)$. It thus suffices to bound

$$D_{n,2,\gamma}(a,b,k,l) := \int_{S_\gamma^c} q_a(x_a) q_b(x_b) \tilde{r}_{a,b,1}(\tilde{x}_a)\tilde{r}_{a,b,2}(\tilde{x}_b) \prod_{1 \le k \le d, k \ne a,b} r_k(\tilde{x}_k) \,\mathrm{d}F_n(x)$$

$$\tilde{D}_{n,2,\gamma}(a,b,k,l) := \int_{S_\gamma^c} q_a(x_a) q_b(x_b) \tilde{r}_{a,b,1}(x_a)\tilde{r}_{a,b,2}(x_b) \prod_{1 \le k \le d, k \ne a,b} r_k(x_k) \,\mathrm{d}F_n(x)$$

By Lemma 5, there exists a constant $M_\varepsilon > 0$ such that

$$\tilde{\Omega}_\varepsilon = \left\{ \tilde{r}_{a,b,1}(\tilde{x}_a)\tilde{r}_{a,b,2}(\tilde{x}_b) \prod_{1\leq k\leq d, k\neq a,b} r_k(\tilde{x}_k) \leq M_\varepsilon \tilde{r}_{a,b,1}(x_a)\tilde{r}_{a,b,2}(x_b) \prod_{1\leq k\leq d, k\neq a,b} r_k(x_k) \right\}$$

with $P(\tilde{\Omega}_\varepsilon) > 1-\varepsilon$ for all n. On $\tilde{\Omega}_\varepsilon$, we have $D_{n,2,\gamma}(a,b,k,l) \leq M_\varepsilon \tilde{D}_{n,2,\gamma}(a,b,k,l)$. As ε is arbitrary, it suffices to bound $\tilde{D}_{n,2,\gamma}(a,b,k,l)$. We have

$$\mathbb{E}\left[|\tilde{D}_{n,2,\gamma}|\right] \leq \int_{S_\gamma^c} q_a(x_a)q_b(x_b)\tilde{r}_{a,b,1}(x_a)\tilde{r}_{a,b,2}(x_b) \prod_{1\leq k\leq d, k\neq a,b} r_k(x_k)\,\mathrm{d}F^\circ(x).$$

By the integrability assumption in eq. (12), this expectation converges to zero by the Dominated Convergence Theorem. □

References

Bickel, P., Klaassen, A., Ritov, Y. & Wellner, J. (1993). *Efficient and adaptive inference in semi-parametric models.* Johns Hopkins University Press, Baltimore.

Chen, X. & Fan, Y. (2005). Pseudo-likelihood ratio tests for semiparametric multivariate copula model selection. *The Canadian Journal of Statistics* **33**, 389–414.

Claeskens, G. & Hjort, N. (2008). *Model Selection and Model Averaging.* Cambridge University Press.

Fermanian, J., Radulović, D. & Wegkamp, M. (2004). Weak convergence of empirical copula processes. *Bernoulli* **10**, 847–860.

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

Grønneberg, S. (2010). The copula information criterion and its implications for the maximum pseudo likelihood estimator. In *Dependence Modeling: Handbook on Vine Copulae*, D. Kurowicka & H. Joe, eds., chap. 6. World Scientific, pp. 131–163.

Joe, H. (1997). *Multivariate Models and Dependence Concepts.* Chapman & Hall.

Pyke, R. & Shorack, G. (1968). Weak convergence of a two-sample empirical process and a new approach to Chernoff-Savage theorems. *The Annals of Mathematical Statistics* **39**, 755–771.

Ruymgaart, F. H. (1974). Asymptotic normality of nonparametric tests for independence. *The Annals of Statistics* **2**, 892–910.

Ruymgaart, F. H., Shorack, G. R. & van Zwet, W. R. (1972). Asymptotic normality of nonparametric tests for independence. *The Annals of Mathematical Statistics* **43**, 1122–1135.

Tsukahara, H. (2005). Semiparametric estimation in copula models. *The Canadian Journal of Statistics* **33**, 357–375.

van der Vaart, A. W. & Wellner, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer Series in Statistics. Springer.

Department of Mathematics, University of Oslo, P.O. Box 1053 Blindern, N-0316 Oslo, Norway
E-mail address: steffeng@math.uio.no

Department of Mathematics, University of Oslo, P.O. Box 1053 Blindern, N-0316 Oslo, Norway
E-mail address: nils@math.uio.no