FIC for copula under two-stage maximum likelihood

Vinnie Ko (PhD candidate)
Supervisors: Nils Lid Hjort & Ingrid Hobæk Haff

BBB, Universitetet i Oslo
23 May 2017

What is Copula?

We all know how the transformation of random variable is done.

Example: bivariate joint density:

$$f_{x,y}(x, y) \to f_{u,v}(u, v)$$

where

$$\begin{array}{c} u = g_1(x) \\ v = g_2(y) \end{array} \quad \text{and} \quad \begin{array}{c} x = g_1^{-1}(u) \\ y = g_2^{-1}(v) \end{array}$$

This can be done by:

$$f_{u,v}(u, v) = f_{x,y}\left(g_1^{-1}(u), g_2^{-1}(v)\right) |J(u, v)|$$

The cdf $F_{u,v}(u, v)$ can be obtained by:

$$F_{u,v}(u, v) = \int_{-\infty}^{v} \int_{-\infty}^{u} f_{u',v'}(u', v')\, du'\, dv'$$

What is Copula?

Consider the case when

$$u = g_1(x) = F_x(x)$$
$$v = g_2(y) = F_y(y)$$

This lead us to the following transformation:

$$F_{x,y}(x, y) \rightarrow F_{F_x,F_y}(F_x(x), F_y(y))$$

Yes, right. We transformed (x, y) with its own marginal distributions $F_x(x)$ and $F_y(y)$. This $F_{F_x,F_y}(F_x(x), F_y(y))$ has a special name **copula** and denoted as **C**.

N.B.: The transformation $F_{x,y}(x, y) \rightarrow F_{F_x,F_y}(F_x(x), F_y(y))$ has a special name:
Probability integral transform (PIT)

What is Copula?

So, we have:

$$F_{1,\cdots,d}(x_1,\cdots,x_d) = C_{1,\cdots,d}(F_1(x_1),\cdots,F_d(x_d))$$

According to the Sklar's theorm [5], there always exists a function C that satisfies this. When $F_{1,\cdots,d}(x_1,\cdots,x_d)$ is continuous, $C_{1,\cdots,d}$ is unique.

Furtheremore, When $F_1(x_1),\cdots,F_d(x_d)$ are absolutely continuous and strictly increasing, C can be differentiated:

$$f_{1,\cdots,d}(x_1,\cdots,x_d) = c_{1,\cdots,d}(F_1(x_1),\cdots,F_d(x_d))\prod_{i=1}^{d} f_i(x_i)$$

where

$$c_{1,\cdots,d} = \frac{\partial^d C_{1,\cdots,d}}{\partial F_1(x_1),\cdots,F_d(x_d)} \quad \text{and} \quad f(x_i) = \frac{\partial F_i(x_i)}{\partial x_i}$$

FIC (Jullum, M. & Hjort, N. L. (2017)): Basic idea

$T(\cdot) =$ focus parameter (e.g. mean, $P(a < y < b)$)
$G =$ true distribution function
$\widehat{F} =$ distribution function from our estimated model

Idea: Not measure the model, but measure the specific performance of the model.

How? Minimize risk!

$$R(T(G), T(\widehat{F})) = \mathrm{E}_G \left[\left(T(G) - T(\widehat{F}) \right)^2 \right] = \mathrm{mse}(T(\widehat{F})) = \mathrm{Var}(T(\widehat{F})) + \mathrm{bias}\left(T(\widehat{F}), T(G) \right)^2$$

Focused information criterion: $\mathrm{FIC} = \widehat{\mathrm{mse}}(T(\widehat{F})) = \widehat{\mathrm{Var}}(T(\widehat{F})) + \widehat{\mathrm{bias}^2}\left(T(\widehat{F}), T(G) \right)$

FIC (Jullum, M. & Hjort, N. L. (2017)): Estimation of $\widehat{\mathrm{bias}^2}$

Focused information criterion: $\mathrm{FIC} = \widehat{\mathrm{mse}}(T(\widehat{F})) = \widehat{\mathrm{Var}}(T(\widehat{F})) + \widehat{\mathrm{bias}^2}\left(T(\widehat{F}), T(G)\right)$

Estimation of $\widehat{\mathrm{bias}^2}$ $\left(\neq \widehat{\mathrm{bias}}^2\right)$:

When model is non-parametric:

$$\lim_{n\to\infty} \mathrm{bias}^2\left(\widehat{T}_{\mathsf{np}}, T_{\mathsf{true}}\right) = \lim_{n\to\infty} \mathrm{E}\left[\widehat{T}_{\mathsf{np}} - T_{\mathsf{true}}\right]^2 = 0$$

When model is parametric:

$$\begin{aligned}
\mathrm{bias}\left(\widehat{T}_{\mathsf{pm}}, T_{\mathsf{true}}\right) &= \mathrm{E}\left[\widehat{T}_{\mathsf{pm}} - T_{\mathsf{true}}\right] \\
&= \mathrm{E}\left[\widehat{T}_{\mathsf{pm}} - T_{0,\mathsf{pm}} + T_{0,\mathsf{pm}} - T_{\mathsf{true}}\right] \\
&= \mathrm{E}\left[\widehat{T}_{\mathsf{pm}} - T_{0,\mathsf{pm}}\right] + T_{0,\mathsf{pm}} - T_{\mathsf{true}} \\
&\approx 0 + T_{0,\mathsf{pm}} - T_{\mathsf{true}}
\end{aligned}$$

Natural choice of $\widehat{\mathrm{bias}}\left(\widehat{T}_{\mathsf{pm}}, T_{\mathsf{true}}\right)$:

$$\widehat{\mathrm{bias}}\left(\widehat{T}_{\mathsf{pm}}, T_{\mathsf{true}}\right) = \widehat{T}_{\mathsf{pm}} - \widehat{T}_{\mathsf{np}}$$

FIC (Jullum, M. & Hjort, N. L. (2017)): Estimation of $\widehat{\text{bias}^2}$

$$\widehat{\text{bias}^2}^{*}\left(\widehat{T}_{\text{pm}}, T_{\text{true}}\right) = \widehat{\text{bias}}\left(\widehat{T}_{\text{pm}}, T_{\text{true}}\right)^2 = \left(\widehat{T}_{\text{pm}} - \widehat{T}_{\text{np}}\right)^2$$

Check whether $\widehat{\text{bias}^2}^{*}\left(\widehat{T}_{\text{pm}}, T_{\text{true}}\right)$ is biased

$$\mathrm{E}\left[\widehat{\text{bias}^2}^{*}\left(\widehat{T}_{\text{pm}}, T_{\text{true}}\right)\right] = \mathrm{E}\left[\left(\widehat{T}_{\text{pm}} - \widehat{T}_{\text{np}}\right)^2\right]$$

Use: $\mathrm{Var}(X) = \mathrm{E}\left[(X)^2\right] - (\mathrm{E}\left[X\right])^2$

$$\begin{aligned}
&= \left(\mathrm{E}\left[\widehat{T}_{\text{pm}} - \widehat{T}_{\text{np}}\right]\right)^2 + \mathrm{Var}(\widehat{T}_{\text{pm}} - \widehat{T}_{\text{np}}) \\
&\approx (T_{0,\text{pm}} - T_{\text{true}})^2 + \mathrm{Var}(\widehat{T}_{\text{pm}} - \widehat{T}_{\text{np}}) \\
&\approx \left(\widehat{T}_{\text{pm}} - T_{0,\text{pm}} + T_{0,\text{pm}} - T_{\text{true}}\right)^2 + \mathrm{Var}(\widehat{T}_{\text{pm}} - \widehat{T}_{\text{np}}) \\
&= \left(\widehat{T}_{\text{pm}} - T_{\text{true}}\right)^2 + \mathrm{Var}(\widehat{T}_{\text{pm}} - \widehat{T}_{\text{np}}) \\
&= \widehat{\text{bias}^2}\left(\widehat{T}_{\text{pm}}, T_{\text{true}}\right) + \mathrm{Var}(\widehat{T}_{\text{pm}} - \widehat{T}_{\text{np}})
\end{aligned}$$

It's biased!

FIC (Jullum, M. & Hjort, N. L. (2017)): Final form

Our final unbiased estimator:

$$\widehat{\text{bias}^2}\left(\widehat{T}_{\mathsf{pm}}, T_{\mathsf{true}}\right) = \left(\widehat{T}_{\mathsf{pm}} - \widehat{T}_{\mathsf{np}}\right)^2 - \text{Var}(\widehat{T}_{\mathsf{pm}} - \widehat{T}_{\mathsf{np}})$$

Finally, FIC becomes:

When model is parametric

$$\begin{aligned}
\text{FIC} &= \widehat{\text{Var}}(\widehat{T}_{\mathsf{pm}}) + \widehat{\text{bias}^2}\left(\widehat{T}_{\mathsf{pm}}, T_{\mathsf{true}}\right) \\
&= \widehat{\text{Var}}(\widehat{T}_{\mathsf{pm}}) + \left(\widehat{T}_{\mathsf{pm}} - \widehat{T}_{\mathsf{np}}\right)^2 - \widehat{\text{Var}}(\widehat{T}_{\mathsf{pm}} - \widehat{T}_{\mathsf{np}}) \\
&= \widehat{\text{Var}}(\widehat{T}_{\mathsf{pm}}) + \left(\widehat{T}_{\mathsf{pm}} - \widehat{T}_{\mathsf{np}}\right)^2 - \widehat{\text{Var}}(\widehat{T}_{\mathsf{pm}}) - \widehat{\text{Var}}(\widehat{T}_{\mathsf{np}}) + 2\widehat{\text{Cov}}(\widehat{T}_{\mathsf{np}}, \widehat{T}_{\mathsf{pm}}) \\
&= \left(\widehat{T}_{\mathsf{pm}} - \widehat{T}_{\mathsf{np}}\right)^2 - \widehat{\text{Var}}(\widehat{T}_{\mathsf{np}}) + 2\widehat{\text{Cov}}(\widehat{T}_{\mathsf{np}}, \widehat{T}_{\mathsf{pm}})
\end{aligned}$$

When model is non-parametric

$$\begin{aligned}
\text{FIC} &= \widehat{\text{Var}}(\widehat{T}_{\mathsf{np}}) + \widehat{\text{bias}^2}\left(\widehat{T}_{\mathsf{np}}, T_{\mathsf{true}}\right) \\
&= \widehat{\text{Var}}(\widehat{T}_{\mathsf{np}})
\end{aligned}$$

So far, nothing specific about model or how model parameters are estimated.
However, we need asymptotic theories of model & estimation to obtain $\widehat{\text{Cov}}(\widehat{T}_{\mathsf{np}}, \widehat{T}_{\mathsf{pm}})$.

Example: copula by maximum likelihood estimator

Maximum likelihood estimator:
The log-likelihood function is defined as

$$\begin{aligned}\ell_n &= \sum_{i=1}^{n} \log f(y_{i,1}, \cdots, y_{i,d}) \\ &= \sum_{i=1}^{n} [\log f_1(y_{i,1}, \alpha_1) + \cdots + \log f_d(y_{i,d}, \alpha_d) + \log c\left(F_1(y_{i,1}, \alpha_1), \cdots, F_d(y_{i,d}, \alpha_d), \theta\right)]\end{aligned}$$

Find the estimate of the parameter vector $\widehat{\eta} = (\widehat{\alpha}_1^{\mathrm{T}}, \cdots, \widehat{\alpha}_d^{\mathrm{T}}, \widehat{\theta}^{\mathrm{T}})^{\mathrm{T}}$ by solving $\left(\frac{\partial \ell_n}{\partial \alpha_1}, \cdots, \frac{\partial \ell_n}{\partial \alpha_d}, \frac{\partial \ell_n}{\partial \theta}\right) = 0.$

Example: copula by maximum likelihood estimator

We are looking for asymptotic covariance: $\mathrm{Cov}(\widehat{T}_{\mathsf{np}}, \widehat{T}_{\mathsf{pm}})$

Main ingredients:

$$\widehat{T}_{\mathsf{np}} - T_{\mathsf{true}} = \frac{1}{n}\sum_{i=1}^{n} \mathrm{IF}(y_i; G) + o_{\mathsf{pr}}(n^{-\frac{1}{2}})$$

$$\widehat{\eta} - \eta_0 = \mathcal{I}^{-1}\frac{1}{n}\sum_{i=1}^{n} u(y_i; \eta_0) + o_{\mathsf{pr}}(n^{-\frac{1}{2}})$$

where

$$u(y, \theta) = \frac{\partial \log f(y, \eta)}{\partial \eta}$$

$$\mathcal{I} = -\mathrm{E}_G\left[H(y, \eta_0)\right] = -\mathrm{E}_G\left[\frac{\partial^2 \log f(y, \eta)}{\partial \eta \partial \eta^{\mathrm{T}}}\Big|_{\eta=\eta_0}\right]$$

Example: copula by maximum likelihood estimator

Main ingredients:

$$\widehat{T}_{\mathrm{np}} - T_{\mathrm{true}} = \frac{1}{n}\sum_{i=1}^{n} \mathrm{IF}(y_i; G) + o_{\mathrm{pr}}(n^{-\frac{1}{2}})$$

$$\widehat{\eta} - \eta_0 = \mathcal{I}^{-1}\frac{1}{n}\sum_{i=1}^{n} u(y_i; \eta_0) + o_{\mathrm{pr}}(n^{-\frac{1}{2}})$$

The multivariate central limit theorem gives us

$$\sqrt{n}\begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} \mathrm{IF}(y_i; G) \\ \\ \frac{1}{n}\sum_{i=1}^{n} u(y_i; \eta_0) \end{pmatrix} \overset{d}{\to} \begin{pmatrix} \Lambda_{\mathrm{np}} \\ \Lambda_{\mathrm{pm}} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \nu & Q \\ Q^{\mathrm{T}} & K \end{pmatrix}\right)$$

where

$$\nu = \mathrm{E}\left[\mathrm{IF}(y; G)^2\right]$$

$$K = \mathrm{E}\left[u(y, \eta_0)\, u(y, \eta_0)^{\mathrm{T}}\right]$$

$$Q = \mathrm{E}\left[\mathrm{IF}(y; G) u(y, \eta_0)\right]$$

Example: copula by maximum likelihood estimator

Apply Slutsky's theorem.

$$
\begin{aligned}
\sqrt{n}\begin{pmatrix}\widehat{T}_{\mathrm{np}} - T_{\mathrm{true}} \\ \widehat{\eta} - \eta_0\end{pmatrix} &= \sqrt{n}\begin{pmatrix}\frac{1}{n}\sum_{i=1}^{n}\mathrm{IF}(y_i; G) + o_p(n^{-\frac{1}{2}}) \\ \mathcal{I}^{-1}\frac{1}{n}\sum_{i=1}^{n} u(y_i;\eta_0) + o_p(n^{-\frac{1}{2}})\end{pmatrix} \\
&= \sqrt{n}\begin{pmatrix}1 & 0 \\ 0 & \mathcal{I}^{-1}\end{pmatrix}\begin{pmatrix}\frac{1}{n}\sum_{i=1}^{n}\mathrm{IF}(y_i; G) \\ \frac{1}{n}\sum_{i=1}^{n} u(y_i;\eta_0)\end{pmatrix} + \begin{pmatrix}o_p(1) \\ o_p(1)\end{pmatrix} \\
&\xrightarrow{d} \begin{pmatrix}1 & 0 \\ 0 & \mathcal{I}^{-1}\end{pmatrix}\begin{pmatrix}\Lambda_{\mathrm{np}} \\ \Lambda_{\mathrm{pm}}\end{pmatrix} \\
&\sim N\left(\begin{pmatrix}0 \\ 0\end{pmatrix}, \begin{pmatrix}1 & 0 \\ 0 & \mathcal{I}^{-1}\end{pmatrix}\begin{pmatrix}\nu & Q \\ Q^{\mathrm{T}} & K\end{pmatrix}\begin{pmatrix}1 & 0 \\ 0 & \mathcal{I}^{-1}\end{pmatrix}\right) \\
&= N\left(\begin{pmatrix}0 \\ 0\end{pmatrix}, \begin{pmatrix}\nu & Q\mathcal{I}^{-1} \\ \mathcal{I}^{-1}Q^{\mathrm{T}} & \mathcal{I}^{-1}K\mathcal{I}^{-1}\end{pmatrix}\right)
\end{aligned}
$$

Example: copula by maximum likelihood estimator

We have now:

$$\sqrt{n}\begin{pmatrix}\widehat{T}_{\mathrm{np}} - T_{\mathrm{true}} \\ \widehat{\eta} - \eta_0\end{pmatrix} \overset{d}{\to} \begin{pmatrix}1 & 0 \\ 0 & \mathcal{I}^{-1}\end{pmatrix}\begin{pmatrix}\Lambda_{\mathrm{np}} \\ \Lambda_{\mathrm{pm}}\end{pmatrix} \sim N\left(\begin{pmatrix}0 \\ 0\end{pmatrix}, \begin{pmatrix}\nu & Q\mathcal{I}^{-1} \\ \mathcal{I}^{-1}Q^{\mathrm{T}} & \mathcal{I}^{-1}K\mathcal{I}^{-1}\end{pmatrix}\right)$$

Apply delta method with the following transformation function

$$S(u, v) = \begin{pmatrix}u \\ T_{\mathrm{pm}}(v)\end{pmatrix} \quad (v \text{ is the parameter vector of the parametric model.})$$

The corresponding Jacobian is

$$\dot{S}(u, v) = \begin{pmatrix}\frac{\partial f_1}{\partial u} & \frac{\partial f_1}{\partial v} \\ \\ \frac{\partial f_2}{\partial u} & \frac{\partial f_2}{\partial v}\end{pmatrix} = \begin{pmatrix}1 & 0 \\ \\ 0 & \left(\frac{\partial T_{\mathrm{pm}}(v)}{\partial v}\right)^{\mathrm{T}}\end{pmatrix}$$

Example: copula by maximum likelihood estimator

The delta method gives

$$\sqrt{n}\begin{pmatrix}\widehat{T}_{\mathrm{np}} - T_{\mathrm{true}} \\ \widehat{T}_{\mathrm{pm}} - T_{0,\mathrm{pm}}\end{pmatrix} = \sqrt{n}\begin{pmatrix}\widehat{T}_{\mathrm{np}} - T_{\mathrm{true}} \\ T_{\mathrm{pm}}(\widehat{\eta}) - T_{\mathrm{pm}}(\eta_0)\end{pmatrix}$$

$$\xrightarrow{d} \left(\begin{pmatrix}0\\0\end{pmatrix}, \dot{S}(T_{\mathrm{true}}, \eta_0)\begin{pmatrix}\nu & Q\mathcal{I}^{-1} \\ \mathcal{I}^{-1}Q^{\mathrm{T}} & \mathcal{I}^{-1}K\mathcal{I}^{-1}\end{pmatrix}\dot{S}(T_{\mathrm{true}}, \eta_0)^{\mathrm{T}}\right)$$

$$\sim N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & 0 \\ 0 & \left(\frac{\partial T_{\mathrm{pm}}(\eta_0)}{\partial \eta_0}\right)^{\mathrm{T}}\end{pmatrix}\begin{pmatrix}\nu & Q\mathcal{I}^{-1} \\ \mathcal{I}^{-1}Q^{\mathrm{T}} & \mathcal{I}^{-1}K\mathcal{I}^{-1}\end{pmatrix}\begin{pmatrix}1 & 0 \\ 0 & \frac{\partial T_{\mathrm{pm}}(\eta_0)}{\partial \eta_0}\end{pmatrix}\right)$$

$$= N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\nu & Q\mathcal{I}^{-1}c \\ c^{\mathrm{T}}\mathcal{I}^{-1}Q^{\mathrm{T}} & c^{\mathrm{T}}\mathcal{I}^{-1}K\mathcal{I}^{-1}c\end{pmatrix}\right)$$

where $c = \frac{\partial T_{\mathrm{pm}}(\eta_0)}{\partial \eta_0}$.

So,

$\mathrm{Var}(\widehat{T}_{\mathrm{np}}) = \nu$

$\mathrm{Var}(\widehat{T}_{\mathrm{pm}}) = c^{\mathrm{T}}\mathcal{I}^{-1}K\mathcal{I}^{-1}c$

$\mathrm{Cov}(\widehat{T}_{\mathrm{np}}, \widehat{T}_{\mathrm{pm}}) = Q\mathcal{I}^{-1}c$

Example: copula by maximum likelihood estimator

We have:

$$\mathrm{Var}(\widehat{T}_{\mathsf{np}}) = \nu$$
$$\mathrm{Var}(\widehat{T}_{\mathsf{pm}}) = c^{\mathrm{T}}\mathcal{I}^{-1}K\mathcal{I}^{-1}c$$
$$\mathrm{Cov}(\widehat{T}_{\mathsf{np}}, \widehat{T}_{\mathsf{pm}}) = Q\mathcal{I}^{-1}\, c$$

where

$$\mathcal{I} = -\mathrm{E}_G\left[H(y,\eta_0)\right]$$
$$\nu = \mathrm{E}\left[\mathrm{IF}(y;G)^2\right]$$
$$K = \mathrm{E}\left[u(y,\eta_0)\,u(y,\eta_0)^{\mathrm{T}}\right]$$
$$Q = \mathrm{E}\left[\mathrm{IF}(y;G)u(y,\eta_0)\right]$$
$$c = \frac{\partial T_{\mathrm{pm}}(\eta_0)}{\partial \eta_0}$$

We define the estimated (co)variances as their sample based plug-in equivalent

We have:

$$\widehat{\mathrm{Var}}(\widehat{T}_{\mathsf{np}}) = \widehat{\nu}$$
$$\widehat{\mathrm{Var}}(\widehat{T}_{\mathsf{pm}}) = \widehat{c}^{\mathrm{T}}\widehat{\mathcal{I}}^{-1}\widehat{K}\widehat{\mathcal{I}}^{-1}\widehat{c}$$
$$\widehat{\mathrm{Cov}}(\widehat{T}_{\mathsf{np}}, \widehat{T}_{\mathsf{pm}}) = \widehat{Q}\widehat{\mathcal{I}}^{-1}\,\widehat{c}$$

where

$$\widehat{\mathcal{I}} = -\frac{1}{n}\sum_{i=1}^{n} H(y_i,\widehat{\eta})$$
$$\widehat{K} = \frac{1}{n}\sum_{i=1}^{n} u(y_i,\widehat{\eta})u(y_i,\widehat{\eta})^{\mathrm{T}}$$
$$\widehat{\nu} = \frac{1}{n}\sum_{i=1}^{n} \mathrm{IF}(y_i;\widehat{G})^2$$
$$\widehat{Q} = \frac{1}{n}\sum_{i=1}^{n} \mathrm{IF}(y_i;\widehat{G})u(y_i,\widehat{\eta})$$

Copula by two-stage maximum likelihood (a.k.a. IFM)

$$\ell_n = \sum_{i=1}^{n} [\log f_1(y_{i,1}, \alpha_1) + \cdots + \log f_d(y_{i,d}, \alpha_d) + \log c\,(F_1(y_{i,1}, \alpha_1), \cdots, F_d(y_{i,d}, \alpha_d), \theta)]$$

Full maximum likelihood is not always feasible when dimension is high.
Popular alternative: two-stage maximum likelihood

- **Stage 1**:
Estimate $\tilde{\alpha}_j$ of each margin, seperately by using ML:
For $1 \le j \le d$, obtain $\tilde{\alpha}_j$ by solving $\frac{\partial \ell_{\alpha_j,n}}{\partial \alpha_j} = 0$ where $\ell_{\alpha_j,n} = \sum_{i=1}^{n} \log f_j(y_{i,j}, \alpha_j)$.

- **Stage 2**:
Plug-in $\tilde{\alpha}_1, \cdots, \tilde{\alpha}_d$ from stage 1 into ℓ_n. The resulting function is

$$\tilde{\ell}_\theta = \ell_n(\tilde{\alpha}, \theta) = \sum_{i=1}^{n} [\log f_1(y_{i,1}, \tilde{\alpha}_1) + \cdots + \log f_d(y_{i,d}, \tilde{\alpha}_d) + \log c\,(F_1(y_{i,1}, \tilde{\alpha}_1), \cdots, F_d(y_{i,d}, \tilde{\alpha}_d), \theta)]$$

Estimate $\tilde{\theta}$ by maximizing $\tilde{\ell}_\theta$.

Resulting estimated parameter vector: $\tilde{\eta} = (\tilde{\alpha}_1^{\mathrm{T}}, \cdots, \tilde{\alpha}_d^{\mathrm{T}}, \tilde{\theta}^{\mathrm{T}})^{\mathrm{T}}$

Copula by two-stage maximum likelihood

To compute FIC, we need $\widehat{\mathrm{Var}}(\widehat{T}_{\mathrm{np}})$, $\widehat{\mathrm{Cov}}(\widehat{T}_{\mathrm{np}}, \widehat{T}_{\mathrm{pm}})$.

Main ingredients:

$$\begin{cases} \text{Non-para:} & \widehat{T}_{\mathrm{np}} - T_{\mathrm{true}} = \frac{1}{n}\sum_{i=1}^{n} \mathrm{IF}(y_i; G) + o_{\mathrm{pr}}(n^{-\frac{1}{2}}) \\ \text{Two-stage ML:} & \tilde{\eta} - \eta_0^{\mathrm{2ML}} = ??? \end{cases}$$

$$\text{Full ML:} \quad \widehat{\eta} - \eta_0^{\mathrm{ML}} = \mathcal{I}^{-1}\frac{1}{n}\sum_{i=1}^{n} u(y_i; \eta_0) + o_{\mathrm{pr}}(n^{-\frac{1}{2}})$$

Same tricks: multivariate CLT + Slutsky's theorem + delta method

Once we have $\tilde{\eta} - \eta_0^{\mathrm{2ML}} = ???$, the job is basically done.

Asymptotic distribution of two-stage maximum likelihood (under development)

$$\text{Lemma 0.1:}\quad \tilde{\theta} - \theta_0 = \mathcal{I}_\theta^{-1} \widetilde{U}_{\theta,n}(\theta_0) + o_p(n^{-\frac{1}{2}})$$

$$\text{Lemma 0.2:}\quad \widetilde{U}_{\theta,n}(\tilde{\alpha}, \theta_0) = U_{\theta,n}(\alpha_0, \theta_0) + L\left(\tilde{\alpha} - \alpha_0\right) + o_p(n^{-\frac{1}{2}})$$

$$\text{Lemma 0.3:}\quad \tilde{\theta} - \theta_0 = \mathcal{I}_\theta^{-1}\left(U_{\theta,n}(\alpha_0, \theta_0) + L\left(\tilde{\alpha} - \alpha_0\right)\right) + o_p(n^{-\frac{1}{2}})$$

$$\text{Lemma 0.4:}\quad \sqrt{n}\left(\tilde{\theta} - \theta_0\right) \overset{d}{\to} \mathcal{I}_\theta^{-1}\left(\Lambda_\theta + L\mathcal{I}_\alpha^{-1}\Lambda_\alpha\right) \sim N\left(0, \mathcal{I}_\theta^{-1} K_\theta^* \mathcal{I}_\theta^{-1}\right)$$

"The master lemma":

$$\text{Lemma 1:}\quad \sqrt{n}\left(\tilde{\eta} - \eta_0\right) \overset{d}{\to} \mathcal{I}_\eta^{-1} L^* \Lambda_\eta \sim N\left(0,\, \mathcal{I}_\eta^{-1} L^* K_\eta (L^*)^{\mathrm{T}} \mathcal{I}_\eta^{-1}\right)$$

Asymptotic distribution of two-stage maximum likelihood (under development)

N.B. We are in two-stage ML setting:
We write simply $\eta_0 = \eta_0^{\mathrm{2ML}}$. (Full) ML estimate will be (later) denoted as η_0^{ML}.

Define:

$$U_{\theta,n}(\alpha, \theta) = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial \log c\left(F_1(y_{i,1}, \alpha_1), \cdots, F_d(y_{i,d}, \alpha_d), \theta\right)}{\partial \theta}$$

$$\widetilde{U}_{\theta,n}(\theta) = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial \log c\left(F_1(y_{i,1}, \tilde{\alpha}_1), \cdots, F_d(y_{i,d}, \tilde{\alpha}_d), \theta\right)}{\partial \theta},$$

Other matrices, including L, will be defined later.

Asymptotic distribution of two-stage maximum likelihood (under development)

$$\text{Lemma 0.1:} \quad \tilde{\theta} - \theta_0 = \mathcal{I}_\theta^{-1} \widetilde{U}_{\theta,n}(\theta_0) + o_p(n^{-\frac{1}{2}})$$

$$\text{Lemma 0.2:} \quad \widetilde{U}_{\theta,n}(\tilde{\alpha}, \theta_0) = U_{\theta,n}(\alpha_0, \theta_0) + L\left(\tilde{\alpha} - \alpha_0\right) + o_p(n^{-\frac{1}{2}})$$

$$\text{Lemma 0.3:} \quad \tilde{\theta} - \theta_0 = \mathcal{I}_\theta^{-1}\left(U_{\theta,n}(\alpha_0, \theta_0) + L\left(\tilde{\alpha} - \alpha_0\right)\right) + o_p(n^{-\frac{1}{2}})$$

$$\text{Lemma 0.4:} \quad \sqrt{n}\left(\tilde{\theta} - \theta_0\right) \overset{d}{\to} \mathcal{I}_\theta^{-1}\left(\Lambda_\theta + L\mathcal{I}_\alpha^{-1}\Lambda_\alpha\right) \sim N\left(0, \mathcal{I}_\theta^{-1} K_\theta^* \mathcal{I}_\theta^{-1}\right)$$

"The master lemma":

$$\text{Lemma 1:} \quad \sqrt{n}\left(\tilde{\eta} - \eta_0\right) \overset{d}{\to} \mathcal{I}_\eta^{-1} L^* \Lambda_\eta \sim N\left(0, \mathcal{I}_\eta^{-1} L^* K_\eta (L^*)^{\mathrm{T}} \mathcal{I}_\eta^{-1}\right)$$

Asymptotic distribution of two-stage maximum likelihood (under development)

$$\text{Lemma 1:} \quad \sqrt{n}\,(\tilde{\eta} - \eta_0) \xrightarrow{d} \mathcal{I}_\eta^{-1} L^* \Lambda_\eta \sim N\left(0, \mathcal{I}_\eta^{-1} L^* K_\eta (L^*)^{\mathrm{T}} \mathcal{I}_\eta^{-1}\right)$$

where

$$\Lambda_\eta = \begin{pmatrix} \Lambda_\alpha \\ \Lambda_\theta \end{pmatrix}, \quad \mathcal{I}_\eta = \begin{pmatrix} \mathcal{I}_\alpha & 0 \\ 0 & \mathcal{I}_\theta \end{pmatrix}, \quad L^* = \begin{pmatrix} 1 & 0 \\ L\,\mathcal{I}_\alpha^{-1} & 1 \end{pmatrix}, \quad K_\eta = \begin{pmatrix} K_\alpha & K_{\alpha,\theta} \\ K_{\alpha,\theta}^{\mathrm{T}} & K_\theta \end{pmatrix}$$

$$\mathcal{I}_\alpha = \begin{pmatrix} \mathcal{I}_{\alpha_1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \mathcal{I}_{\alpha_d} \end{pmatrix}, \quad K_\alpha = \begin{pmatrix} K_{\alpha_1} & K_{\alpha_1,\alpha_2} & \cdots & K_{\alpha_1,\alpha_d} \\ K_{\alpha_2,\alpha_1} & K_{\alpha_2} & \cdots & K_{\alpha_2,\alpha_d} \\ \vdots & \vdots & \ddots & \vdots \\ K_{\alpha_d,\alpha_1} & K_{\alpha_d,\alpha_2} & \cdots & K_{\alpha_d} \end{pmatrix},$$

$$\begin{aligned}
&\mathcal{I}_{\alpha_j} = -\mathrm{E}_G\left[H_{\alpha_j}(y, \alpha_{0,j})\right] \\
&\mathcal{I}_\theta = -\mathrm{E}_G\left[H_\theta(Y, \alpha_0, \theta_0)\right] \\
&K_{\alpha_j} = \mathrm{E}_G\left[u_{\alpha_j}(y; \alpha_{0,j}) u_{\alpha_j}(y; \alpha_{0,j})^{\mathrm{T}}\right] \\
&K_\theta = \mathrm{E}_G\left[u_\theta(y, \alpha_0, \theta_0)\, u_\theta(y, \alpha_0, \theta_0)^{\mathrm{T}}\right] \\
&K_{\alpha,\theta} = E\left[u_\alpha(y; \alpha_0)\, u_\theta(y, \alpha_0, \theta_0)^{\mathrm{T}}\right].
\end{aligned}$$

$$L = \int g \frac{\partial^2 \log c\left(F_1(y_1, \alpha_1), \cdots, F_d(y_d, \alpha_d), \theta\right)}{\partial\theta\alpha^{\mathrm{T}}}\Big|_{\alpha=\alpha_0, \theta=\theta_0} \mathrm{d}y$$

Asymptotic distribution of two-stage maximum likelihood (under development)

Comparision: full ML vs two-stage ML

$$\text{Full ML:} \quad \sqrt{n}\left(\tilde{\eta} - \eta_0^{\mathrm{ML}}\right) \overset{d}{\to} \mathcal{I}_\eta^{-1}\Lambda_\eta \sim N\left(0, \mathcal{I}_\eta^{-1}K_\eta\mathcal{I}_\eta^{-1}\right)$$

$$\text{Two-stage ML:} \quad \sqrt{n}\left(\tilde{\eta} - \eta_0\right) \overset{d}{\to} \mathcal{I}_\eta^{-1}L^*\Lambda_\eta \sim N\left(0, \mathcal{I}_\eta^{-1}L^*K_\eta(L^*)^{\mathrm{T}}\mathcal{I}_\eta^{-1}\right) \quad \text{(lemma 1)}$$

Note that $\mathcal{I}_\eta^{-1}$ and K_η will be different between *full ML* and *two-stage ML* in general, because $\eta_0^{\mathrm{ML}} \neq \eta_0^{\mathrm{2ML}}$ unless our model contains the true model.

$$L^* = \begin{pmatrix} 1 & 0 \\ L\,\mathcal{I}_\alpha^{-1} & 1 \end{pmatrix}, \quad L = \int g \frac{\partial^2 \log c\left(F_1(y_1, \alpha_1), \cdots, F_d(y_d, \alpha_d), \theta\right)}{\partial\theta\alpha^{\mathrm{T}}}\Big|_{\alpha=\alpha_0, \theta=\theta_0} \mathrm{d}y$$

In the special case when $L = 0$, we have: *full ML* $=$ *two-stage ML*.

Is $\eta_0^{\mathrm{2ML}} = \eta_0^{\mathrm{ML}}$?

Is $\eta_0^{\mathrm{2ML}} = \eta_0^{\mathrm{ML}}$?
Answer: only if the our model captures the true model

$$\begin{aligned} \text{True density:} \quad & g(y_1, \cdots, y_d) = g_1(y_1) \cdots g_d(y_d) c_0 \left(G_1(y_1), \cdots, G_d(y_d) \right) \\ \text{Model density:} \quad & f(y_1, \cdots, y_d) = f_1(y_1, \alpha_1) \cdots f_d(y_d, \alpha_d) c \left(F_1(y_1, \alpha_1), \cdots, F_d(y_d, \alpha_d), \theta \right) \end{aligned}$$

Recall that maximum likelihood is aming for η_0 that minimizes Kullback-Leibler divergence. Under the true model assumption (i.e. the true model lines within our model space), $\eta_0^{\mathrm{ML}} = \left(\alpha_0^{\mathrm{ML}}, \theta_0^{\mathrm{ML}} \right)$ is the value that satisfies

$$\begin{aligned} \mathrm{KL}(g, f) &= \mathrm{KL}(g_1(y_1) \cdots g_d(y_d) c_0 \left(G_1(y_1), \cdots, G_d(y_d) \right), \\ &\qquad f_1(y_1, \alpha_{0,1}) \cdots f_d(y_d, \alpha_{0,d}) c \left(F_1(y_1, \alpha_{0,1}), \cdots, F_d(y_d, \alpha_{0,d}), \theta_0 \right)) \\ &= 0 \end{aligned}$$

Is $\eta_0^{\mathrm{2ML}} = \eta_0^{\mathrm{ML}}$?

$$\text{True density:} \quad g(y_1, \cdots, y_d) = g_1(y_1) \cdots g_d(y_d) c_0\left(G_1(y_1), \cdots, G_d(y_d)\right)$$
$$\text{Model density:} \quad f(y_1, \cdots, y_d) = f_1(y_1, \alpha_1) \cdots f_d(y_d, \alpha_d) c\left(F_1(y_1, \alpha_1), \cdots, F_d(y_d, \alpha_d), \theta\right)$$

$\mathrm{KL}(g, f) = 0$ is only possible when $f_j(Y, \alpha_{0,j}) = g_j$. (i.e. our model marginal density contains the true marginal density.)
So,

$$\begin{aligned} \alpha_0^{\mathrm{2ML}} &= \arg\min_{\alpha_j} \mathrm{KL}(g_j, f_j(Y, \alpha_j)) \\ &= \arg\max_{\alpha_j} \left\{ \int g_j \log f_j(y_j, \alpha_j) \, \mathrm{d}y_j \right\} \end{aligned}$$

should be equal to

$$\begin{aligned} \alpha_0^{\mathrm{ML}} &= \arg\min_{\eta} \mathrm{KL}(g, f(Y, \eta)) \\ &= \arg\max_{\eta} \left\{ \int g \left(\log f_1(y_1, \alpha_1) + \cdots + \log f_d(y_d, \alpha_d) + \log c\left(F_1(y_1, \alpha_1), \cdots, F_d(y_d, \alpha_d), \theta\right)\right) \mathrm{d}y \right\}. \end{aligned}$$

This automatically implies that $\theta_0^{\mathrm{2ML}} = \theta_0^{\mathrm{ML}}$.
So, we have $\eta_0^{\mathrm{2ML}} = \eta_0^{\mathrm{ML}}$.

When our model doesn't contain the true model, $\eta_0^{\mathrm{2ML}} \neq \eta_0^{\mathrm{ML}}$

Asymptotic distribution of two-stage maximum likelihood (under development)

Comparision: full ML vs two-stage ML

$$\text{Full ML:} \quad \sqrt{n}\left(\tilde{\eta} - \eta_0^{\mathrm{ML}}\right) \xrightarrow{d} \mathcal{I}_\eta^{-1}\Lambda_\eta \sim N\left(0, \mathcal{I}_\eta^{-1}K_\eta\mathcal{I}_\eta^{-1}\right)$$

$$\text{Two-stage ML:} \quad \sqrt{n}\left(\tilde{\eta} - \eta_0\right) \xrightarrow{d} \mathcal{I}_\eta^{-1}L^*\Lambda_\eta \sim N\left(0, \mathcal{I}_\eta^{-1}L^*K_\eta(L^*)^{\mathrm{T}}\mathcal{I}_\eta^{-1}\right) \quad \text{(lemma 1)}$$

Note that $\mathcal{I}_\eta^{-1}$ and K_η will be different between *full ML* and *two-stage ML* in general, because $\eta_0^{\mathrm{ML}} \neq \eta_0^{\mathrm{2ML}}$ unless our model contains the true model.

$$L^* = \begin{pmatrix} 1 & 0 \\ L\,\mathcal{I}_\alpha^{-1} & 1 \end{pmatrix}, \quad L = \int g \frac{\partial^2 \log c\left(F_1(y_1, \alpha_1), \cdots, F_d(y_d, \alpha_d), \theta\right)}{\partial\theta\alpha^{\mathrm{T}}}\Big|_{\alpha=\alpha_0, \theta=\theta_0} \mathrm{d}y$$

$L = 0$ leads to *full ML* $=$ *two-stage ML*.

End

To do:
- Proofread & implement lemma 1
- Possible bonus: a new Copula Information Criterion (CIC) under two-stage ML: $\hat{\ell}_{n,\max} - \hat{p}^*$

Questions:
- What does $L = 0$ mean?
- When is $L = 0$?
- When is L small in size?

Reference

[1] Ingrid Hobæk Haff et al.
Parameter estimation for pair-copula constructions.
Bernoulli, 19(2):462–491, 2013.

[2] Harry Joe.
Multivariate models and multivariate dependence concepts.
CRC Press, 1997.

[3] Martin Jullum and Nils Lid Hjort.
Parametric or nonparametric: the fic approach.
Statistica Sinica, 2017.

[4] Martin Jullum and Nils Lid Hjort.
Supplementary material: Parametric or nonparametric: the fic approach.
Statistica Sinica, 2017.

[5] Roger B Nelsen.
An introduction to copulas.
Springer Science & Business Media, 2007.

[6] Aad W Van der Vaart.
Asymptotic statistics, volume 3.
Cambridge university press, 2000.