# Estimation of Copulas ....

... under a Parametric Assumption on the Copula

Artur Bille

February 4, 2021

Institute of Stochastics
Ulm University

# Overview

(i) Setting & Outlook

(ii) Parametric Copula Families

(iii) Parametrically Estimated Margins

- Maximum Likelihood Estimator
- Inference Functions for Margins Estimator

(iv) Non-Parametrically Estimated Margins

- Method of Moments
- Kendall's Tau Representation

# Setting & Outlook

**Our situation:**

- $n$ copies of a $d$–dimensional random vector $X = (X_1, \dots, X_d)^\intercal$:

$$X_1 = \begin{pmatrix} X_1^1 \\ \vdots \\ X_d^1 \end{pmatrix}, \dots, X_n = \begin{pmatrix} X_1^n \\ \vdots \\ X_d^n \end{pmatrix}$$

- $X$ has an unknown (multivariate) distribution function $H$ and absolutely **continuous**, but unknown, margins $F_1, \dots, F_d$.

**Aim**

Estimate the Copula $C$, s.t. $H(x) = C(F_1(x_1), \dots, F_d(x_d))$, $x \in \mathbb{R}^d$.

**Today:** The wanted Copula is an element of a parametric family.
**Tomorrow:** The Non-Parametric case by Viet.

# Why unknown margins?

**Theorem (Stochastic Analog of Sklar's Theorem)**

*Let $X = (X_1, \ldots, X_d)^\intercal$ be a d–dimensional random vector with continuous univariate margins $F_1, \ldots, F_d$.*
*Then $X$ has copula $C$ if and only if $(F_1(X_1), \ldots, F_d(X_d)) \sim C$.*

# Element of a parametric family

A parametric family of absolutely continuous copulas will be denoted as

$$\mathfrak{C} = \{C_\theta \mid \theta \in \Theta\},$$

where $\Theta \subseteq \mathbb{R}^p$ for some $p \in \mathbb{N}$, i.e.

$$\theta = (\theta_1, \ldots, \theta_p)^\mathsf{T}.$$

Let us first recall two examples of such parametric families.

# Three Examples of Parametric Families

(i) **Bivariate Gumbel-Hougaard Copula**
This is a special case of an Archimedean Copula
(with generator $\psi(t) = \exp\left(-t^{1/\theta}\right)$):

$$\mathfrak{C} = \left\{ C_\theta(u_1, u_2) = \exp\left( -\left( (-\log u_1)^\theta + (-\log u_2)^\theta \right)^{1/\theta} \right) \mid \theta \geq 1 \right\}$$

(ii) **Clayton Copula**

This is a special case of an Archimedean Copula (with generator $\psi(t) = (1+t)^{-1/\theta}$):

$$\mathfrak{C} = \left\{ C_\theta(u_1, u_2) = \left(\max\{u_1^{-\theta} + u_2^{-\theta} - 1; 0\}\right)^{-1/\theta} \mid \theta > 0 \right\}$$

# Element of a parametric family

So, assuming $C \in \mathfrak{C}$, it follows that

$$\exists! \theta^* : C(F_1(\cdot), \ldots, F_d(\cdot)) = C_{\theta^*}(F_1(\cdot), \ldots, F_d(\cdot)).$$

**(New) Aim**

Estimate the parameter vector $\theta^*$, s.t. $H(x) = C_{\theta^*}(F_1(x_1), \ldots, F_d(x_d))$

**Additional Assumption:**

$$F_1 \in \mathfrak{F}_1 = \{F_{1,\lambda_1} \mid \lambda_1 \in \Lambda_1\}, \quad \Lambda_1 \subseteq \mathbb{R}^{p_1} \text{ and } p_1 \in \mathbb{N},$$
$$\vdots$$
$$F_d \in \mathfrak{F}_d = \{F_{d,\lambda_d} \mid \lambda_d \in \Lambda_d\}, \quad \Lambda_d \subseteq \mathbb{R}^{p_d} \text{ and } p_d \in \mathbb{N}.$$

Hence, it holds

$$\exists! \lambda_i^* \in \Lambda_i \ : \ F_i = F_{i,\lambda_i^*}(\cdot), \quad i \in \{1, \ldots, d\}$$

Therefore, the wanted Copula looks as follows

$$\boxed{C(F_1(\cdot), \ldots, F_d(\cdot)) = C_{\theta^*}\left(F_{1,\lambda_1^*}(\cdot), \ldots, F_{d,\lambda_d^*}(\cdot)\right)}$$

# Maximum Likelihood Estimator

Sklar's theorem says that

$$H(x_1, \dots, x_d) = C_\theta\left(F_{1,\lambda_1}(x_1), \dots, F_{d,\lambda_d}(x_d)\right)$$

and since $C_\theta, F_{1,\lambda_1}, \dots, F_{d,\lambda_d}$ are absolutely continuous, we compute the density

$$\begin{aligned} h(x_1, \dots, x_d) &= \frac{\partial^d}{\partial x_1 \cdots \partial x_d} H(x_1, \dots, x_d) \\ &= \frac{\partial^d}{\partial x_1 \cdots \partial x_d} C_\theta(F_{1,\lambda_1}(x_1), \dots F_{d,\lambda_d}(x_d)) \\ &= c_\theta(F_{1,\lambda_1}(x_1), \dots, F_{d,\lambda_d}(x_d)) \cdot \prod_{i=1}^{d} f_{i,\lambda_i}(x_i). \end{aligned}$$

# Maximum Likelihood Estimator

Likelihood-function:

$$\mathcal{L}(\lambda_1, \dots, \lambda_d, \theta) = \prod_{i=1}^{n} h(X_1^i, \dots, X_d^i)$$

Log-Likelihood-function:

$$\begin{aligned}
\log\left(\mathcal{L}(\lambda_1, \dots, \lambda_d, \theta)\right) &= \log\left(\prod_{i=1}^{n} h(X_1^i, \dots, X_d^i)\right) \\
&= \sum_{i=1}^{n} \log\left(c_\theta(F_{1,\lambda_1}(x_1), \dots, F_{d,\lambda_d}(x_d)) \cdot \prod_{i=1}^{d} f_i(x_i)\right) \\
&= \sum_{i=1}^{n} \log\left(c_\theta(F_{1,\lambda_1}(x_1), \dots, F_{d,\lambda_d}(x_d))\right) \\
&+ \sum_{i=1}^{n} \sum_{j=1}^{d} \log\left(f_i(x_j^i)\right)
\end{aligned}$$

# Maximum Likelihood Estimator

$$(\widehat{\lambda}_1, \ldots, \widehat{\lambda}_d, \widehat{\theta}) = \underset{\substack{\lambda_1 \in \Lambda_1, \\ \ldots, \\ \lambda_d \in \Lambda_d, \\ \theta \in \Theta}}{\arg\max} \log\left(\mathcal{L}(\lambda_1, \ldots, \lambda_d, \theta)\right)$$

**Main drawbacks**

(i)

(ii)

# Maximum Likelihood Estimator - Example

Lets assume that

- $d = 10$,
- $\mathfrak{C}$ Normal Copula Family
- $\mathfrak{F}_1 = \ldots = \mathfrak{F}_{10}$ family of $\Gamma(p, b)$–distributions.

What is the number of parameters ?

# Inference Functions for Margins Estimator

**Main idea:**

First, estimate the parameters $\lambda_1^*, \dots, \lambda_d^*$ as

$$\tilde{\lambda}_1 = \arg\max_{\lambda_{\mathbf{1}} \in \Lambda_{\mathbf{1}}} \sum_{i=1}^{n} \log(f_{1\lambda_{\mathbf{1}}}(X_1^i)),$$

$$\vdots$$

$$\tilde{\lambda}_d = \arg\max_{\lambda_d \in \Lambda_d} \sum_{i=1}^{n} \log(f_{d\lambda_d}(X_d^i))$$

Create *pseudo-observations*

$$U_{\tilde{\lambda}}^1 := \left(F_{1,\tilde{\lambda}_{\mathbf{1}}}(X_1^1), \dots, F_{d,\tilde{\lambda}_{\mathbf{1}}}(X_d^1)\right)$$

$$\vdots$$

$$U_{\tilde{\lambda}}^n := \left(F_{1,\tilde{\lambda}_{\mathbf{1}}}(X_1^n), \dots, F_{d,\tilde{\lambda}_{\mathbf{1}}}(X_d^n)\right)$$

# Inference Functions for Margins Estimator

Estimate the wanted parameter $\theta^*$ using a likelihood-like function

$$\begin{aligned}\mathcal{L}(\theta) &= \prod_{i=1}^{n} c_\theta(U^i_{\tilde{\lambda}}), \\ \log(\mathcal{L}(\theta)) &= \sum_{i=1}^{n} \log\left(c_\theta(U^i_{\tilde{\lambda}})\right) \\ \rightarrow \tilde{\theta} &= \end{aligned}$$

# Asymptotic Efficiency

Denote by

$$\begin{aligned} \eta^* &= (\lambda_1^*, \dots, \lambda_d^*, \theta^*)^\intercal && \text{the real parameters,} \\ \widehat{\eta} &= (\widehat{\lambda}_1, \dots, \widehat{\lambda}_d, \widehat{\theta})^\intercal && \text{all estimators of MLE,} \\ \widetilde{\eta} &= (\widetilde{\lambda}_1, \dots, \widetilde{\lambda}_d, \widetilde{\theta})^\intercal && \text{all estimators of IFME.} \end{aligned}$$

One can show that

$$\sqrt{n}(\widehat{\eta} - \eta^*) \xrightarrow[n\to\infty]{d} \mathcal{N}(0, \mathcal{I}^{-1}),$$

$$\sqrt{n}(\widetilde{\eta} - \eta^*) \xrightarrow[n\to\infty]{d} \mathcal{N}(0, V).$$

Numerically, one can show that in many cases

$$V - \mathcal{I}^{-1} \approx 0.$$

# One drawback is left

IFM Eis computationally very efficient.
But what about the second drawback?

# Non-Parametrically Estimated Margins

Drop the assumption:

~~Margins belong to parametric families~~

**Instead:** (Rescaled) Empirical distributions functions

# Pseudo-Observations

Plugging in yields the following *pseudo-observations*

$$\widehat{U}_1 = \left(\widehat{F}_1(X_1^1), \dots, \widehat{F}_d(X_d^1)\right)$$
$$\widehat{U}_2 = \left(\widehat{F}_1(X_1^2), \dots, \widehat{F}_d(X_d^2)\right)$$
$$\vdots$$
$$\widehat{U}_n = \left(\widehat{F}_1(X_1^n), \dots, \widehat{F}_d(X_d^n)\right)$$

**Note:**

- Not true observations, just estimators.
- Not independent.
- Use them as observations to estimate $C$.

# Method of Moments

We discuss the bivariate case and one-parameter copulas only.
Let $(X, Y)$ be a vector, and $(x_1, y_1), \ldots, (x_n, y_n)$ be a random sample of observationbs of $(X, Y)$.
How many distinct pairs exist?

$$\text{Concordant pair: } (x_i, y_i), (x_j, y_j) \text{ s.t. } \operatorname{sgn}(x_i - x_j) = \operatorname{sgn}(y_i - y_j)$$
$$\text{Discordant pair: } (x_i, y_i), (x_j, y_j) \text{ s.t. } \operatorname{sgn}(x_i - x_j) = -\operatorname{sgn}(y_i - y_j)$$

**Sample version of Kendall's tau:**

$$\tau_n := \frac{c - d}{\binom{n}{2}} = P(\text{Concordant pair}) - P(\text{Discordant pair})$$

**Population version of Kendall's tau:**

$$\tau_{X,Y} := P((X_2 - X_1)(Y_2 - Y_1) > 0) - P((X_2 - X_1)(Y_2 - Y_1) < 0)$$

# Kendall's Tau Representation

**Theorem**

*Let $(X_1, Y_1), (X_2, Y_2)$ be independent vectors. All random variables are continuous with joint distribution function $H_1$ and $H_2$, and margins $F$ (of $X_1, X_2$) and $G$ (of $Y_1, Y_2$). Let $C_1$ and $C_2$ the copulas associated with $H_1$ and $H_2$.*
*Then,*

$$\begin{aligned}\mathcal{Q} &= \mathcal{Q}(C_1, C_2) \\ &:= P((X_2 - X_1)(Y_2 - Y_1) > 0) - P((X_2 - X_1)(Y_2 - Y_1) < 0) \\ &= 4 \int_{[0,1]} \int_{[0,1]} C_2(u, v) dC_1(u, v) - 1\end{aligned}$$

# Proof

# Proof

# Proof

# Method-of-Moments Estimator

Next, we define

$$
\begin{aligned}
g_\tau(\theta) &:= \tau(C_\theta) \\
\rightarrow \widehat{\theta} &= g_\tau^{-1}(\tau_n)
\end{aligned}
$$

**Example:** Farlie-Gumberl-Morgenstern

# References i


Marius Hofert, Ivan Kojadinovic, Martin Machler, Jun Yan
**Elements of Copula Modeling with R.**
Springer, 2018

Roger B. Nelsen
**An introduction to copulas.**
Second Edition, Springer, 2006.

E.L. Lehmann, George Casella
**Theory of Point Estimation.**
Springer texts in statistics, Edition 2, 2003

- Ivan Kojadinovic and Jun Yan
  **Comparison of three semiparametric methods for estimating dependence parameters in copula models.**
  Insurance: Mathematics and Economics, Volume = 47, PP 52 - 63, 2010
- Harry Joe and James Jianmeng Xu
  **The Estimation Method of Inference Functions for Margins for Multivariate Models.**
  Technical Report, Vancouver : University of British Columbia Library, 1996 PP 401 - 419, 2005
- Harry Joe
  **Asymptotic efficiency of the two-stage estimation method for copula-based models.**
  Journal of Multivariate Analysis, Volume 94, PP 401 - 419, 2005

Thank you for your attention!

Any questions?