Vector Copulas and Vector Sklar Theorem

Yanqin Fan and Marc Henry
Department of Economics
University of Washington and Penn State University

January 30, 2021
PIHOT Kickoff Event

A Review of Copulas and Sklar's Theorem

- **Definition**: *A copula is a multivariate distribution function with uniform marginals on* $[0, 1]$.

- Let $(Y_1, Y_2) \sim F$, a bivariate cdf with *continuous* marginals F_1, F_2.

- **Sklar's Theorem (i)**: Given F, there exists a unique copula $C : [0, 1]^2 \rightarrow [0, 1]$ such that

$$F(y_1, y_2) = C(F_1(y_1), F_2(y_2)), \text{ where}$$

$$C(u_1, u_2) = F\left(F_1^{-1}(u_1), F_2^{-1}(u_2)\right). \tag{1}$$

- The copula function in (1) is called the copula function of F or of (Y_1, Y_2):

$$C(u_1, u_2) = \Pr(F_1(Y_1) \leq u_1, F_2(Y_2) \leq u_2).$$

- Since $F_1(Y_1) \sim U[0,1]$ and $F_2(Y_2) \sim U[0,1]$, C characterizes *rank dependence* between Y_1 and Y_2.

- **Example**: Let Φ_ρ denote the standard bivariate normal cdf with correlation coefficient $\rho \in (0,1)$. Then (1) yields the Gaussian copula:

$$C^{\text{Gaussian}}(u_1, u_2; \rho) = \Phi_\rho\left(\Phi^{-1}(u_1), \Phi^{-1}(u_2)\right).$$

- **Sklar's Theorem (ii)**: For any copula C and any marginal cdfs F_1, F_2, $C(F_1(y_1), F_2(y_2))$ is a bivariate distribution function with marginals F_1, F_2 and copula C.

  – Let $(U_1, U_2) \sim C$. Then

  $$C(F_1(y_1), F_2(y_2)) = \mathsf{Pr}\left(F_1^{-1}(U_1) \leq y_1, F_2^{-1}(U_2) \leq y_2\right), \quad (2)$$

  where $F_1^{-1}(U_1) \sim F_1$ and $F_2^{-1}(U_2) \sim F_2$.

- **Example**: Let $F_1(y_1)$ be lognormal and $F_2(y_2)$ be χ^2_ν. Then

  $$C^{\text{Gaussian}}(F_1(y_1), F_2(y_2); \rho) = \Phi_\rho\left(\Phi^{-1}(F_1(y_1)), \Phi^{-1}(F_2(y_2))\right)$$

  is a bivariate cdf with lognormal and χ^2_ν marginals respectively and the Gaussian copula with parameter ρ.

- Let $\{C(u_1, u_2; \theta) : \theta \in \Theta\}$ denote a parametric family of copulas. Then $\{C(F_1(y_1), F_2(y_2); \theta) : \theta \in \Theta\}$ is a semiparametric family of bivariate cdfs with density function

$$f_1(y_1)f_2(y_2)c(F_1(y_1), F_2(y_2); \theta),$$

  where c is the copula density function and $f_1(y_1), f_2(y_2)$ are pdfs of $F_1(y_1), F_2(y_2)$.

- Copulas provide a flexible approach to constructing semiparametric multivariate distributions
  - there exist rich classes of parametric copulas (Gaussian, Archimedean,...).

- Suppose a random sample $\{Y_{1i}, Y_{2i}\}_{i=1}^{n}$ is drawn from the pdf above for some $\theta_0 \in \Theta$.

- Estimation and inference can be done via either full MLE or two-step MLE (e.g. Genest and Rivest 2003; Chen and Fan, 2006a,b; Chen, Fan, and Tsyrennikov, 2006; Joe 1997, 2015)

$$\ln \mathcal{L}\left(f_1, f_2, \theta\right) = \sum_{i=1}^{n} \left[\ln f_1(Y_{1i}) + \ln f_2(Y_{2i})\right] + \sum_{i=1}^{n} \ln c(F_1(Y_{1i}), F_2(Y_{2i}); \theta)$$

  - needs to have estimators of the marginal cdfs and/or pdfs (empirical cdf, kernel, sieve estimates,...)

- Many empirical applications in Economics and Finance, see Fan and Patton (2014) for a review.

Vector Copulas and Vector Sklar Theorem

- Consider two *random vectors* Y_1 and Y_2 such that $(Y_1, Y_2) \sim P\ (F)$, where $P\ (F)$ denotes a probabilty measure (cdf) on $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ with absolutely continuous marginals $P_k\ (F_k)$ on $\mathbb{R}^{d_k}$ with support contained in a convex set $\mathcal{Y}_k$ for $k = 1, 2$.

- **Questions**: how to
  - characterize rank dependence between random vectors Y_1 and Y_2;
  - construct multivariate distributions with given multivariate marginals and rank dependence.

- Some existing proposals

  - copula impossibility result (e.g. Genest, 1995): the only copula C such that $C(F_1(y_1), F_2(y_2))$ defines a $(d_1 + d_2)$-dimensional cdf with d_1-dimensional marginal F_1 and d_2-dimensional F_2 for all d_1 and d_2 such that $d_1 + d_2 \geq 3$, and for all F_1 and F_2, is $C(u_1, u_2) = u_1 u_2$.

  - linkage function: Li et al (1996) makes use of Knothe-Rosenblatt transform of F_k to define *a linkage function* analogously to a copula function. Unlike copulas, no known flexible parametric families of linkage functions are available.

- **This talk introduces** vector copulas and vector Sklar Theorem

- **Definition**: A *vector copula* C is defined as a joint distribution function on $[0,1]^d$ with uniform marginals μ_k on $\mathcal{U}_k \equiv [0,1]^{d_k}$, $k = 1, 2$, where $d = d_1 + d_2$.

- **How to extract the vector copula from P (F)?**

- When $d_1 = d_2 = 1$, we rely on *ranks/quantiles*

- For $d_k > 1$, we rely on (generalized) vector ranks/vector quantiles: **Brenier maps between P_k and μ_k**, see Chernozhukov et al. (2017)

- **Brenier-McCann Theorem:**
  - there exists a convex function $\psi_k : \mathcal{U}_k \rightarrow \mathbb{R} \cup \{+\infty\}$ such that $\nabla\psi_k\#\mu_k = P_k$. The function $\nabla\psi_k$ exists and is unique, μ_k-almost everywhere. $\nabla\psi_k$ is called the *vector quantile* of P_k.
  - there exists a convex function $\psi_k^* : \mathcal{Y}_k \rightarrow \mathbb{R} \cup \{+\infty\}$ such that $\nabla\psi_k^*\#P_k = \mu_k$. The function $\nabla\psi_k^*$ exists, is unique and equal to $\nabla\psi_k^{-1}$, P_k-almost everywhere. $\nabla\psi_k^*$ is called the *vector rank* of P_k.

- **Generalized Vector Quantiles and Ranks**

- Let $\psi_{k,l}$, $l \le L$ for some finite integer L, be convex functions such that the following hold.

  – the map $T_k := \nabla\psi_{k,L} \circ \nabla\psi_{k,L-1} \circ \ldots \circ \nabla\psi_{k,1}$ exists and satisfies $T_k \# \mu_k = P_k$. The map T_k is called *generalized vector quantile* associated with P_k.

  – the map $T_k^- := \nabla\psi^*_{k,1} \circ \nabla\psi^*_{k,2} \circ \ldots \circ \nabla\psi^*_{k,L}$ exists and satisfies $T_k^- \# P_k = \mu_k$. The map T_k^- is called *generalized vector rank* associated with P_k.

- By choosing L and $\psi_{k,l}$, $l \le L$, we construct generalized vector quantile and rank with closed-form expressions.

- **Example.** Let $Y_k \sim \Phi_{d_k}(\cdot; \Sigma_k)$, where $\Sigma_k > 0$. The generalized Gaussian vector rank is

$$T_k^- = \nabla\psi_{1k}^* \circ \nabla\psi_{2k}^*,$$

where

  - $\nabla\psi_{1k}^*(u_k) = \Phi(u_k)$, $u_k \in (0,1)^{d_k}$, is the OT map between $\Phi_{d_k}\left(\cdot; I_{d_k}\right)$ and μ_k;

  - $\nabla\psi_{2k}^* \equiv \Sigma_k^{-1/2}$ is the OT map between $\Phi_{d_k}(\cdot; \Sigma_k)$ and $\Phi_{d_k}\left(\cdot; I_{d_k}\right)$.

- **Three versions of Sklar's Theorem**

- For $d_1 = d_2 = 1$, the Sklar's theorem states that

$$F(y_1, y_2) = C(F_1(y_1), F_2(y_2)). \quad (3)$$

- The above expression is equivalent to

$$\begin{aligned} f(y_1, y_2) &= f_1(y_1)f_2(y_2)c(F_1(y_1), F_2(y_2)) \text{ or} \quad (4) \\ P_F(A_1 \times A_2) &= P_C(F_1(A_1) \times F_2(A_2)) \quad (5) \end{aligned}$$

  for any collection (A_1, A_2), where A_k is a Borel subset of $\mathcal{Y}_k$. Here P_C is the probability measure induced by C.

  – For any $A_k = (-\infty, y_k]$, $F_k((-\infty, y_k]) = (0, F_k(y_k)]$.

- **Vector Sklar Theorem (i)** Given P, there exists a unique vector copula C such that for any collection (A_1, A_2), where A_k is a Borel subset of $\mathcal{Y}_k$,

$$P(A_1 \times A_2) = P_C\left(T_1^{-}(A_1) \times T_2^{-}(A_2)\right), \tag{6}$$

  and for all Borel sets B_1, B_2 in $\mathcal{U}_1, \mathcal{U}_2$,

$$P_C(B_1 \times B_2) = P(T_1(B_1) \times T_2(B_2)). \tag{7}$$

  - The vector copula of P is the joint distribution of $\left(T_1^{-}(Y_1), T_2^{-}(Y_2)\right)$ for $(Y_1, Y_2) \sim P$.

  - Since $T_k^{-} \# P_k = \mu_k$, the vector copula of P measures the rank dependence between Y_1 and Y_2.

- **Vector Sklar Theorem (ii)** For any vector copula C and any distributions P_k on $\mathbb{R}^{d_k}$ with (generalized) vector quantiles T_k, (6) defines a distribution on $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ with marginals P_k and vector copula C.

- The Vector Sklar theorem extends Sklar's theorem (5).

- A direct extension of Sklar's theorem (3) would be

$$F(y_1, y_2) = C\left(T_1^-(y_1), T_2^-(y_2)\right)?$$

  - Let $A_k = (-\infty, y_k]$, $y_k \in \mathbb{R}^{d_k}$. (6) implies that

$$F(y_1, y_2) = C\left(T_1^-(A_1), T_2^-(A_2)\right)$$

    but in general $T_k^-(A_k) \neq (0, T_k^-(y_k)]$ when $d_k > 1$.

- Thanks to the *Monge Ampère Equation*,

$$\det\left(DT_k\left(u_k\right)\right) = \frac{1}{f_k\left(T_k\left(u_k\right)\right)},$$

we obtain the following extension of Sklar's theorem (4):

$$f\left(y_1, y_2\right) = \left[\prod_{k=1}^{2} f_k\left(y_k\right)\right] c\left(T_1^{-}(y_1), T_2^{-}(y_2)\right). \tag{8}$$

- (8) offers a unified approach to constructing and estimating semiparametric multivariate distributions with prespecified multivariate marginals and parametric vector copulas
  - need parametric families of vector copulas (Gaussian, Archimedean,...) but more are needed!

- **Gaussian Vector Copula.** Let $(Y_1, Y_2) \sim \Phi_d(\cdot; \Sigma)$, where $d = d_1 + d_2$ and $\Sigma = \begin{pmatrix} \Sigma_1 & \Sigma_{12} \\ \Sigma_{21} & \Sigma_2 \end{pmatrix}$, $\Sigma_k > 0$. The Gaussian vector copula is

$$C^{Ga}(u_1, u_2; \Omega) = \Phi_d(\nabla\psi_{11}(u_1), \nabla\psi_{12}(u_2); \Omega), \tag{9}$$

where

$$\nabla\psi_{1k}(u_k) = \Phi^{-1}(u_k) = \left(\Phi^{-1}(u_{k1}), ..., \Phi^{-1}(u_{kd_k})\right) \text{ and}$$

$$\Omega = \begin{pmatrix} I_{d_1} & \Sigma_1^{-1/2}\Sigma_{12}\Sigma_2^{-1/2} \\ \Sigma_2^{-1/2}\Sigma_{21}\Sigma_1^{-1/2} & I_{d_2} \end{pmatrix} \tag{10}$$

- When $d_1 = d_2 = 1$, $C^{Ga}(u_1, u_2; \Omega) = C^{\text{Gaussian}}(u_1, u_2; \rho)$, where $\rho = \Sigma_{12}/(\Sigma_1\Sigma_2)^{1/2}$.

- **Proof:** The vector copula is the distribution function of $\left(T_1^-(Y_1), T_2^-(Y_2)\right)$, where $T_k^- := \nabla\psi_{1k}^* \circ \nabla\psi_{2k}^*$,

$$\nabla\psi_{1k}^*(u_k) = \Phi(u_k) \text{ and } \nabla\psi_{2k}^* \equiv \Sigma_k^{-1/2}.$$

Since $\left(\Sigma_1^{-1/2}Y_1, \Sigma_2^{-1/2}Y_2\right) \sim \Phi_d(\cdot;\Omega)$, we obtain that

$$\begin{aligned}
& C^{Ga}(u_1, u_2; \Omega) \\
= \; & \Pr\left(T_1^-(Y_1) \le u_1, T_2^-(Y_2) \le u_2\right) \\
= \; & \Pr\left(\nabla\psi_{11}^*\left(\Sigma_1^{-1/2}Y_1\right) \le u_1, \nabla\psi_{12}^*\left(\Sigma_2^{-1/2}Y_2\right) \le u_2\right) \\
\le \; & \Pr\left(\Sigma_1^{-1/2}Y_1 \le \nabla\psi_{11}(u_1), \Sigma_2^{-1/2}Y_2 \le \nabla\psi_{12}(u_2)\right) \\
= \; & \Phi_d\left(\nabla\psi_{11}(u_1), \nabla\psi_{12}(u_2); \Omega\right).
\end{aligned}$$

Current Research

- Suppose a random sample $\{Y_{1i}, Y_{2i}\}_{i=1}^{n}$ is drawn from the pdf below for some $\theta_0 \in \Theta$:

$$f(y_1, y_2) = \left[\prod_{k=1}^{2} f_k(y_k)\right] c\left(T_1^{-}(y_1), T_2^{-}(y_2); \theta_0\right),$$

where T_k^{-} is the vector rank of F_k for $k = 1, 2$.

- A two-step estimator of θ_0 is given by

$$\widehat{\theta} = \arg\max_{\theta \in \Theta} \left[\frac{1}{n}\sum_{i=1}^{n} \ln c\left(\widehat{T}_1^{-}(Y_{1i}), \widehat{T}_2^{-}(Y_{2i}); \theta\right)\right],$$

where $\widehat{T}_k^{-}$ is a nonparametric estimator of T_k^{-}.

– many candidates for $\widehat{T}_k^-$ are available in the OT literature,

– significant progress on computation has been made recently, but

– asymptotic theory for $\widehat{T}_k^-$ is less developed (Flamary et al 2019, Hutter and Rigollet 2019, Harchaoui, Liu, and Pal (2020),...)

- Under regularity conditions,

$$\begin{aligned}
\sqrt{n}\left(\widehat{\theta}-\theta_0\right) &\approx \left[\frac{1}{n}\sum_{i=1}^{n} D_\theta^2 \ln c\left(\widehat{T}_1^{-}(Y_{1i}), \widehat{T}_2^{-}(Y_{2i}); \theta_0\right)\right]^{-1} \\
&\quad \times \left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n} D_\theta \ln c\left(\widehat{T}_1^{-}(Y_{1i}), \widehat{T}_2^{-}(Y_{2i}); \theta_0\right)\right] \\
&\approx \left[E\left\{D_\theta^2 \ln c\left(T_1^{-}(Y_{1i}), T_2^{-}(Y_{2i}); \theta_0\right)\right\}\right]^{-1} \\
&\quad \times \left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n} D_\theta \ln c\left(\widehat{T}_1^{-}(Y_{1i}), \widehat{T}_2^{-}(Y_{2i}); \theta_0\right)\right] \\
&\Longrightarrow \left[E\left\{D_\theta^2 \ln c\left(T_1^{-}(Y_{1i}), T_2^{-}(Y_{2i}); \theta_0\right)\right\}\right]^{-1} N\left(0, ???\right)???
\end{aligned}$$