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 $C : [0, 1]^2 \to [0, 1]$ such that

\[
F (y_1, y_2) = C (F_1(y_1), F_2(y_2)), \quad \text{where}
\]

\[
C (u_1, u_2) = F \left( F_1^{-1} (u_1), F_2^{-1} (u_2) \right). \tag{1}
\]
The copula function in (1) is called the copula function of $F$ or of $(Y_1, Y_2)$:

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

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 $\Phi_{\rho}$ 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 $F_1, F_2$, $C\left(F_1(y_1), F_2(y_2)\right)$ is a bivariate distribution function with marginals $F_1, F_2$ and copula $C$.

- Let $(U_1, U_2) \sim C$. Then

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

  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 $\chi^2_v$. Then

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

  is a bivariate cdf with lognormal and $\chi^2_v$ marginals respectively and the Gaussian copula with parameter $\rho$. 
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)

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

- 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 probability 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 $C(u_1, u_2) = u_1u_2$.

  – 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 $\mu_k$ 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 $\mu_k$, see Chernozhukov et al. (2017)
• Brenier-McCann Theorem:

- there exists a convex function $\psi_k : \mathcal{U}_k \to \mathbb{R} \cup \{+\infty\}$ such that $\nabla \psi_k \# \mu_k = P_k$. The function $\nabla \psi_k$ exists and is unique, $\mu_k$-almost everywhere. $\nabla \psi_k$ is called the vector quantile of $P_k$.

- there exists a convex function $\psi_k^* : \mathcal{Y}_k \to \mathbb{R} \cup \{+\infty\}$ 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 $\psi_{k,l}$, $l \leq L$ 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 \textit{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 \textit{generalized vector rank} associated with $P_k$.

• By choosing $L$ and $\psi_{k,l}$, $l \leq L$, 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

$$T_k^- = \nabla \psi^*_1 k \circ \nabla \psi^*_2 k,$$

where

- $\nabla \psi^*_1 k (u_k) = \Phi(u_k), u_k \in (0, 1)^{d_k}$, is the OT map between $\Phi_{d_k} (\cdot; I_{d_k})$ and $\mu_k$;

- $\nabla \psi^*_2 k \equiv \Sigma_k^{-1/2}$ is the OT map between $\Phi_{d_k} (\cdot; \Sigma_k)$ and $\Phi_{d_k} (\cdot; I_{d_k})$. 
Three versions of Sklar’s Theorem

- For \( d_1 = d_2 = 1 \), the Sklar’s theorem states that

\[
F(y_1, y_2) = C(F_1(y_1), F_2(y_2)).
\]  

(3)

- The above expression is equivalent to

\[
f(y_1, y_2) = f_1(y_1)f_2(y_2)c(F_1(y_1), F_2(y_2)) \text{ or }  
P_F(A_1 \times A_2) = P_C(F_1(A_1) \times F_2(A_2))
\]  

(4)

(5)

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 $\mathcal{Y}_k$,

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

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

$$P_C(B_1 \times B_2) = P\left(T_1(B_1) \times T_2(B_2)\right). \quad (7)$$

– The vector copula of $P$ is the joint distribution of $(T_1^{-}(Y_1), T_2^{-}(Y_2))$ 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 \( T_k^-\left(A_k\right) \neq (0, T_k^-\left(y_k\right)] \) 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):

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

(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 
\[
\Sigma = \begin{pmatrix}
\Sigma_1 & \Sigma_{12} \\
\Sigma_{21} & \Sigma_2
\end{pmatrix}, \quad \Sigma_k > 0.
\]
The Gaussian vector copula is
\[
C^{Ga} (u_1, u_2; \Omega) = \Phi_d (\nabla \psi_{11} (u_1), \nabla \psi_{12} (u_2); \Omega),
\]
where
\[
\begin{align*}
\nabla \psi_{1k} (u_k) &= \Phi^{-1} (u_k) = (\Phi^{-1} (u_{k1}), \ldots, \Phi^{-1} (u_{kd_k})) \\
\Omega &= \begin{pmatrix}
I_{d_1} & \Sigma_1^{-1/2} \Sigma_{12} \Sigma_2^{-1/2} \\
\Sigma_2^{-1/2} \Sigma_{21} \Sigma_1^{-1/2} & I_{d_2}
\end{pmatrix}
\end{align*}
\]

• When \(d_1 = d_2 = 1\), 
\[
C^{Ga} (u_1, u_2; \Omega) = C^{Gaussian} (u_1, u_2; \rho),
\]
where \(\rho = \Sigma_{12} / (\Sigma_1 \Sigma_2)^{1/2}\).
Proof: The vector copula is the distribution function of \( (T_1^- (Y_1), T_2^- (Y_2)) \),
where \( T_k^- := \nabla \psi_{1k}^* \circ \nabla \psi_{2k}^* \),

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

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

\[
C^\text{Ga} (u_1, u_2; \Omega) = \Pr \left( T_1^- (Y_1) \leq u_1, T_2^- (Y_2) \leq u_2 \right) = \Pr \left( \nabla \psi_{11}^* \left( \Sigma_1^{-1/2} Y_1 \right) \leq u_1, \nabla \psi_{12}^* \left( \Sigma_2^{-1/2} Y_2 \right) \leq u_2 \right) \leq \Pr \left( \Sigma_1^{-1/2} Y_1 \leq \nabla \psi_{11} (u_1), \Sigma_2^{-1/2} Y_2 \leq \nabla \psi_{12} (u_2) \right) = \Phi_d (\nabla \psi_{11} (u_1), \nabla \psi_{12} (u_2); \Omega) .
\]
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 \( \theta_0 \) is given by

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

where \( \hat{T}_k^- \) is a nonparametric estimator of \( T_k^- \).
- many candidates for $\hat{T}_k^{-}$ are available in the OT literature,

- significant progress on computation has been made recently, but

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

\[
\sqrt{n} \left( \hat{\theta} - \theta_0 \right) \approx \left[ \frac{1}{n} \sum_{i=1}^{n} D_{\theta}^2 \ln c \left( \hat{T}_1^-(Y_{1i}), \hat{T}_2^-(Y_{2i}); \theta_0 \right) \right]^{-1} \\
\times \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} D_{\theta} \ln c \left( \hat{T}_1^-(Y_{1i}), \hat{T}_2^-(Y_{2i}); \theta_0 \right) \right] \\
\approx \left[ E \left\{ D_{\theta}^2 \ln c \left( T_1^-(Y_{1i}), T_2^-(Y_{2i}); \theta_0 \right) \right\} \right]^{-1} \\
\times \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} D_{\theta} \ln c \left( \hat{T}_1^-(Y_{1i}), \hat{T}_2^-(Y_{2i}); \theta_0 \right) \right] \\
\implies \left[ E \left\{ D_{\theta}^2 \ln c \left( T_1^-(Y_{1i}), T_2^-(Y_{2i}); \theta_0 \right) \right\} \right]^{-1} N(0, ???) ???
\]