Modelling and estimation of multivariate densities in a copula-based model

Eckhard Liebscher

University of Applied Sciences Merseburg

Germany

contents:

1. Construction of multivariate copulas
2. Parametric estimators for copulas
3. Semiparametric density estimators

Motivation

- recent trends: large datasets, highdimensional data
- nonparametric density estimation $\longrightarrow$ curse of dimensionality
- applications from multivariate statistics
- lack of asymmetric multivariate copula models

Copulas

X, Y real random variables
H, h distribution function and density of (X, Y).
F, G marginal distribution functions of X, Y

- Sklar's Theorem $\Longrightarrow$ $\boxed{H(x,y) = C(F(x), G(y))}$

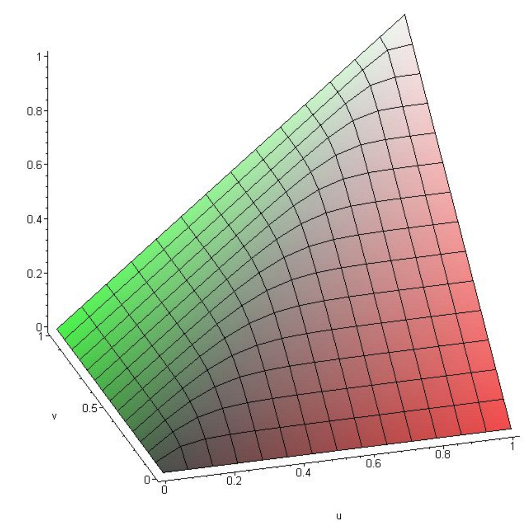
C Copula von (X, Y)

C is a distribution function on $[0,1] \times [0,1]$,

marginal distributions of C are uniform on $[0,1]$

Joe 1997, Nelsen 1999

figure of a copula



Multivariate Copulas

H joint distribution function, h density of $(X^{(1)}, \dots, X^{(d)})$,
F_m, f_m marginal distribution function/density of $X^{(m)}$

- Sklar's Theorem $\Longrightarrow$

$$H(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d))$$

C copula $\rightarrow$ continuous distribution function on $[0,1]^d$

$$C(u_1, \dots, 0, \dots, u_d) = 0, \qquad C(1, \dots, v, \dots, 1) = v$$

- copula density:

$$\varphi(u_1, \dots, u_d) = \frac{\partial^d}{\partial u_1 \dots \partial u_d} C(u_1, \dots, u_d)$$

Product of copulas

Theorem: Assume that $C_1, C_2 : [0,1]^d \to [0,1]$ are continuous copulas, $g_{j1}, \ldots, g_{jd} : [0,1] \to [0,1]$ $(j = 1, 2)$ be bijective, strictly increasing functions or identically equal 1 such that

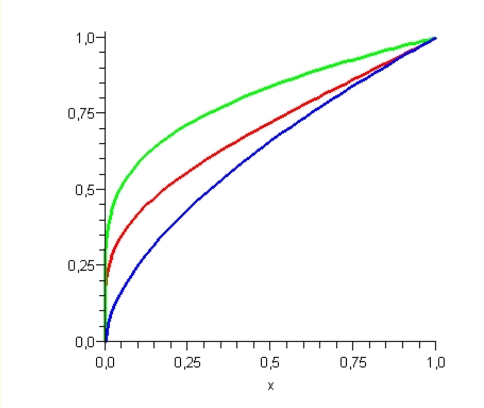
$$g_{1i}(v) \cdot g_{2i}(v) = v \text{ for } v \in [0,1].$$

$\Longrightarrow$

$$\boxed{\bar{C}(u_1, \ldots, u_d) = C_1(g_{11}(u_1), \ldots, g_{1d}(u_d)) \cdot C_2(g_{21}(u_1), \ldots, g_{2d}(u_d))}$$

for $u_i \in [0,1]$. special case: Khoudraji 1995

- Extension to more than two factors



- Examples for g_{ji}:

$g_{1i}(v) = v^\theta,\ g_{2i}(v) = v^{1-\theta},\ \theta \in (0,1)$

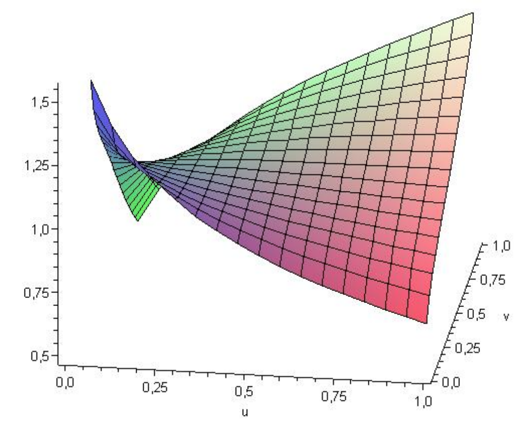
$g_{1i}(v) = \exp(\gamma - \sqrt{|\ln v| + \gamma^2}),\ g_{2i}(v) = v \exp(-\gamma + \sqrt{|\ln v| + \gamma^2})$

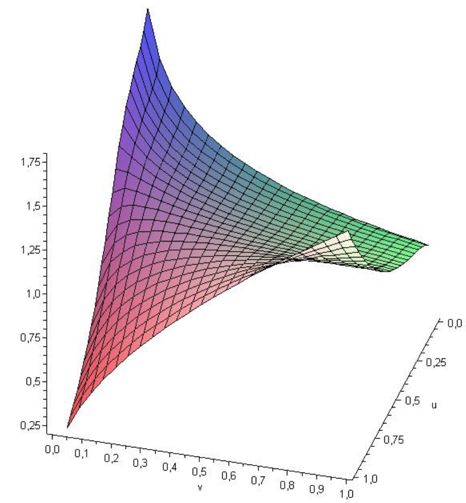
Product of Clayton (Cook-Johnson) copulas

$$\bar{C}(u_1,\ldots,u_d) = C(g_{11}(u_1),\ldots,g_{1d}(u_d) \mid \gamma)\cdot C(g_{21}(u_1),\ldots,g_{2d}(u_d) \mid \delta),$$

$$C(u_1,\ldots,u_d|\gamma) = \left(1+\sum_{i=1}^{d}(u_i^{-\gamma}-1)\right)^{-1/\gamma}$$

Copula densities





- Kendall's tau for several cases

γ	δ	θ_1	θ_2	Kendall's tau
0.6	6	0.0	0.0	0.75
		0.0	0.2	0.6
		0.4	0.4	0.379485
		0.4	0.9	0.154278
2	25	0.0	0.0	0.925965
		0.2	0.2	0.689878
		0.4	0.4	0.579587
		0.2	0.9	0.190630

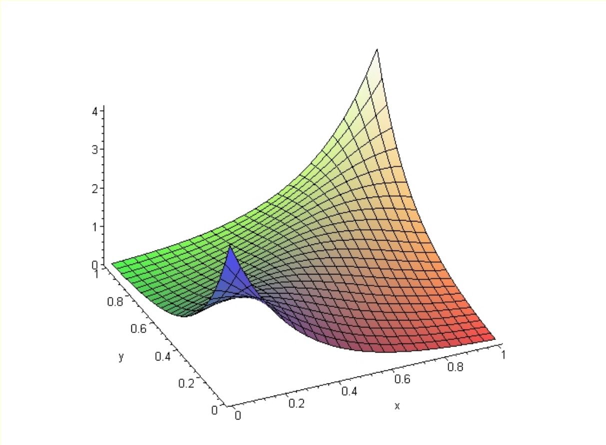
Generalised Archimedean copulas

- **Archimedean copula:** $\psi : (0,1] \to [0,+\infty)$ strictly decreasing with $\psi(0)=\infty, \psi(1)=0$,

$$C^{(d)}(u_1,\ldots,u_d) = \psi^{-1}(\psi(u_1)+\ldots+\psi(u_d))$$

- Frank copula

copula density



$$\psi(t) = \ln\left(\frac{1-e^{-\gamma}}{1-e^{-\gamma t}}\right)$$

parameter: $\gamma > 0$

multiplicative generator $\bar{\psi}(u) = \exp(-\psi(u)) \Longrightarrow$

$$C^{(d)}(u_1,\ldots,u_d) = \bar{\psi}^{-1}(\bar{\psi}(u_1)\cdot\ldots\cdot\bar{\psi}(u_d))$$

Replace the product $\bar{\psi}(u_1)\cdot\ldots\cdot\bar{\psi}(u_d)$ by a sum of such products$\Rightarrow$

Generalised Archimedean copulas

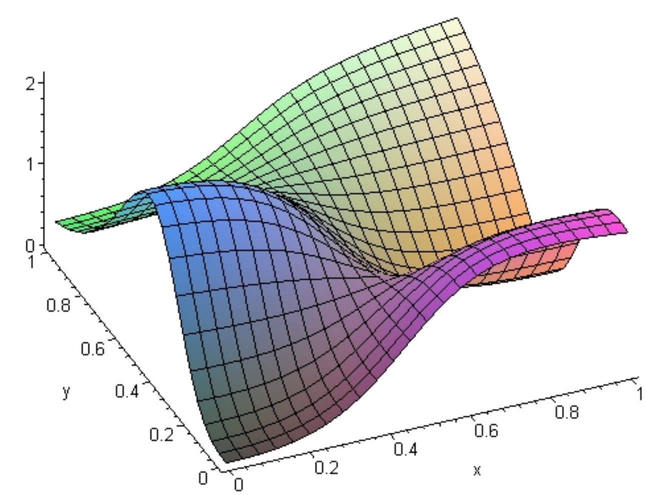
$$C(u) = \Psi\left(\frac{1}{m}\sum_{j=1}^{m} h_{j1}(\psi(u_1)) \cdot \ldots \cdot h_{jd}(\psi(u_d))\right), \qquad \Psi = \psi^{-1}$$

$\Psi : [0,1] \to [0,1]$ increasing function, $\Psi(0) = 0$, $\Psi(1) = 1$

Theorem: Assume that $\Psi^{(d)}$ exist, $\Psi^{(k)}(u) \geq 0$ for $k = 1, \ldots, d$, $u \in [0,1]$, $h_{jk} : [0,1] \to [0,1]$ is differentiable with $h'_{jk}(u) > 0$, $h_{jk}(0) = 0$, $h_{jk}(1) = 1$, and $\frac{1}{m}\sum_{j=1}^{m} h_{jk}(v) = v$ for $k = 1, \ldots, d, v \in [0,1]$.

$\Longrightarrow C$ is a copula.

copula density - general. Frank family



- Example:

$$h_{1k}(x) = x^{\delta_k},$$
$$h_{2k}(x) = 2x - x^{\delta_k}$$

Advantages of product/generalised Archimedean copulas

- asymmetric

  symmetry means $C(u_1, \ldots, u_d) = C(\pi(u_1, \ldots, u_d))$ for all $u_i \in [0, 1]$ and all permutations π

- flexible number of parameters

- cover a wide range of dependencies

- product copulas: simple procedure for the generation of random vectors, positive quadrant dependence

Multivariate densities and copulas

h joint density of $(X^{(1)}, \ldots, X^{(d)})$,
F_m, f_m marginal distribution function/density of $X^{(m)}$,

- **multivariate density**

$$h(x_1, \ldots, x_d) = \varphi(F_1(x_1), \ldots, F_d(x_d)) f_1(x_1) \ldots f_d(x_d)$$

copula density $\varphi(u_1, \ldots, u_d) = \dfrac{\partial^d}{\partial u_1 \ldots \partial u_d} C(u_1, \ldots, u_d)$

▷ **Idea: Parametrisation**

$$\varphi(u_1, \ldots, u_d) = \varphi(u_1, \ldots, u_d \mid \theta) \text{ for } u_i \in [0, 1]$$

parameter $\theta \in \Theta$

Maximum likelihood estimators

$\{\varphi(. \mid \theta)\}_{\theta\in\Theta}$ family of copula densities, $\Theta \subset \mathbb{R}^q$ parameter space
$X_1, \ldots, X_n$ sample of i.i.d. random vectors having copula density φ

- complexity of the distribution $\Longrightarrow$ often $\varphi \notin \{\varphi(. \mid \theta)\}_{\theta\in\Theta}$.

Let $\varphi(. \mid \theta_0)$ be the best approximation for φ in the sense of Kullback-Leibler divergence:

$$\theta_0 = \underset{\theta\in\Theta}{\operatorname{argmax}} \int_{[0,1]^d} \ln\big(\varphi(t \mid \theta)\big)\, \varphi(t) dt$$

case $\varphi \in \{\varphi(. \mid \theta)\}_{\theta\in\Theta}$: θ_0 is the true parameter

Aim: estimation of parameter θ_0

Maximum likelihood estimators

- $X_1, \ldots, X_n$ sample of i.i.d. random vectors having copula density φ, F_{jn} empirical marginal distribution function

$Y_{ni} = (F_{1n}(X_i^{(1)}), .., F_{dn}(X_i^{(d)}))^T$ transformed sample item

- **Canonical maximum likelihood estimator:**

$$\hat{\theta}_n \in \underset{\theta \in \Theta}{\operatorname{argmax}} \sum_{i=1}^{n} \ln \big(\varphi(Y_{ni} \mid \theta)\big)$$

- **Approximate maximum likelihood estimator:** Hess 1996

$$\Phi_n(\hat{\theta}_n) \geq \max_{\theta \in \Theta} \Phi_n(\theta) - \varepsilon_n, \quad \Phi_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \ln \big(\varphi(Y_{ni} \mid \theta)\big)$$

$\varepsilon_n \to 0 \ a.s.$

Consistency of maximum likelihood estimators

$$\hat{\theta}_n \in \underset{\theta \in \Theta}{\operatorname{argmax}} \sum_{i=1}^{n} \ln\big(\varphi(Y_{ni} \mid \theta)\big)$$

Theorem: Assume that $\varphi : [0,1]^d \times \Theta \to [0,+\infty)$ is continuous, Θ is compact, $d(x,A) = \min_{y \in A} d(x,y)$.

(i) $\Longrightarrow$

$$\boxed{\lim_{n\to\infty} d(\hat{\theta}_n, \Gamma) = 0 \quad a.s., \quad \Gamma = \underset{\theta\in\Theta}{\operatorname{argmin}}\, \Phi(\theta)}$$

Consistency of maximum likelihood estimators

$$\hat{\theta}_n \in \operatorname*{argmax}_{\theta\in\Theta} \sum_{i=1}^{n} \ln\big(\varphi(Y_{ni} \mid \theta)\big)$$

Theorem: Assume that $\varphi : [0,1]^d \times \Theta \to [0,+\infty)$ is continuous, Θ is compact.

(i) $\Longrightarrow$

$$\lim_{n\to\infty} d(\hat{\theta}_n, \Gamma) = 0 \quad a.s., \quad \Gamma = \operatorname*{argmin}_{\theta\in\Theta} \Phi(\theta)$$

(ii) If, in addition, $\varphi(. \mid \theta_1) \neq \varphi(. \mid \theta_2)$ for $\theta_1 \neq \theta_2$, identifiabil.

$\Longrightarrow$

$$\lim_{n\to\infty} \hat{\theta}_n = \theta_0 \quad a.s.$$

minimum Kullback-Leibler information consistency

Suzukawa et al. 2001

Asymptotic normality of ML estimators

$$G_{ij}(u) = \frac{\partial^2}{\partial\theta_i \partial u_j} \ln \varphi(u \mid \theta)\Big|_{\theta=\theta_0}, \quad \bar{G}_i(u) = \frac{\partial}{\partial\theta_i} \ln \varphi(u \mid \theta)\Big|_{\theta=\theta_0}$$

$$I(\theta) = \Big(I_{ij}(\theta)\Big)_{i,j=1\dots d}, \quad I_{ij}(\theta) = \int_{[0,1]^d} \frac{\partial^2}{\partial\theta_i \partial\theta_j} \ln \varphi(u \mid \theta) \varphi(u) \, \mathrm{d}u$$

Theorem: θ_0 interior point of Θ. Under regularity assumptions,

$$\boxed{\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, I(\theta_0)^{-1} \Sigma_1 I(\theta_0)^{-1})}$$

where $\Sigma_1 = \Big(\operatorname{cov}\big(\Gamma_i(X_{11}, \dots, X_{d1}), \Gamma_k(X_{11}, \dots, X_{d1})\big)\Big)_{i,k=1\dots q}$,

$$\Gamma_i(z) = \bar{G}_i(F_1(z_1), \dots, F_d(z_d)) + \int_{[0,1]^d} \sum_{j=1}^{d} G_{ij}(t) \mathbf{1}(F_j(z_j) \le t_j) \, \mathrm{d}C(t).$$

default term | from the plug-in F_n

Genest et al. 1995, Joe 2005, Chen et al. 2006

Semiparametric density estimators

$X_1, \ldots, X_n$ sample, $X_i = (X_i^{(1)}, \ldots, X_i^{(d)})^T$ has density h,
f_j density of $X^{(j)}$, F_{jn} empirical distribution function of $X^{(j)}$,
$\varphi(. \mid \theta_0)$ copula density of $(X^{(1)}, \ldots, X^{(d)})^T$,

- **Density estimators for f_j:**

$$\hat{f}_{jn}(x) = n^{-1} b^{-1} \sum_{i=1}^{n} K\left((x - X_i^{(j)}) b^{-1} \right) \quad (x \in \mathbb{R}),$$

for $j = 1 \ldots d$, $b = b(n)$ bandwidth, K kernel function

- **Estimator for h:**

$$\hat{h}_n(x_1, \ldots, x_d) = \varphi(F_{1n}(x_1), \ldots, F_{dn}(x_d) \mid \hat{\theta}_n) \cdot \hat{f}_{1n}(x_1) \ldots \hat{f}_{dn}(x_d)$$

$(x_i \in \mathbb{R})$ where $\hat{\theta}_n$ is a consistent estimator for θ_0.

Semiparametric density estimators

$$\hat{h}_n(x_1, \ldots, x_d) = \varphi(F_{1n}(x_1), \ldots, F_{dn}(x_1) \mid \hat{\theta}_n) \cdot \hat{f}_{1n}(x_1) \ldots \hat{f}_{dn}(x_d)$$

Assumption: $\left\| \hat{\theta}_n - \theta_0 \right\| = O\left(n^{-1/2} \sqrt{\ln(n)} \right) \quad a.s.$

Theorem: Assume that $(f_j)''$ is bounded. For any compact set $C \subset \mathbb{R}^d$,

$$\sup_{x \in C} \left| \hat{h}_n(x) - h(x) \right| = O\left(\sqrt{\ln n}(nb)^{-1/2} + b^2 \right) \; a.s.$$

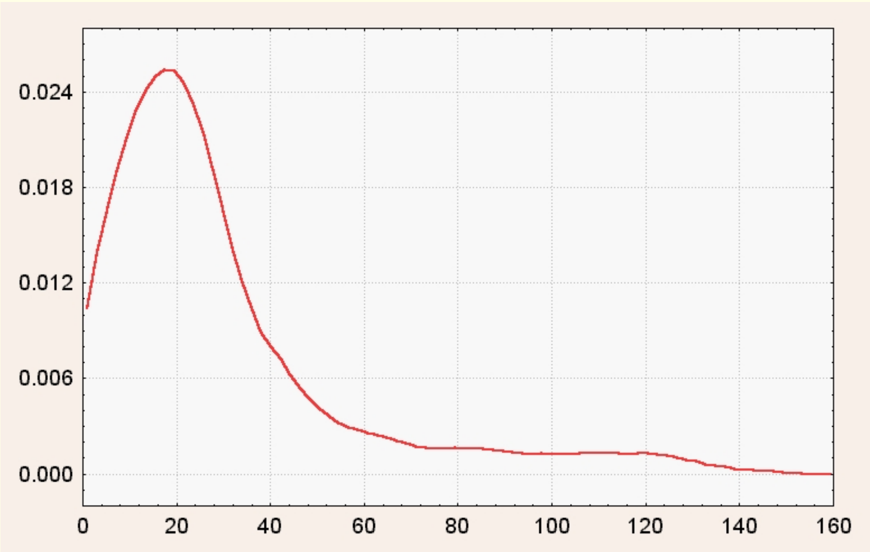
optimised rate: $O((\ln n/n)^{2/5})$

• The estimator $\hat{h}_n$ has the same convergence rate as known from univariate nonparametric density estimation.

Liebscher 2005

Semiparametric density estimators - Dataset liver disorders group 2

Marginal density of $X^{(3)}$



Sectional view of the density